USING OPTIMISATION TECHNIQUES TO SOLVE A PRODUCTION PROBLEM

WITH APPLICATION TO THE DEEP DRAWING PROCESS OF AN AUTOMOTIVE PART

ELSE VELDMAN

14th December 2006

Preface

This master thesis describes an application of the optimisation strategy for metal forming processes using time consuming Finite Element Method (FEM) calculations. This strategy is developed at the University of Twente in the project MC1.03162, *Optimisation of Forming Processes*. This project is part of the research programme of the Netherlands Institute of Metal Research (NIMR). The aim of the project is to develop an generally applicable optimisation strategy for metal forming processes which can be used in combination with any FEM programme. The problem described in this report shows the applicability of this strategy to an industrial problem.

The industrial problem concerns an automotive part of Volkswagen. Volkswagen and INPRO made this problem available for application in the optimisation project. I would like to thank INPRO for offering me the useful information on the work they did on the automotive part. In particular, I would like to thank Mr. Weiher, Dr. Kraska, Dr. Rietman, Mr. Zubert and Mr. Nitsche of INPRO for their input in the meeting they arranged to discuss the problem. Mr. Maschke of Volkswagen has been very helpful too and delivered useful information on the part. I thank FIuKA for the welcoming Martijn Bonte and me to their factory where we could see the process of the automotive part in practice. Also, I would like to thank Bart Carleer who always has been very helpful when I had any questions regarding the simulations in AutoForm.

During the performance of the research and the writing of the Master Thesis I worked at the research group of Applied Mechanics. Everybody at this group has always been very kind and helpful to me. I would like to thank professor Han Huétink and Ton van den Boogaard for their support. They have always shown real interest in my research.

In particular, I would like to thank Martijn Bonte. He has always been very involved in my work and challenged me to get the most out of my master thesis. He attributed directly and indirectly to the results I achieved. Besides that, he supported me in good and bad times and I would here like to thank him again for the cookies I received on a certain dull day in September.

Else Veldman
Enschede, December 2006

Samenvatting

Op de Universtiteit Twente wordt in samenwerking met de industrie een optimalisatiestrategie voor metaalomvormprocessen ontworpen. Eén van de indistriële partners is INPRO, een dochteronderneming van onder andere Volkswagen. Volkswagen is geïnteresseerd in de optimalisatie van het dieptrekproces van een automobiel onderdeel waarin scheurtjes ontstaan. Voor dit afstudeeronderzoek is deze strategie toegepast om dit probleem op te lossen.

Eerst is een korte studie gedaan naar de oorzaak van de scheuren in het onderdeel. Vervolgens is de optimalisatiestrategie toegepast. Deze strategie bevat een gestructureerde procedure die gevolgd kan worden om het probleem zodanig te modelleren dat er een wiskundig optimalisatie algoritme op toepasbaar is. Daarnaast bevat het een selectie procedure om het aantal variabelen te verminderen en een wiskundig algoritme dat gekoppeld kan worden aan een willekeurig Eindig Elementen Methode (EEM) simulatie pakket om het optimalisatie probleem op te lossen.

Voor het automobiel onderdeel twee processen worden onderscheiden: het *referentie proces* and het *gemodificeerde proces*. Dit gemodificeerde proces is het referentie proces aangepast door INPRO. Dit gemodificeerd proces heeft bewezen dat het het aantal scheurtjes in het onderdeel vermindert. De optimalisatiestrategie is toegepast om op twee manieren optimalisatie toe te passen:

1. Optimalisatie van het referentie proces om de resultaten te vergelijken met het gemodificeerde proces
2. Optimalisatie van het gemodificeerde proces om het schrootafval verder te verminderen

In eerste instantie is een optimalisatie van het referentie proces met zes variabelen uitgevoerd. Deze optimalisatie convergeerde niet naar een optimum. Uit nadere analyse bleek er sprake te zijn van numerieke ruis. Om de resultaten te verbeteren is de aandacht gericht op de modellering van het probleem. Dit leidde tot een vereenvoudigd probleem waarin slechts één variabele werd meegenomen. Dit probleem is opgelost met twee verschillende sets van waarden voor de onbelangrijke variabelen: één set bevat de orginele waarden van de variabelen en één set bevat de optimale waarden van de variabelen gebaseerd op de selectie resultaten. Voor deze selectie is een aantal sommen gedaan waaruit het effect van deze variabelen op de response functies is geschat. Na oplossen van het probleem is een optimum gevonden dat het proces verbetert ten opzichte van het referentie proces. De optimalisatie met de optimale waarden voor de variabelen resulteerde in een beter optimum dan de optmalisatie met de orginele waarden. Het gemodificeerde proces boekt nog altijd betere resultaten dan het geoptimaliseerde referentie proces.

Daarnaast is nog een tweede optimalisatie van het gemodificeerde proces uitgevoerd. Het optimalisatiemodel verschilt niet veel van het model van het referentie proces: de definities van de response functies zijn hetzelfde, alleen de variabelen verschillen. De selectie procedure is toegepast om onbelangrijke variabelen te elimineren. De selectie resultaten lieten zien dat er geen dominante variabelen zijn, daarom zijn er geen variabelen uitgesloten van het optimalisatie model en het resultaat is een optimalisatie model met elf variabelen. Omdat dit veel variabelen zijn voor de optimalisatiestrategie, wordt deze strategie verder niet toegepast. Voor optimalisatie worden de variabelen ingesteld op hun minimale of maximale waarde zoals deze zijn gedefiniëerd in het optimalisatiemodel. De keuze voor deze instellingen wordt gebaseerd op de resultaten die uit de selectie procedure volgden. Deze simpele optimalisatie leidde snel tot een grote verbetering ten opzichte van het aangepaste proces.

Voor de optimalisaties is het EEM programma AutoForm gebruikt. Gebaseerd op deze simulaties en de geïdentificeerde oorzaak voor de scheurtjes, leveren de optimalisaties goede resultaten. De AutoForm simulaties kwamen echter niet overeen met de realiteit en met simulaties in een ander programma. Bovendien is de oorzaak van de scheurtjes nog steeds onzeker. Dit maakt het moeilijk om sterke conclusies te trekken over het effect van de optimalisatie op het verminderen van de scheurtjes in realiteit.

Summary

At the University of Twente an optimisation strategy for metal forming processes is being developed in cooperation with the metals industry. One of the industrial partners of the UT project is INPRO in Berlin, which is a subsidiary of amongst others Volkswagen. Volkswagen is interested in optimising the deep drawing process of an automotive part in which cracks occur. For the research described in this report, the optimisation strategy is applied to solve this problem.

First, the cause of the cracks has been investigated and it was concluded that large deformation due to forming is causing the cracks. Subsequently, the optimisation strategy is applied. This strategy comprises a structured procedure to model the problem in such a way that a mathematical optimisation algorithm can be applied to it. Besides the modelling part, the strategy comprises a screening procedure to reduce the number of variables and a mathematical algorithm to solve the optimisation model. This algorithm can be coupled to any Finite Element Method (FEM) programme.

For the automotive part two processes are distinguished: the *reference process* and the *modified process*. The modified process is the reference process modified by INPRO. This modified process proved to reduce the number of cracks that occur in the part. The optimisation strategy is applied to two optimisations of the process of the automotive part:

1. Optimise the reference process to compare these results with the modified process
2. Optimise the modified process to further reduce scrap

Initially, for the first optimisation of the reference process an optimisation model with six variables has been defined. After an analysis of the responses, it appeared that the responses suffer from numerical noise. To improve the results the optimisation model was adjusted. This led to a simplified optimisation problem with only one variable. This problem has been solved with two different sets of variable settings for the unimportant variables: one set is contains the original settings for the design variables and one set contains the optimal variables settings based on screening results. For screening, a number of calculations are executed to estimate the effect of the variables on the responses. After solving this problem an optimum was found that improved the process with respect to the reference process. The optimisation with the optimal variable settings resulted in a better optimum that the optimisation with the original settings. The modified process still outperforms the optimised reference process.

Furthermore, a second optimisation of the modified process was performed. The optimisation model does not differ much from the one for the reference process: the definitions of the responses are equal, only the variables are different. Screening was performed to eliminate the unimportant variables. The screening results showed that there are no dominant variables. Therefore, no variables are excluded from the optimisation model and the result is an optimisation model with eleven variables. Because this are many variables for the optimisation strategy, this strategy is no further applied. For optimisation the variables are set to their minimum or maximum values as defined in the optimisation model. The choice for these settings is based on the screening results. This simple optimisation led to a large improvement with respect to the modified process.

For the optimisations the FEM programme AutoForm is used. Based on these simulations and the identified cause for the cracks, the optimisations delivered good results. However, the AutoForm simulations did not coincide with reality and simulations in other simulation programme and

the cause of the cracks is still uncertain. This makes it difficult to draw strong conclusions on the effect of the optimisation on reducing the cracks in reality.

Contents

Chapter 1

Introduction

At the University of Twente (UT) an optimisation strategy for metal forming processes is being developed in cooperation with the metals industry. For the project described in this report, this strategy is applied to an industrial problem; the application is the optimisation of a deep drawing process of an automotive part. In this chapter a short introduction to the optimisation project is given and the problem is described. In the last section the structure of the report is given.

1.1 Optimisation in the automotive industry

Until the 1980s trial-and-error was used to produce products which fulfil the demands of the customer. The operator chose initial settings for the process on the basis of his experience and if the product did not satisfy the demands he changed the settings once or several times until the product was sufficient. In the late 1980s the increasing power of computers allowed scientific researchers to develop methods for simulating metal forming processes using the computer. During the 1990s the Finite Element Method (FEM) became a well-known method for simulating forming processes. The simulations agreed better with reality and more and more use was made of them in the metal forming business. However, the iterative procedure of simulating still resulted in a sufficiently good final product, whereas competitive industrial companies became more interested in an optimised product instead of just a sufficient one. This is where the scientific field of mathematical optimisation can assist. By coupling optimisation algorithms to FEM simulations, product or process optimality comes into sight. From the year 2000 onwards optimisation in this field is picked up by scientific researchers and is the next challenge in the simulation of forming processes [11].

With the possibility to couple optimisation to FEM simulations optimisation becomes interesting for industrial companies. For example, the automotive industry is a very competitive business where optimisation can play an important role. Product improvement and cost reduction are major issues to ensure customer satisfaction and thus, to outperform competition. To achieve the desired product improvement and cost reduction, the production processes of cars need to be optimised. A large percentage of the total amount of car parts is made by metal forming processes. For example, each car contains about 300 deep drawn parts and its engine is built up from numerous forged products. Since metal forming processes are playing such an important role, optimising these type of processes will lead to significant product improvement and cost reduction in the automotive industry.

At the UT, the project *Optimisation of Forming Processes* aims at developing an optimisation strategy for these metal forming processes. This optimisation strategy comprises a guideline how to model the optimisation problem and a mathematical algorithm which can be coupled to a FEM package to solve the problem. Software with the name *OptformV2.0* has been written that implements this algorithm in MATLAB. This software can be used in combination with any FEM code for simulating metal forming processes. For this project the strategy is tested on the deep drawing process of an automotive part.

1.2 Problem description

One of the industrial partners of the UT project is INPRO in Berlin, which is a subsidiary of amongst others Volkswagen. Volkswagen is interested in optimising the deep drawing process of the automotive part shown Figure 1.1. How this bearing tube is part of a car is also shown in this figure.

Optimising a process can be done for two reasons:

- to improve an already feasible process;
- to improve an infeasible process by solving a problem.

In this case the process suffers from a problem. During or after the production of the bearing tubes cracks appear. The location of the cracks are indicated by the black lines in Figure 1.2. The production process has been modified by adding an annealing step, which reduces the scrap rate, and by adding a 100% crack control to preclude that a bearing tube with a crack leaves the factory.

INPRO already investigated the problem of the bearing tube. The process was simulated with FEM software and on the basis of this simulation several adjustments were tested. The outcome of this investigation led to modifying the process by adding an extra deep drawing step with a blank holder force. More information on this investigation can be found in Appendix A. The modification resulted in a better process and significantly reduced the scrap rate. But still, some cracks occurred. In the project described in this report optimisation techniques are applied to prevent the cracks from occurring. These techniques are as well applied to the original process as to the modified process. In the remaining of the report we will refer to the original process as the *reference process*.

For the optimisation the deep drawing process is simulated in the FEM simulation software AutoForm. To the simulation results of the deep drawing process the optimisation strategy developed at the UT is applied. There are two possibilities for optimisation:

1. Optimise the reference process to compare these results with the modified process

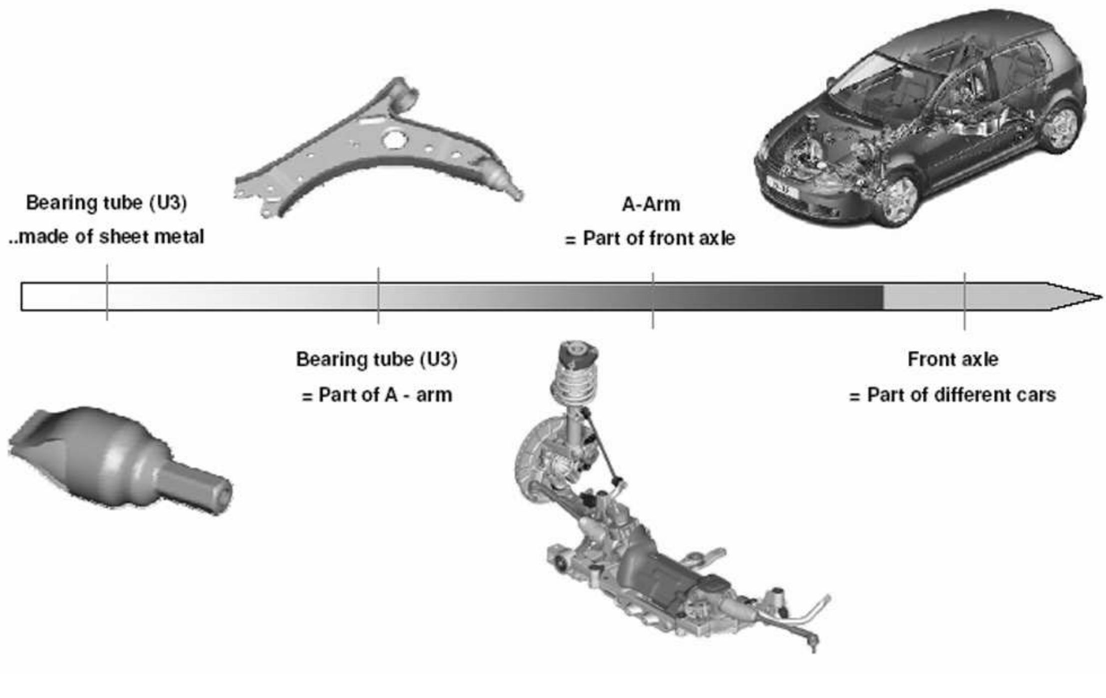


Figure 1.1: The position of the bearing tube in the car

Figure 1.2: The locations where the cracks occur

2. Optimise the modified process to further reduce the scrap rate

Both optimisations are performed during this project. It shows whether the at the UT developed strategy is good applicable for an industrial case as the bearing tube and whether it is successful. The knowledge which is gained during this project can be used for further improvement of the optimisation strategy for metal forming processes using FEM.

1.3 Outline report

This report describes the project of optimising the production process of a bearing tube. The cause of the cracks is investigated in Chapter 2. In Chapter 3 an overview is given of the theory behind the optimisation strategy which is being applied. Subsequently, the optimisation of the process of the bearing tube is treated in Chapters 4 and 5. The former describes the optimisation of the reference process and the latter describes the optimisation of the modified process. The report is completed with some conclusions and recommendations for future work in Chapter 6.

Chapter 2

Problem analysis: failure in deep drawing

In order to effectively remove the cracks in the bearing tube by optimisation techniques, it is essential to clearly understand the problem and the quantities (e.g. strains, residual stresses) that cause these cracks. If the wrong quantity is optimised, applying optimisation techniques is useless since it will never solve the occurrence of the cracks. The cause of the cracks is therefore thoroughly investigated in this chapter. Besides identifying which quantity needs to be optimised, it is important to know which boundaries limit the process. After all, the optimal process needs to be a feasible one. Also, for optimisation we need to know which parameters influence the process: which parameters influence the quantities that cause the cracks and which parameters influence the process boundaries. All these matters are treated in this chapter.

In Section 2.1 the deep drawing process of a simple cylindrical cup is treated to get a feeling of deep drawing processes in general. In this section possible failure mechanisms which limit this process are also described. This gives insight into the process boundaries and the parameters that are involved. The process of the bearing tube is the subject of Sections 2.2 and 2.3. In the first of these two sections, a complete picture of the situation is made and the simulation of the process is discussed. In the next section on the bearing tube, possible causes for the cracks are discussed. Finally, Section 2.4 contains some conclusions.

2.1 A cylindrical cup

This section handles the deep drawing process of a simple cylindrical cup. First, the process is treated in Section 2.1.1, including the forces which influence this process. In Section 2.1.2 different failure mechanisms which limit the deep drawing process of a simple cylindrical cup are discussed.

2.1.1 The deep drawing process

In Figure 2.1 the deep drawing process of a simple cylindrical cup is shown. It is a forming process under combined compression and tension. A flat plate is forced trough a die by a punch to form a closed-contoured, thin-walled product (circular, rectangular, oval, etc.). The material from the flange which is drawn inside forms the wall and is stretched in axial direction ($\varepsilon_z > 0$), while the material in the flange is stretched in radial direction, but compressed in tangential direction ($\varepsilon_r > 0, \varepsilon_\phi < 0$). An example of a deep drawn product is a sink or an aluminium beverage can. Also, a lot of automotive parts, like body panels, are formed by the process of deep drawing.

For a better understanding of deep drawing process the forces which influence the process will be examined. These forces limit the process: when maximum allowable forces are exceeded the process may fail. The most common failure mechanisms of deep drawing processes will be treated in Section 2.1.2 to get a good scope of the process boundaries of the deep drawing process. Here the forces are treated to get a feeling which parameters influence a simple deep drawing process.

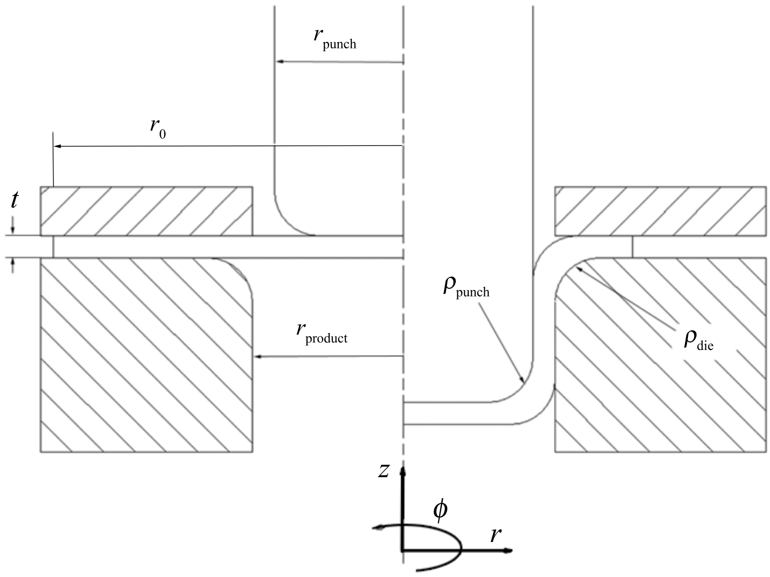


Figure 2.1: The deep drawing process of a cylindrical cup

The process of deep drawing includes radial drawing between the die and the blank holder (the deformation of the flange), bending and unbending under tension over the rounding of the die and the rounding of the punch, stretching of the blank between the die and the punch and stretching over the punch nose.

The total needed deep drawing force can be split up in the different partial forces needed to accomplish these parts of the process. These forces are the force needed to deform the flange F_1, the force to bend the blank F_2 and the forces to overcome the friction between the flange and the tools F_3 and between the blank and the rounding of die F_4. From the derivation of these (dimensionless) forces in [24], it follows that:

$$F_{1\,\max} = F(n,\beta,\varepsilon_0) \approx (0,75+0,15\sqrt{n}-0,35n^2)(\beta-1)\cdot\left(\frac{n}{e}\right)^n e^{\varepsilon_0} \tag{2.1}$$

$$F_2 = F(\varepsilon_0,\rho_{\mathrm{die}},t) \approx \frac{1}{2\frac{\rho_{\mathrm{die}}}{t}+1}\cdot\left(\frac{n}{e}\right)^n e^{\varepsilon_0} \tag{2.2}$$

$$F_3 = F(\mu_{\mathrm{flange}},r_0,t,n,\varepsilon_0) \approx 0,012\mu_{\mathrm{flange}}\frac{r_0}{t}\left(\frac{n}{e}\right)^n e^{\varepsilon_0} \tag{2.3}$$

$$F_4 = F(\mu_{\mathrm{die\,rounding}},\mu_{\mathrm{flange}},r_0,t,n,\beta,\varepsilon_0) \approx 1,6\mu_{\mathrm{die\,rounding}}(F_1+\frac{1}{2}F_2+F_3) \tag{2.4}$$

All parameters in these equations are summarised in Table 2.1. The total deep drawing force exists from these partial forces and thus depends on all these parameters. For the derivation of these forces, the strain hardening function defined by Equation 2.5 has been applied.

$$\sigma_f = (\overline{\varepsilon}-\varepsilon_0)^n \tag{2.5}$$

In this equation σ_f denotes the deformation resistance, C a characteristic value depending on the material and $\overline{\varepsilon}$ the true strain.

2.1.2 Failure mechanisms

A large part of the process is determined by the forces, as mentioned in Section 2.1.1. On one side the total force is needed to get the product deep drawn. On the other side, the force limits

Material parameters	
n	strain-hardening coefficient
ε_0	prestrain
Process parameters	
μ_{flange}	coefficient of friction along the flange
$\mu_{\text{die rounding}}$	coefficient of friction near the rounding of the die
Design parameters	
β	deep drawing ratio $\frac{r_0}{r_{\text{product}}}$
r_0	original radius of the blank
t	original thickness of the blank
ρ_{die}	rounding of the die

Table 2.1: Parameters of the deep drawing process of a simple cup

the process, because the product will fail if the total force exceeds the maximum allowable force. Failing due to too large a force is a common failure mechanism. However, there are other situations in which the product does not satisfy the designed shape. In this section we will take a look into these different failure mechanisms. The most common failure mechanisms which determine the constraints of a deep drawing process are:

- fracture;
- necking;
- primary and secondary wrinkling;
- earing.

Fracture

The most important process boundary which limits the critical deep drawing force is fracture. Fracture occurs when the maximum allowable tensile stress is exceeded. The fractures mostly occur in the wall or in the bending region of the product, because the wall and the bending region have to transmit the punch load to the flange and this causes a large tensile stress. In the literature three different types of fractures during the deep drawing process of a cylindrical cup are distinguished [24]:

1. The most common fracture which acts in the transition zone between wall and bottom in the bending region near the punch radius
2. The fracture just above the bending region in the wall
3. The premature fracture which acts in the bottom at the start of the deep drawing process

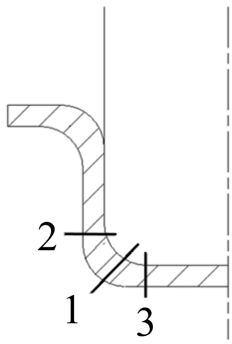


Figure 2.2: Places where fracture can occur in the deep drawn product

The places where these fractures occur is shown in Figure 2.2. A critical force F_{cr} which should not be exceeded by the total deep drawing force F can be derived for the first two cases [24]. In case of the fracture in the transition zone the critical force appears to depend on the anisotropy factor R, on the strain-hardening coefficient n and on the prestrain ε_0 of the material. For anisotropic material this critical force can be defined as follows:

$$F_{\mathrm{cr}} = F(R, n, \varepsilon_0) = \left(\frac{R+1}{\sqrt{2R+1}}\right)^{n+1} \left(\frac{n}{e}\right)^{n} e^{\left(\frac{\sqrt{2R+1}}{R+1}\varepsilon_0\right)} \tag{2.6}$$

The anisotropy factor R is in this formula based on the Hill'48 yield criterion. When the fracture occurs in the bending region, this means that the curvatures of the punch have influence on the tensile stress. The critical force then becomes:

$$F_{\mathrm{cr}} = F(R, n, \varepsilon_0, t, \rho_{\mathrm{punch}}, r_{\mathrm{punch}}) = \left(\frac{R+1}{\sqrt{2R+1}}\right)^{n+1} n^n \left\{\frac{t}{\rho_{\mathrm{punch}}} + \frac{t}{r_{\mathrm{punch}}} + e^{\left(n - \frac{\sqrt{2R+1}}{R+1}\varepsilon_0\right)}\right\}^{-1} \tag{2.7}$$

where ρ_{punch} is the rounding of the punch and r_{punch} the radius. The premature fracture is caused by badly chosen process conditions, for example the deep drawing ratio β or the rounding of the punch ρ_{punch}. Hence, the first two types of fracture are the most common ones. The size of the total deep drawing force depends on the parameters mentioned in Section 2.1.1. With a look at these partial forces it can be concluded that for a chosen material (with a given n, ε_0 and R) the friction coefficients may not exceed certain limits and also the ratios $\frac{t}{\rho_{\mathrm{die}}}$, $\frac{r_0}{t}$ and β may not be too large. Besides that, the critical force tells us that the ratios $\frac{\rho_{\mathrm{punch}}}{t}$ and $\frac{r_{\mathrm{punch}}}{t}$ are limiting the process.

Necking

An engineering stress-strain curve for the most common steels looks like Figure 2.3 [15]. Until the stress σ_{e} the material behaves elastically. After this point the material starts to deform plastically and gets a higher yield strength due to hardening. However, the cross section decreases. At first the decrease of the cross section is compensated by the higher yield strength, but at a certain point the yield strength cannot compensate for it anymore. Now, the cross section reduces heavily until the material fails. This phenomenon is called necking. Also in deep drawing processes necking can be the cause of product failure: it causes the wall of the deep drawn product to thin and after excessive thinning the cup will finally break. However, this necking is localised necking.

The difference between diffuse and localised necking is shown in Figure 2.4. Diffuse necking precedes localised necking. After the diffuse neck has started to form, it is accompanied by contraction strains in both the width and thickness direction. With a wide specimen, the width strain cannot localise rapidly, so the whole neck develops gradually and considerable extension is still possible after onset of diffuse necking. A condition will finally be reached where a sharp localised neck can form at an angle θ to the loading axis. Typically, the width b of the neck is of the order of the blank thickness t, so very little additional elongation is possible before failure [14].

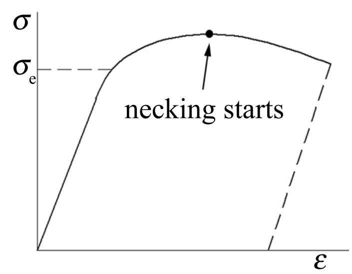


Figure 2.3: A typical stress-strain curve

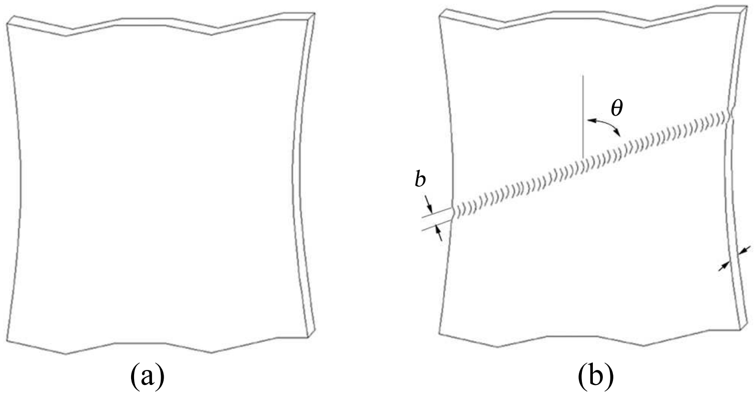


Figure 2.4: Diffuse necking (a) and localised necking (b) [23]

Necking is mainly influenced by the n-value of the material, by the prestrain ε_0 and by the anisotropy factor R. For material with higher R-values necking occurs less rapidly and thus high R-values positively affect the height of the deep drawn product.

It is not easy to express necking with a formula including these variables. To overcome this, Keeler and Goodwin introduced the Forming Limit Diagram (FLD). An example of an FLD is given in Figure 2.5. The Forming Limit Curve (FLC) which is shown in this figure indicates the limiting strains that sheet metal can sustain over a wide range of major to minor strain ratios. Usually the FLC is measured experimentally because, as mentioned, the theory for FLDs is not that straightforward. Once the FLC is determined the curve is supposed to stay the same. However, because the limit curve is dependent on several factors you should be careful assuming this.

In case of low carbon steel, it can be shown that the n- and R-values are not subject to a significant change during deformation and indeed can be assumed constant. However, when a product is deep drawn in different steps, the prestrain ε_0 differs in these steps. The influence of different values of the prestrain on the shape of the FLC of aluminium alloy 2008-T6 is shown in Figure 2.6. In Figure 2.6a the prestrains were in uniaxial tension (for $\varepsilon_2 < 0$ the prestrains were along the 1-direction and for $\varepsilon_2 > 0$ the prestrains were along the 2-direction). In each case the minimum corresponded to plane strain ($\varepsilon_2 = 0$) during testing. In the Figure 2.6b the prestrains were in biaxial tension.

It can also happen that there are some deviations between an experimental FLC and a theoretical approach of the FLC. These deviations can be caused by the strain rate exponent m which is

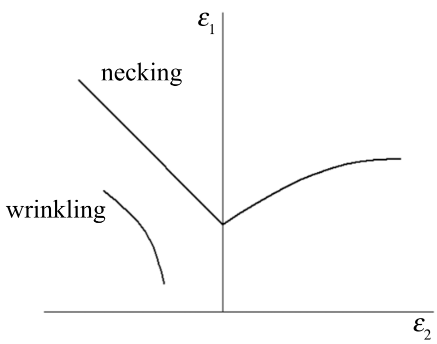
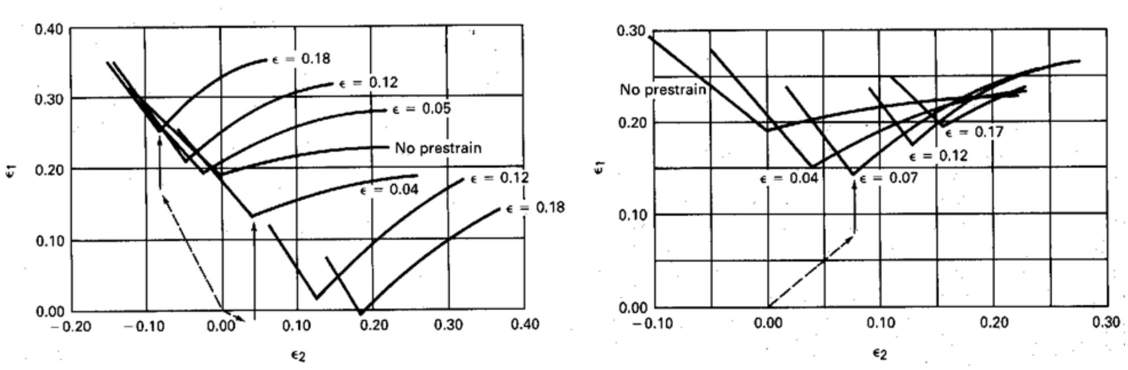


Figure 2.5: A Forming Limit Diagram



(a) (b)

Figure 2.6: Influence of prestrains in (a) uniaxial tension and (b) biaxial tension on the FLC of 2008-T6 aluminium [14]

not taken into account. The strain rate exponent can be defined as $m = \frac{\delta \ln \sigma}{\delta \ln \dot{\varepsilon}}$ and is in his turn dependent on the temperature. This m-value has influence on the minimum of the FLC. A change in m can cause critical strains to be up to 30% larger. Another deviation can be that the minimum of the FLC is not situated at $\varepsilon_2 = 0$ like theory tells us, but at slightly positive values of ε_2. This happens when the strains are measured at the outside surface of the blank while the plane strain situation is at the mid-plane of the blank.

Primary wrinkling

The compressive stress in tangential direction of the flange can be a cause of wrinkling. The remedy for this is to apply a blank holder. However, if the blank holder force is too large, the friction between the blank and the blank holder will become so large that the force to overcome this friction makes the total deep drawing force to exceed the maximum allowable force and fracture will occur. In Figure 2.7 it is shown how wrinkling depends on the blank holder force F and the deep drawing ratio β. For thinner blanks, the first critical line moves upwards (so wrinkling still occurs for a larger hold-down force than thicker blanks) while the second critical line moves downwards (so fracture occurs for a smaller hold-down force than for thicker blanks), narrowing the range where a good draw occurs. Wrinkling can also be indicated with help of an

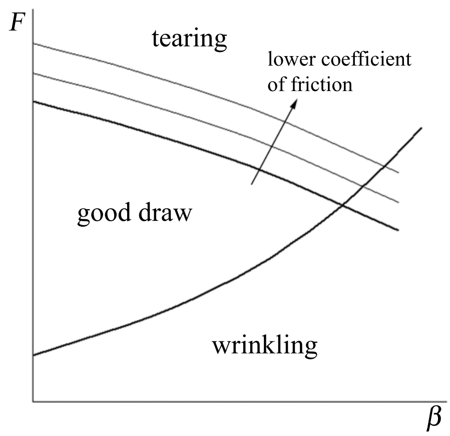


Figure 2.7: Wrinkling and tearing in the relation to F and β [5]

FLD, this is shown in Figure 2.5.

Besides the thickness t, experiments have shown that higher values of the Young's modulus E and the strain-hardening exponent n reduce the tendency of wrinkling. Also, it is shown that the R-value has influence on this type of wrinkling; higher R-values reduce the problem of wrinkling [23].

Another measure which is used to prevent wrinkling, is applying draw beads. Draw beads limit the amount of material drawn into the die, which has a beneficial effect on preventing compressive instabilities.

Secondary wrinkling

Another form of compressive instability is called secondary wrinkling. When the rounding of the die ρ_{die} is too large compared to the thickness of the blank t, a large part of the blank is not supported by the die which causes secondary wrinkling. In the same way, the ratio between the rounding of the punch and the thickness $\frac{\rho_{\text{punch}}}{t}$ may not be too large, to prevent secondary wrinkling near the rounding of the punch.

Earing

Anisotropy causes uneven flow, resulting in an uneven rim at the open end of the cup. When an even number of valleys and peaks are evident on the rim of the cup, the defect is called earing. A measure for the differences in the planar anisotropy, and therefore a measure for earing, is:

$$\Delta(R) = \frac{1}{2}(R_0 - 2R_{45} + R_{90}) \tag{2.8}$$

When $\Delta(R)$ differs more from 0, the earing will be stronger. More earing means that the blank needs to be larger, which results in a larger β and the process becomes more critical [1].

2.2 The bearing tube: the process

In this section the deep drawing process of the bearing tube is treated. In Section 2.2.1 the process of the bearing tube is discussed and in Section 2.2.2 the simulation of the process is treated.

2.2.1 The process

The bearing tube is being deep drawn in ten different steps. The process is shown in Figure 2.8. In OP02 to OP10 the bolt is formed. Next, the flange is trimmed and thereafter the flanges are further formed in two more deep drawing steps, OP16 and OP18. The blank is made of 3 mm thick steel; for deep drawing processes this is rather thick. The steel is a hot-rolled low carbon steel which has good deep drawing abilities. The European norm EN 10111 indicates the different qualities of hot-rolled steel for cold forming with the numbers DD10 through DD14 [3]. The higher the number, the better the formability of the steel. The steel used for this automotive part satisfies the norm for DD13. The material properties of the steel used for the bearing tube can be found in Appendix B.

The production of the parts takes place in the factory of Fischer & Kaufmann (FIuKA) which is the supplier of the product. After the production the parts are transported to a Volkswagen (VW) plant.

After the cracks had been signalised, FIuKA added an annealing step. During this annealing step, which takes place directly after the production, the parts are kept at a temperature of 200°C for two hours. Also, VW asked INPRO to take a better look on the matter. After an analysis with help of FEM simulations (see Appendix A) INPRO suggested to modify the process by adding a

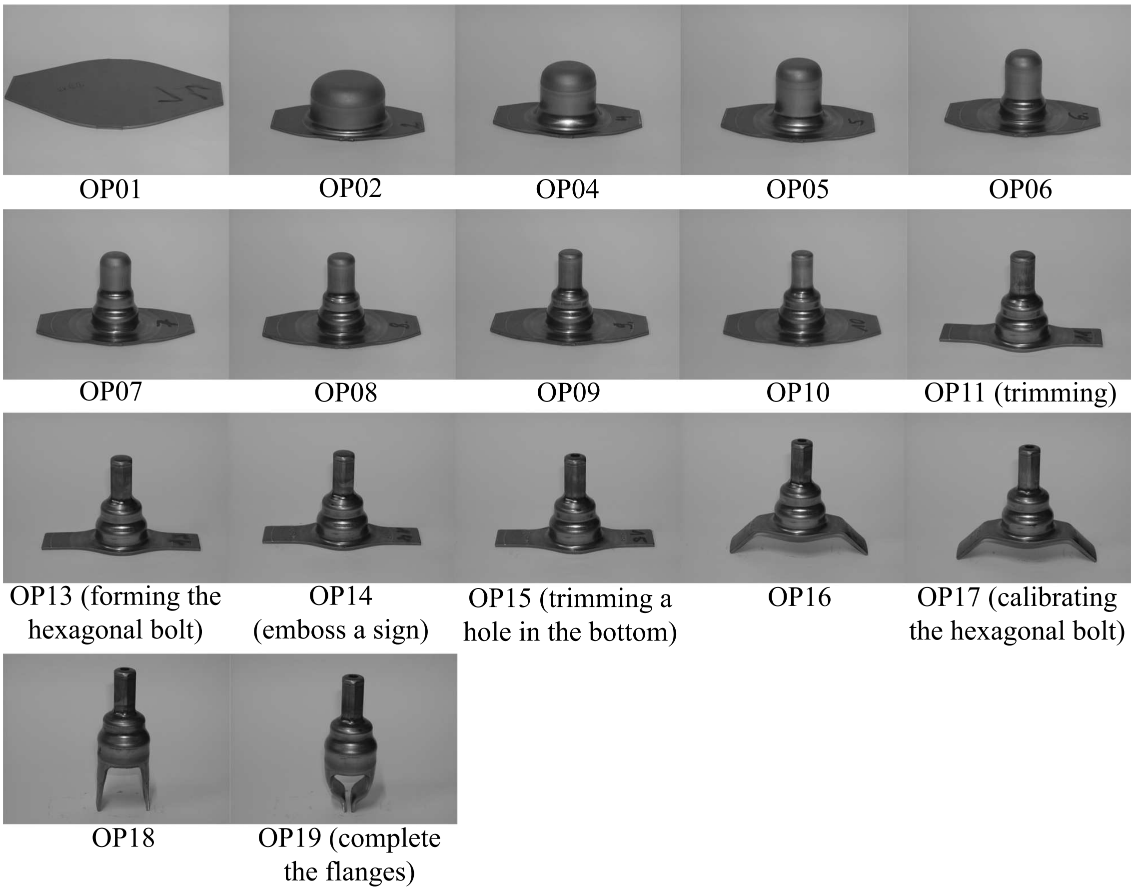


Figure 2.8: The process

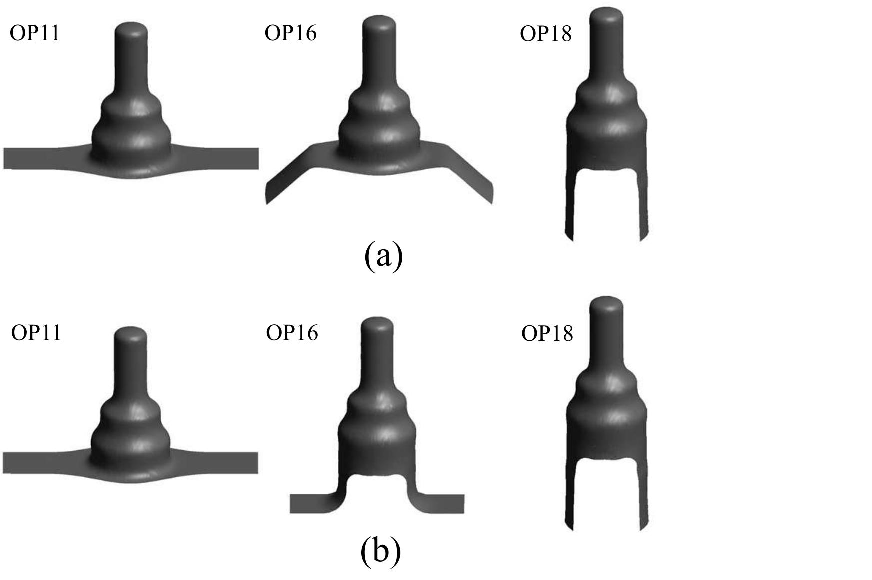


Figure 2.9: The last process steps of (a) the reference process and of (b) the modified process

deep drawing step with a blank holder force before the last deep drawing step OP18. In this step the first part of the flanges are deformed while a blank holder forces keeps the second part of the flanges in place. The original process, referred to as the reference process, and the modified process are shown in Figure 2.9.

These two adjustments reduced the scrap rate significantly, but did not prevent all cracks from occurring [29]. Now, VW is applying a 100% crack control to prevent a part with a crack from leaving the plant. This control takes place two days after the production of the parts in the FIuKA factory when the parts arrive at the VW plant.

2.2.2 Finite Element simulation of the process

The optimisation of complex forming processes like the bearing tube is becoming possible through the use of Finite Element (FEM) simulations. Therefore, simulating the process is an important part of the optimisation procedure. INPRO already simulated the process in the simulation programme INDEED, but since INDEED is not available at the University of Twente, AutoForm is chosen to be the FEM simulation software which is being used for optimising the process of the bearing tube. Both the reference process and the modified process were modelled in AutoForm and the results were validated by comparing them to the results of INDEED and to reality.

Comparison between the AutoForm and INDEED simulations

The process was already simulated in the computer programme INDEED by INPRO. Some results of the simulation in INDEED can be found in Appendix A. To be able to compare the simulation in INDEED with the simulation in AutoForm the used material models were made in

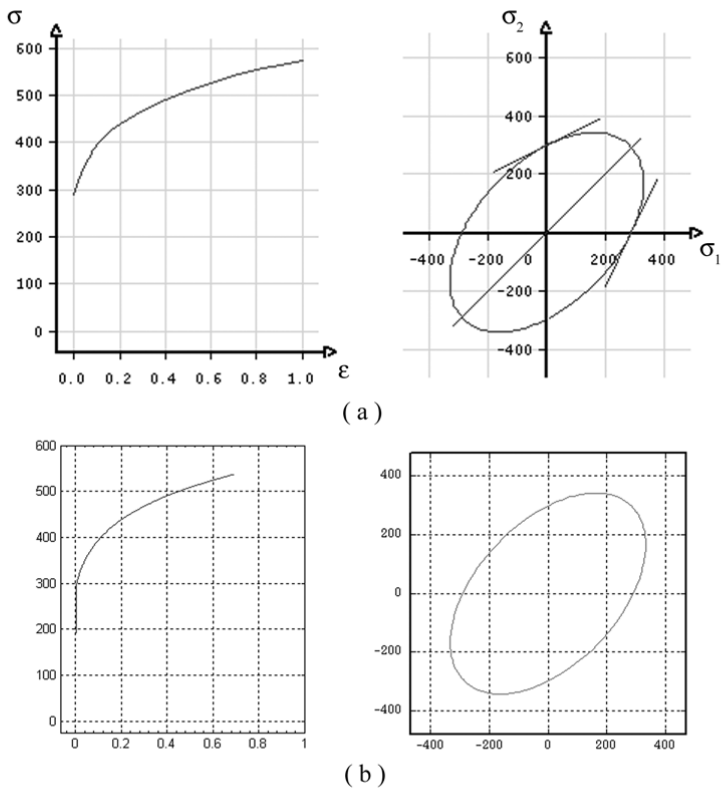


Figure 2.10: The stress-strain curves and yield loci used in AutoForm (a) and in INDEED (b)

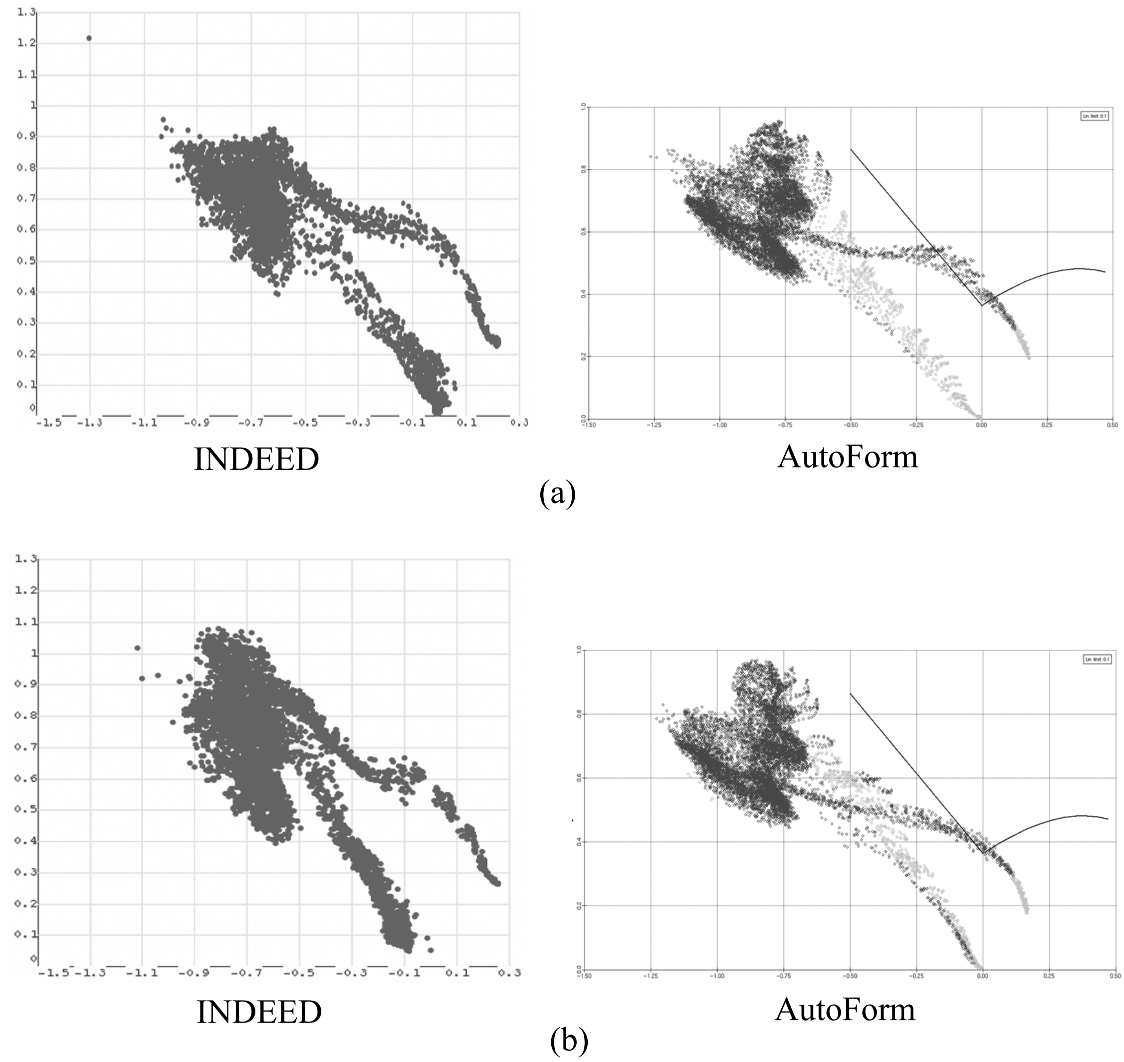


Figure 2.11: The forming limit diagrams after OP18 (a) in case of the reference process and (b) in case of the modified process

accordance with each other. In Figure 2.10 the stress-strain curves and the yield loci which are used for the simulation in AutoForm and for the simulation in INDEED can be seen. Besides the used material models, differences between the outcome of both simulations can be due to for example the use of different elements or different contact algorithms. Moreover, the blank used in the INDEED simulation has a slightly different shape than the blank used in the AutoForm simulation.

The simulations are compared by looking at the thickness of the final part and at the Forming Limit Diagrams (FLDs) following from both simulations. In Figure 2.12 the change in thickness is shown for the reference as well as the modified process. It can be seen that thinning and thickening take place on the same locations, but that the magnitude is quite different. The FLDs of both simulations are shown in Figure 2.11. Some differences exist, but it can also be seen that in both cases the exceptionally high strains in the upper-left corner of the FLDs, are decreased in case of the modified process.

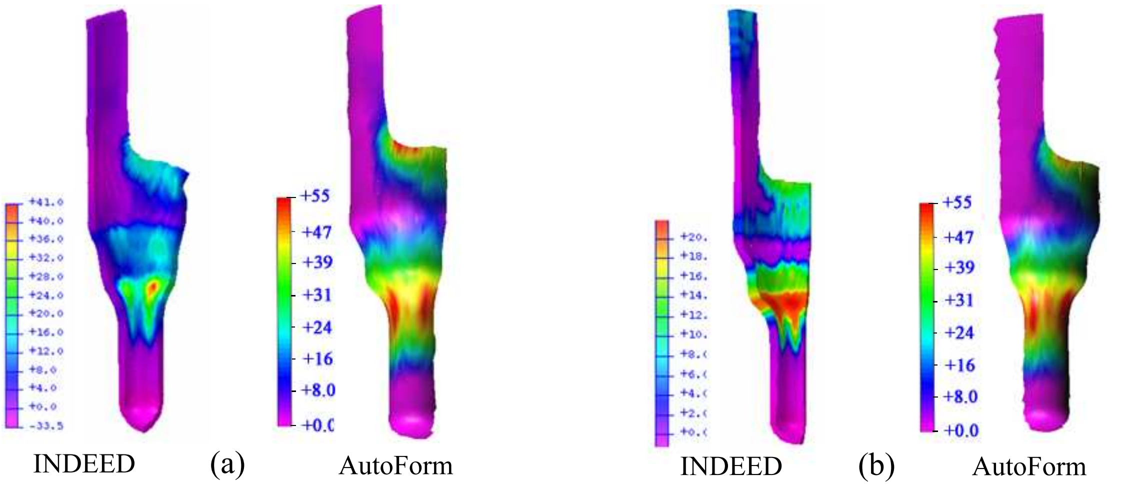


Figure 2.12: Change in thickness in % (a) in case of the reference process and (b) in case of the modified process

Comparison between AutoForm and reality

The thickness over the length of the bearing tube was measured with a part of a cross-section of the bearing tube that has been produced by the modified process. This thickness is compared to the thickness resulting from the AutoForm simulation of the modified process. The measured values of the thickness can be found in Figure 2.13c and the values of the thickness as simulated by AutoForm in Figure 2.13b. It appears that the AutoForm simulation underestimates the thickness in the top of the tube and overestimates the thickness near the edge.

Because we are most interested in thickness over the edge where the cracks occur (the edge is indicated in Figure 2.13a), this thickness was measured for more bearing tubes. Eleven bearing tubes which are made with the modified process and two which are made with the reference process are used for this measurement. The average thickness over the edge of these tubes for both the reference and the modified process are presented in Figure 2.14. The thickness over the edge that AutoForm simulates is also presented in this figure for both processes. It is clear that magnitude of the thickness over the edge between AutoForm and reality is rather different. However, the trend in the thickness is more or less comparable and for the modified process the thickness reduced in both cases.

For optimisation, the AutoForm simulations are used as a reference situation and the differences that exist between AutoForm and INDEED and between AutoForm and reality are no further

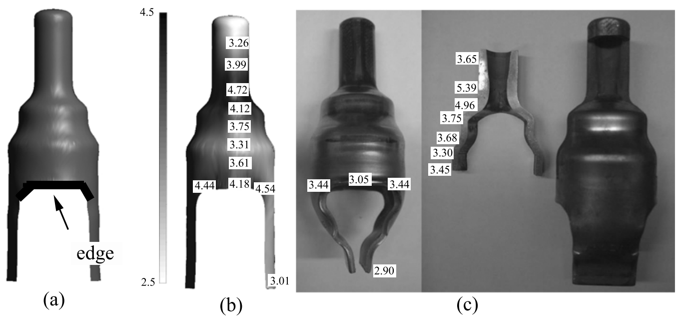


Figure 2.13: Comparison of the thickness in reality and in AutoForm

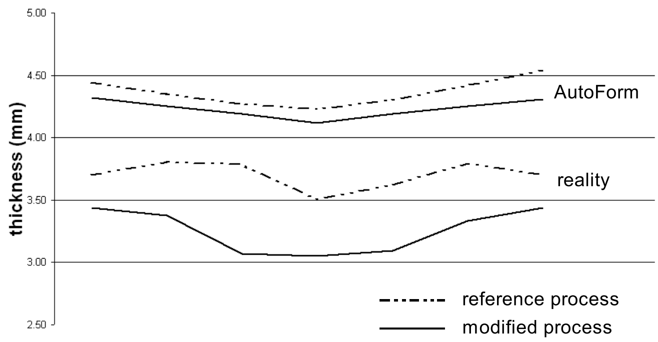


Figure 2.14: The thickness over the edge of the bearing tube, from the left till the right corner

discussed for this project.

2.3 The bearing tube: possible causes of failure

As could be read in Section 2.2.1, some modifications to the process have been implemented after the cracks were discovered. Although this reduced the scrap rate, it did not prevent all the cracks from occurring. There have been some tests, but it was not exactly clear yet what the problem is. To gather information that could reveal the cause of the cracks several initiatives have been undertaken. These initiatives and the resulting sources of information are indicated below.

- The meeting with Mr Maschke (VW) and INPRO in Berlin
  - Presentation Mr Maschke [18]
  - Presentation Mr Nitsche [22]
  - Material data DD13 (Norm EN 10111 - DD13) [3]
  - Metallographic photos
  - Bearing tubes with cracks and A-arm
- E-mail contact with Mr Maschke [17]
  - Metallographic photos
  - Report Volkswagen Qualitätssicherung [25]
- Communication with metallurgists
  - Conversation with metallurgists [4]
  - E-mail contact with a metallurgist [2]
  - Scientific publication on Cold Work Embrittlement [20]
- A visit to the manufacturer of the bearing tubes (FIuKA, M. Schröder and J. Müller) in Finnentrop [29]
  - Report Access [19]
  - Metallographic photos
  - Material data Hoesch (see Appendix B and [13])
  - Raw material

These contacts led to two possible causes for the cracks in the bearing tube.

1. Brittle fracture or Cold Work Embrittlement
2. Large deformations due to forming

These two possible causes are further discussed in Sections 2.3.1 and 2.3.2. Section 2.3.3 finishes with a discussion resulting in a final assumption on the cause of the cracks.

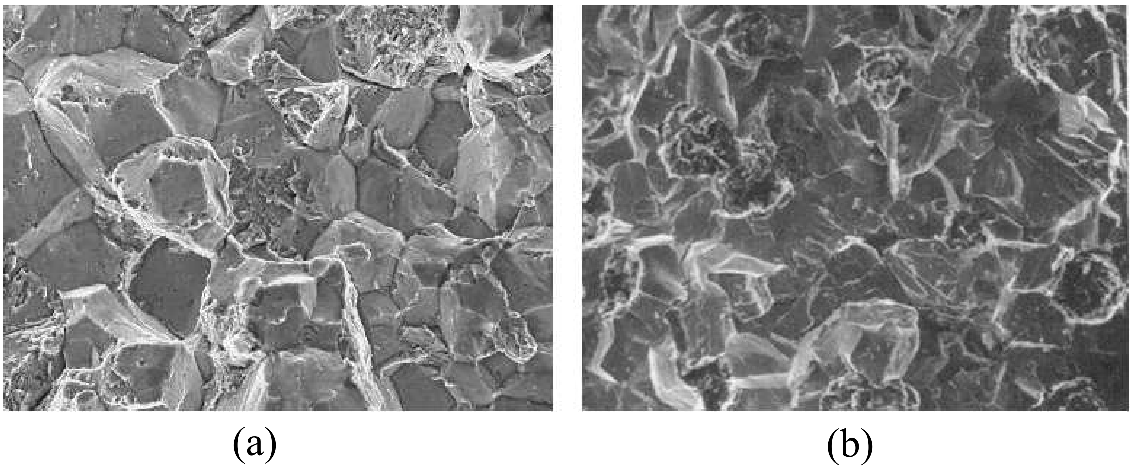


Figure 2.15: A transgranular fracture surface (a) and an intergranular fracture surface (b) [12]

2.3.1 Brittle fracture in deep drawing processes

The most common failure mechanisms in deep drawing processes have been discussed in Section 2.1.2. These failure mechanisms do not include brittle fracture. Brittle fracture does normally not occur in deep drawing processes, because in the process the material is subject to plastic deformation and therefore it needs to be ductile. However, with the development of stronger steels which are well deformable brittle fracture is an increasing problem. In this section brittle fracture within ductile materials will be treated.

A brittle fracture occurs by rapid crack propagation and without appreciable macroscopic deformation [12]. There are two types of brittle fracture that differ in relation to the strain rate and the temperature. In case of the first type, a change in temperature and in strain rate causes a ductile to brittle transition. The fracture is characterised by the propagation of a brittle crack along certain crystallographic planes and is *transgranular* (or *transcrystalline*) because the fracture passes through the grains. The other type is an *intergranular* fracture (crack propagation is along grain boundaries). This type of fracture normally results subsequent to processes that weaken or embrittle grain boundary regions [27]. Examples of a fracture surface for both these types of fracture are shown in Figure 2.15.

As mentioned before, deep drawing processes become more susceptible to brittle fracture. This is because more low carbon interstitial free (IF) steels are used. These steels have excellent deep drawing abilities which result from extra low carbon and nitrogen contents (<50 ppm) in addition to titanium and niobium microalloying for stabilising interstitial elements such as carbon. Thus these strong, good deformable steels seem very suitable for deep drawing processes and therefore it is more and more used. However, IF steels are susceptible to *Cold Work Embrittlement* (CWE). CWE is a phenomenon which occurs in secondary forming operations and causes a brittle fracture which is a combination of both types of fractures. Such a fracture starts as an intergranular crack due to weak grain boundaries and can propagate, after some time, as a combined intergranular and transgranular crack. Because it can take some time before the crack propagates, this fracture is called a delayed fracture. The delayed fracture develops in three stages: nucleation of the crack, growth to critical size and rapid propagation. The fracture starts as an intergranular crack. Subsequently a low applied stress can increase the concentration of elastic stresses in areas of large residual stresses what makes forming and propagation of cracks possible. The rapid crack propagation can be energetically more profitable through the grains, so the intergranular nucleation of cracks can be followed by a combination of intergranular and transgranular propagation.

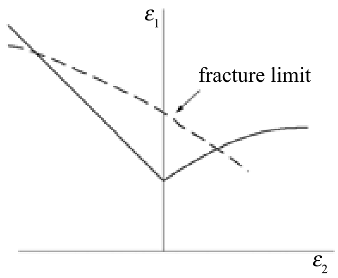


Figure 2.16: A fracture limit in an FLD

The propagation of the crack is dependent on the temperature and the applied stress (residual stresses in the material and externally applied stresses). Besides that, grain boundary weakening elements, such as phosphorus, increase the susceptibility to CWE, where grain boundary strengthening elements, such as boron and free carbon, decrease the susceptibility. Another interesting note is that there seems to exist a relation between the equivalent plastic strain and the ductile to brittle transition temperature which is of importance in case of CWE [20].

2.3.2 Large deformations

Large deformations due to forming can cause fracture. The ten step deep drawing process of the bearing tube causes the prestrain ε_0 to differ. In Section 2.1.2 it was mentioned that this influences the position of the Forming Limit Curve (FLC). Figure 2.6a showed that prestraining in 1-direction, which is the case in the flange during deep drawing processes, causes the FLC to raise for larger minor strains and the necking limit will be exceeded at higher strains. For small minor strains the FLC barely changes. The variation in prestrain is thus not expected to be the cause of the cracks in the critical areas which are dealing with small minor cracks after the last deep drawing step.

However, besides the necking limit it is also possible to show a fracture limit in a Forming Limit Diagram (FLD). An example of a fracture limit curve in an FLD is shown in Figure 2.16. Usually local necking occurs before fracture, but with an unfavorable combination of load and material structure, fracture can occur before necking becomes a problem.

2.3.3 Discussion

In case of the bearing tube there are reasons to believe that the brittle fracture as described in Section 2.3.1 is causing the cracks. Brittle fracture as cause for the cracks in the bearing tube is supported by the following observations:

- Delayed cracking is reported for the bearing tubes [18, 29];
- Metallographic photos present intergranular cracks (see Figure 2.18);
- The DD13 material norm indicates a quite high percentage of phosphorus (0.030%) [3], which weakens the grain boundaries;
- An annealing heat treatment to reduce the residual stresses has significantly reduced the scrap rate [29, 17].

However, a closer look at the actual material data provided by Hoesch reveals a significantly lower percentage of phosphorus (see Appendix B [13]) than mentioned in the norm, which makes the material less sensitive to Cold Work Embrittlement (CWE). Besides that, it is observed that the material is well suited for the application [25] and metallurgists indicated the material is not likely to be susceptible to CWE [4]. Furthermore, the influence of modifying the process on the equivalent plastic strain is evaluated. Figure 2.17 present the equivalent plastic strain for both the reference and the modified process. The plots show that there is no change in the equivalent

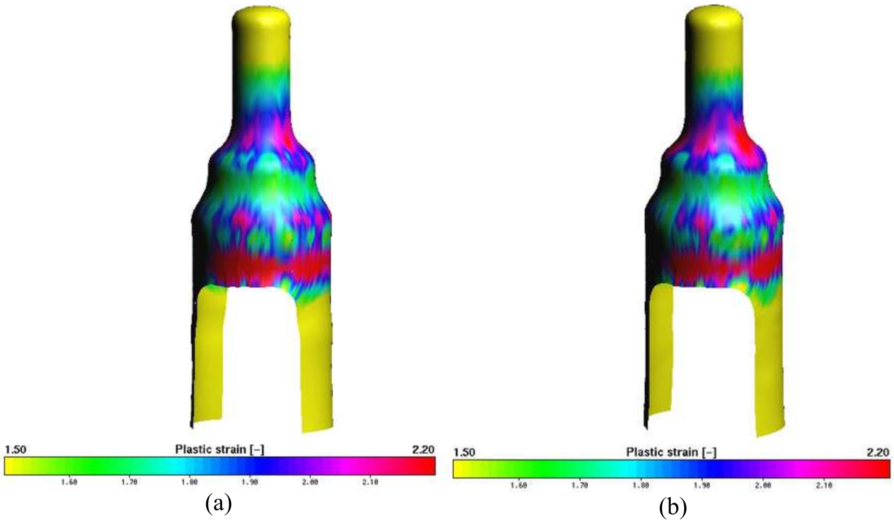


Figure 2.17: The equivalent strain in (a) the reference process and (b) the modified process.

plastic strain. This supports the conclusion that the cause of the cracks is not influenced by the equivalent plastic strain which seems to be related to CWE [20].

On the other hand, an investigation of VW tells us that the cause of the cracks should be sought in a critical forming step when the flange is drawn in the direction of the bolt (OP18). During this forming step the strains at the critical areas increase considerably. Due to work hardening at this location, the material loses its ductility which can finally lead to cracks [25]. Metallurgist confirm that this can be a possible cause [4]. This cause is supported by the following observations:

- Finite Element simulations performed by INPRO have shown large strains in the critical areas (see Figure 2.11. Reducing these strains by adding another forming step before Operation 18 has significantly reduced the scrap rate [29, 22];
- An annealing heat treatment to reduce the effect of work hardening in the critical areas has also reduced the scrap rate [29, 17].

The above observations lead us to believe that the large deformations due to forming is the cause to focus on for optimisation. The cracks are not (always) there directly after the deep drawing process (in that case, annealing would not help), but the large amount of deformation and work hardening is responsible for the delayed crack occurrence. Therefore, we will - just as was done by INPRO - focus on reducing the major- and minor-strains, rather than the equivalent plastic strain and residual stresses that would have been more interesting in case brittle fracture were the cause.

2.4 Conclusions

This chapter provided background on deep drawing processes by treating the deep drawing process of a simple cylindrical cup. Also, the most common failure mechanisms for deep drawing processes are treated:

- fracture;
- necking;
- primary and secondary wrinkling;
- earing.

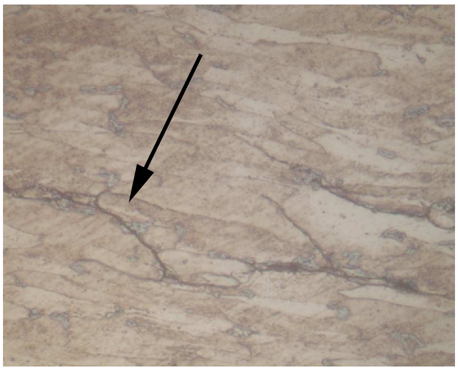


Figure 2.18: Intergranular crack

This gave insight into the different parameters that influence the deep drawing process. Table 2.2 gives an overview of the variables that influence the deep drawing process. In this table F is the deep drawing force and F_{cr} the critical force related to fracture. r_{punch}, r_0, ρ_{punch} and ρ_{die} denote respectively the radius of the punch and the original radius of the blank and the rounding of the punch and the die.

This chapter also introduced the process of the bearing tube. Both the reference and the modified process as proposed by INPRO are mentioned. FEM simulations of both processes have been made in AutoForm. These simulations are compared to reality and to the simulations made by INPRO in the FEM programme INDEED. The differences are discussed, however the AutoForm simulations are used as a reference situation for the optimisations that are performed in Chapters 4 and 5.

After having gathered further information on the process, two possible causes for the cracks were identified:

1. Brittle fracture or Cold Work Embrittlement (CWE)
2. Large deformation due to forming

parameter		**influence parameter on the deep drawing force F and on the different failure mechanisms**
Material parameters		
n	strain-hardening coefficient	F, F_{cr}, necking, primary wrinkling
ε_0	prestrain	F, F_{cr}, necking
R	anisotropy factor	F_{cr}, necking, primary wrinkling, earing
E	Young's modulus	primary wrinkling
Process parameters		
μ	friction coefficient	F
Design parameters		
β	deep drawing ratio	F
t	thickness	primary wrinkling
$\frac{r_{\mathrm{punch}}}{t}$		F_{cr}
$\frac{r_0}{t}$		F
$\frac{\rho_{\mathrm{punch}}}{t}$		F_{cr}, secondary wrinkling
$\frac{\rho_{\mathrm{die}}}{t}$		F, secondary wrinkling

Table 2.2: Parameters of the deep drawing process of a simple cup

These two causes are further discussed and different observations led to the assumption that large deformation in the last deep drawing step are causing the cracks.

Chapter 3

Optimisation theory

For this project the optimisation strategy for metal forming processes which is being developed at the University of Twente (UT) is applied. In this chapter the theory behind this strategy is treated. Section 3.1 starts with a short overview of the optimisation strategy. The other sections treat the theory behind the different steps in this optimisation strategy as can be seen in Figure 3.3. Section 3.9 finalises this chapter with some conclusions.

3.1 The optimisation strategy developed at the University of Twente

The optimisation strategy couples a mathematical optimisation procedure to Finite Element Method (FEM) calculations. Making the optimisation problem suitable for optimisation is an important part of the procedure, this is done by modelling. The interaction between modelling the problem and solving it is shown in Figure 3.1.

The modelling should be done cleverly to prevent that the optimisation problem that was modelled can finally be solved efficiently by a suitably chosen optimisation algorithm. Modelling should result in a set of variables to describe the design alternatives, an objective function and a set of constraints. How these quantities relate to FEM is displayed in Figure 3.2. Some quantities are known beforehand, where others are not. Design variables and explicit constraints which explicitly depend on these design variables are known beforehand and are the input for FEM. The output of the FEM simulation delivers quantities which implicitly depend on the design variables. The objective function is generally such an implicit quantity and it is also possible to have implicit constraints.

After the problem is modelled it can be solved by applying a suitable optimisation algorithm. For the strategy a metamodel based sequential approximate optimisation algorithm is chosen. This algorithm is extensively described by Bonte in [11]. Background on choices made for this algorithm is provided in [6, 8, 9, 16]. The steps the algorithm comprises are shown in Figure

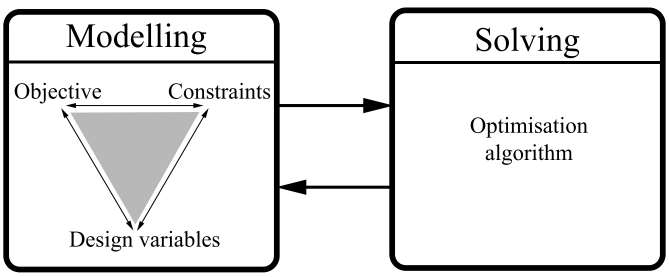


Figure 3.1: Interaction in mathematical optimisation [11]

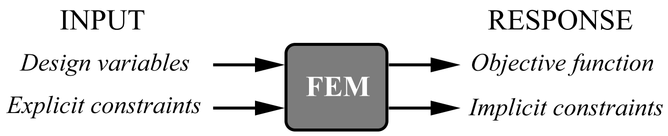


Figure 3.2: Input-response model for FEM in relation to optimisation [11]

3.3. It starts with a Design Of Experiments for cleverly choosing experimental points for which responses should be calculated by running FEM simulations. Next, a metamodel is fitted through these responses. This metamodel approximates the true response function. Optimising this model by applying an algorithm will give the solution of the optimisation problem. Because this is an approximation it needs to be validated by running a FEM calculation with the optimal settings. When the solution is not accurate enough more experimental points need to be added and the process starts over again until an optimum is found which is accurate enough. This process is called sequential improvement. Finally, the result of the optimisation procedure provides the optimal values for the design variables.

As shown in Figure 3.1 there exists interaction between the modelling and the solving of the optimisation problem. It means that the optimisation algorithm should be compatible with the model of the optimisation problem in order to solve the optimisation problem both effectively and efficiently. With this in mind it should be noted that the optimisation algorithm which is used for the strategy suffers from the *curse of dimensionality*, which means that it becomes exponentially more time consuming to solve when more design variables are taken into account. Therefore the number of variables will be limited as much as possible. This is done by screening the variables. In the screening process the influence of every variable on the objective function is evaluated (taking into account the set of constraints), making it possible to choose an appropriate set of variables for the optimisation.

The different steps of the optimisation strategy are treated in the next sections as displayed in Figure 3.3.

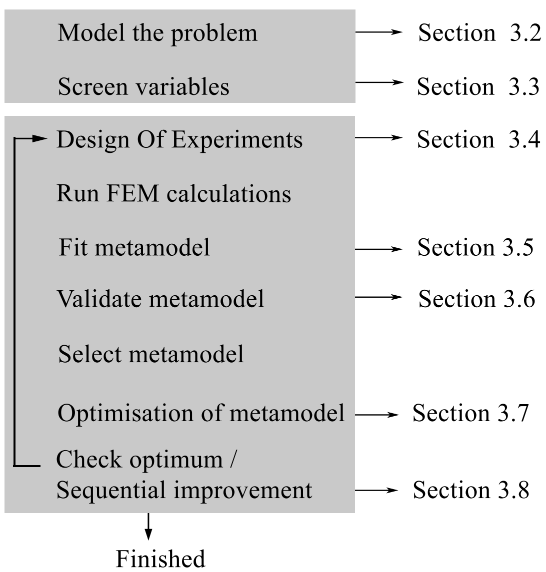


Figure 3.3: The optimisation strategy developed at the University of Twente

3.2 Modelling

The first step of any mathematical optimisation procedure is modelling the problem. This is an important step which always should be done carefully, because if the problem is not properly modelled, the optimisation will not give a useful result. A mathematical optimisation model consists of:

- design variables;
- an objective function;
- constraints.

Bonte proposes a structured procedure to model optimisation problems in metal forming processes [10]. With this procedure the variables, the objective function and the constraints are identified by following different steps. This procedure is applied to the optimisation strategy. In this section this procedure as elaborated in [10] is summarised.

The procedure is based on the Product Development Cycle, which is a part of the Product Life Cycle. The Product Development Cycle is equal to the Product Life Cycle minus the stages 6 and 7 as presented in Figure 3.4. The Product Development Cycle is applicable to any kind of product. When applied to metal formed products, there are five groups of quantities which can be identified to be influencing the Production Development Cycle (these are indicated in Figure 3.4):

- Functional Requirements: these are product properties that are critical to customer satisfaction and product functionality;
- Design Parameters: these define the product design;
- Process Variables: these are process settings necessary to manufacture the product;
- Defects;
- Costs.

Relevant metal forming quantities can be subdivided into these groups. With help of top-down structures which are made for each group all metal forming related quantities can be defined. Examples of these top-down structures can be found in [[10]]. For modelling optimisation problems in metal forming a seven step procedure based on this Product Development Cycle can be

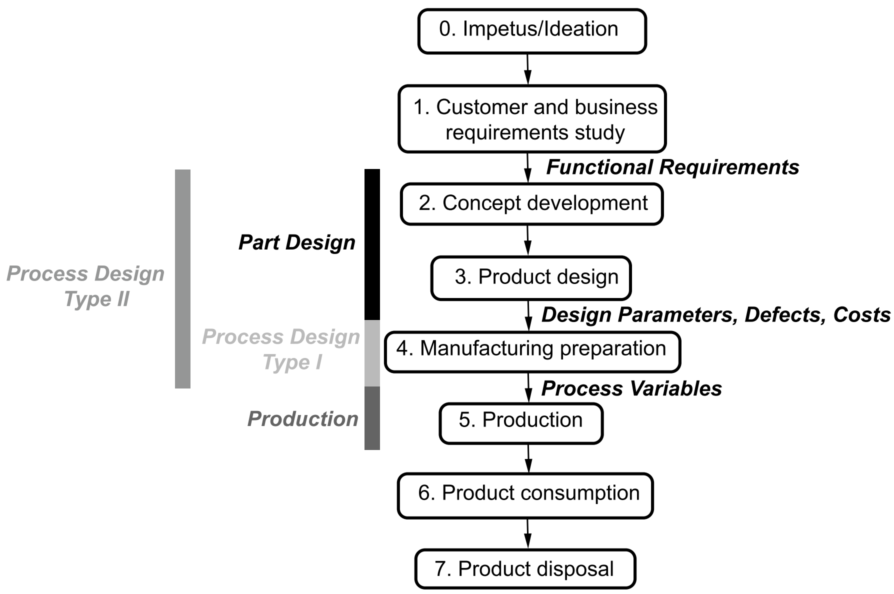


Figure 3.4: The Product Life Cycle applied to metal forming [10]

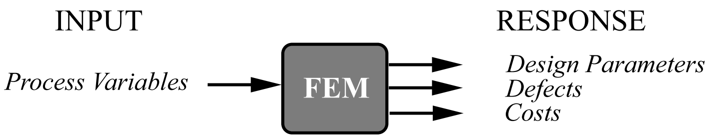


Figure 3.5: Input-response model for FEM during the Process Type I situation [10]

applied to determine which quantities determine the objective function, the design variables and the constraints. It includes the following steps:

1. Determine the appropriate optimisation situation
2. Select only the necessary responses
3. Select one response as objective function, the others as implicit constraints
4. Quantify the objective function and implicit constraints
5. Select possible design variables
6. Define the ranges on the design variables
7. Identify explicit constraints

Step 1: Determine the appropriate optimisation situation

For modelling optimisation problems in metal forming processes, we focus on the stages 2 through 5 of of the Product Life Cycle. Within this part of the Product Design Cycle, four situations in which the optimisation of metal products and their manufacturing processes can play a role are distinguished:

- Part Design, where it is aimed to optimise the metal formed product's Functional Requirements by determining the Design Parameters.
- Process Design Type I, where it is aimed to exactly obtain the Design Parameters set by the part designer by determining the Process Variables. Alternatively, one can aim to manufacture a defect free product or minimise production costs by determining the Process Variables.
- Process Design Type II, where it is aimed to optimise the metal formed product's Functional Requirements by determining the Process Variables. Design Parameters, Defects and Costs can still play a role in parallel to the Functional Requirements.
- Production, where optimisation techniques can be used to solve manufacturing problems.

These situations and their relations to the Product Development Cycle are also presented in Figure 3.4. To demonstrate how optimisation can be applied to one of these four situations, compare the input-response model for optimisation in Figure 3.2 to that for a Process Design Type I problem in Figure 3.5. Note the resemblance: one can immediately observe that for a Process Design Type I problem, the objective function and implicit constraints are the Design Parameters, Defects and Costs, whereas the design variables are related to the Process Variables.

Steps 2 to 4: Define responses

In steps 2 to 4 the responses are defined. In step 2, top-down structures can be used to select the necessary responses for the specific problem. Subsequently, in step 3, one of the defined response quantities is selected as objective function, the others as implicit constraints. Next, the responses need to be quantified. This is done in step 4.

Table 3.1 is proposed for selecting the final mathematical formulations of the objective function and implicit constraints. These depend on the sort of FEM output that is used for the responses (nodal/element values or not) and on the question if the objective function aims at a target or not. For the implicit constraints upper and lower limits are defined which may not be exceeded by the response. Furthermore, Table 3.1 subdivides the nodal/element values related responses

Type of response	No nodal/element value	Nodal/element value, critical	Nodal/element value, non-critical
Objective function, no target	$\min X$	$\min \max_N X$	$\min \frac{\sum_N(X)}{N}$
Objective function, target = X_0	$\min \lvert X - X_0 \rvert$	$\min \max_N \lvert X - X_0 \rvert$	$\min \frac{\lvert\lvert X-X_0 \rvert\rvert_2}{\sqrt{N}}$
Implicit constraint, USL	$X - USL \leq 0$	$\max_N(X - USL) \leq 0$	
Implicit constraint, LSL	$LSL - X \leq 0$	$\max_N(LSL - X) \leq 0$	

Table 3.1: Response quantification [10]

further into critical and non-critical values. Critical values are the values for which none of the nodal/element values is allowed to exceed a specified level. If it is acceptable that some of the nodal/element values exceed this specified level, but important that the average response value performs well, the response is non-critical. Constraints are by definition critical values.

For the usage of the software belonging to the applied optimisation strategy, the objective function needs to be defined in such a way that it is minimised. The formulation of a constraint should be such that it delivers a positive value when the limit is exceeded and the response does not satisfy the constraint.

Step 5 to 7: Define input quantities

In steps 5 to 7 the FEM inputs are defined. These are the design variables and explicit constraints which depend on these variables. Steps 5 and 6 comprise the selection and quantification of the design variables. The optimisation problem selected in step 1 determines the groups of design variables to take into account. Subsequently, top-down structures can be used to select the variables. This is done in step 5. In step 6 the upper and lower bounds of all design variables are determined. Step 7 deals with the explicit constraints. Explicit constraints describe impossible combinations of the design variables.

This 7 step methodology is generally applicable to any metal forming problem and yields a specific mathematical optimisation model, which can subsequently be solved using a suitable optimisation algorithm.

3.3 Screening

Since the applied optimisation strategy suffers from the *curse of dimensionality*, it is useful to limit the number of variables taken into account for optimisation. This is done by a screening experiment. Such an experiment is designed to investigate variables with a view toward eliminating the unimportant ones. The influences of the different variables on the responses are investigated and these influences are evaluated by a Pareto analysis. A Pareto analysis is based on the 80-20 rule which is supported by experiences in many fields. For example, in many stores 80% of the profit is realised by 20% of the products. Likewise, it is often the case that 80% of quality problems are caused by 20% of the possible causes. In this case, it will be tried to obtain the variables that determine 80% of the variations in the responses by a Pareto analysis. After the number of variables is reduced, optimisation will be more efficient and require fewer FEM runs.

First, a screening experiment is set up to investigate the influence of the variables on the objective function and the constraints. For this screening experiment experimental points are cleverly chosen. Experimental points are points in the design space for which FEM runs are performed so the design space is projected by a number of numerical experiments. Goal of this screening experiment is to demonstrate which variables are most important by only having to run a few FEM simulations. Widely used experimental designs for this purpose are so-called factorial designs. Myers and Montgomery give more background on the experimental designs treated in

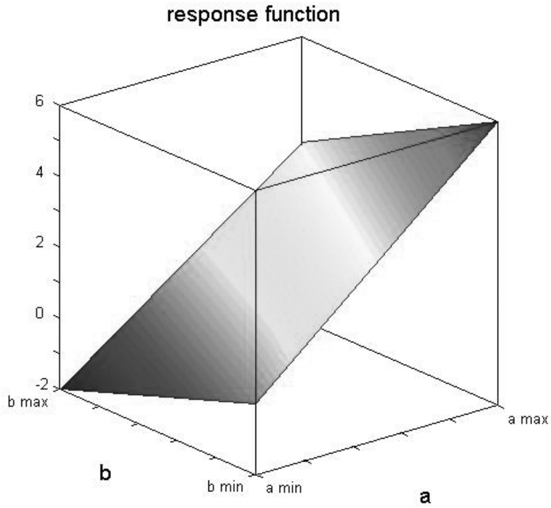


Figure 3.6: The response function of a 2^2 factorial design

this section [21]. After running the FEM simulations, the effect of the variables on the responses is analysed by Pareto and effect plots.

3.3.1 Two-level full factorial designs

In case of screening for the optimisation strategy the variables vary between only two values, their minimum and their maximum value. It is tried, by setting up several computer experiments with different combinations of these input values, to determine the influence of the different variables on the optimisation problem. This is an efficient way for getting a rough estimation on the influence, because it requires only few calculations.

Such an experimental design is called 2^k factorial design because it defines exactly 2^k runs. We will demonstrate this by giving the example of a 2^2 full factorial design which can be used in case of two variables. This design denotes four experiments with the following combinations of the variables a and b:

- $(a_{\min}, b_{\min})$
- $(a_{\min}, b_{\max})$
- $(a_{\max}, b_{\min})$
- $(a_{\max}, b_{\max})$

An example of such an experimental design and its response can be seen in Figure 3.6. With the response of these runs the main effects and interaction effect of the variables a and b on the response can be evaluated. The magnitude and direction of the effects can be examined to determine which variables are likely to be important. For example, in Figure 3.6 it can be seen that the main effect of variable a is significantly larger than the main effect of variable b.

3.3.2 Fractional factorial designs

With an increasing number of variables in a 2^k factorial design, the number of required runs for a complete replicate of the design grows rapidly. For example, a complete replicate of the 2^6 design requires 64 runs. In this design only 6 of the 63 degrees of freedom are used to estimate the main effects, and only 15 degrees of freedom are used to estimate the two-factor interactions. The remaining 42 degrees of freedom are associated with three-factor and higher interactions. If certain high-order interactions are negligible, then information on the main effects and low-order interactions may be obtained by running only a fraction of the complete factorial experiment. These fractional factorial designs are among the most widely used types of design in industry. The purpose of screening is to identify the variables that have large effects and therefore it can be satisfactory only to look at a fraction of the factorial experiment.

The successful use of fractional factorial designs is based on three key ideas [21]:

- The sparsity-of-effects principle. When there are several variables, the system or process is likely to be driven primarily by some of the main effects and low-order interactions.
- The projection property. Fractional factorial designs can be projected into stronger (larger) designs in the subset of significant variables.
- Sequential experimentation. It is possible to combine the runs of two (or more) fractional factorials to assemble sequentially a larger design to estimate the factor effects and interactions of interest.

On account of the sparsity-of-effects principle, in many cases we learn enough from the fractional design to proceed to the next stage of experimentation, which might involve adding or removing variables, changing responses, or varying some of the variables over new ranges. And if we have not learned enough, we can apply sequential experimentation and expand the experiment.

3.3.3 The resolution of a design

A 2^k fractional design containing 2^{k-p} runs is called a $1/2^p$ fraction of the 2^k design or, more simply, a 2^{k-p} fractional factorial design. An example of a fractional factorial design is a 2^{3-1} resolution III design (see Figure 3.7). A resolution III design means that main effects are aliased with two-factor interactions. A design is of resolution R if no p-factor effect is aliased with another effect containing less than $R-p$ factors. Designs of resolution III, IV and V are used for screening for the optimisation strategy treated in this chapter. The definitions of these designs follow below [21].

- Resolution III designs. These are designs in which no main effects are aliased with any other effect, but main effects are aliased with two-factor interactions and two-factor interactions may be aliased with each other.
- Resolution IV designs. These are designs in which no main effect is aliased with any other main effect or with any two-factor interaction, but two-factor interactions are aliased with each other.
- Resolution V designs. In these designs main effects are not aliased with each other nor with two-factor interactions, two-factor interactions are not aliased with each other either, but they are aliased with three-factor interactions.

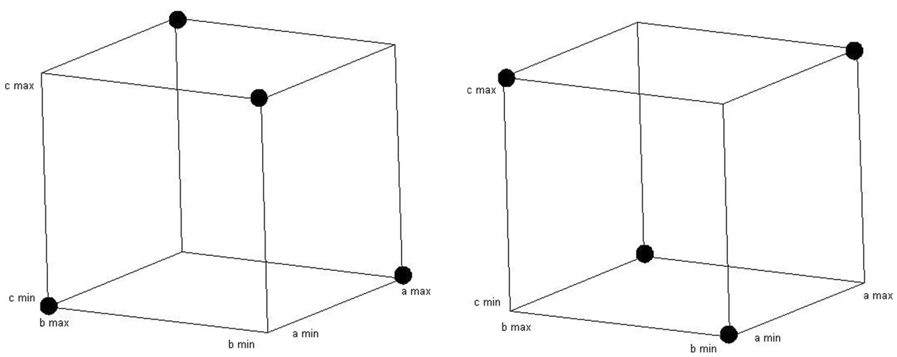
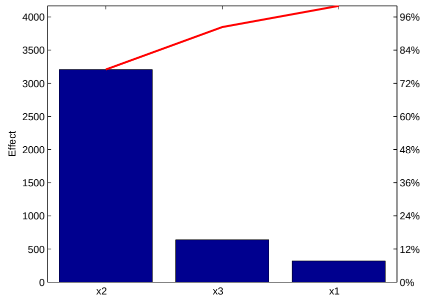
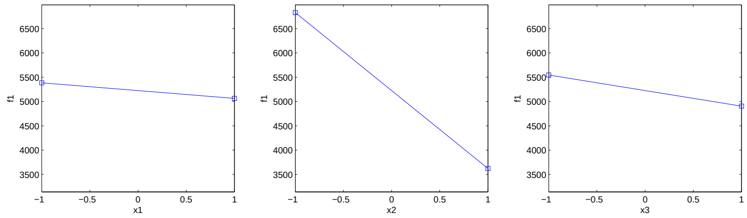


Figure 3.7: The two possibilities for a 2^{3-1} resolution III design for the variables a, b and c



(a)



(b)

Figure 3.8: Analysis of the effect of the variables on the response by (a) a Pareto plot and (b) effect plots

We usually like to employ fractional designs that have the highest possible resolution consistent with the degree of fractionation required. The higher the resolution, the less restrictive the assumptions that are required regarding which interactions are negligible in order to obtain a unique interpretation of the data. However, for a higher resolution design more calculations are required. The choice thus depends on the desired accuracy considering the calculation time.

3.3.4 Analysis with Pareto and effect plots

The responses of the the design points are analysed by applying an ANalysis Of VAriance (ANOVA), see Myers and Montgomery [21]. This statistical technique is used to determine the effects of the different variables on the responses. The figures following from the ANOVA are used for Pareto and effect plots. Examples of a Pareto and effect plots are presented in Figure 3.8. The proportional effects on the responses are displayed by the Pareto plot and the amount and direction of the effect of each variable on the response is displayed in the effect plots. In the example of Figure 3.8 the variable x_2 determines almost 80% of the effect on the response and could thus be chosen to be kept in the optimisation model whereas the variables x_1 and x_3 may be omitted.

3.4 Design Of Experiments

If the objective function, the constraints and the design variables are determined and the modelling is completed by a screening experiment to reduce the number of design variables to a sufficient number to describe the problem, then the solving of the optimisation problem can begin. This starts with a Design Of Experiments (DOE). This is a structured method for choosing a set of experimental points in the design space which gives a good reflection of the response. Since running a set of FEM calculations is a time consuming step, one would like to keep the number of experiments as low as possible. For selecting a suitable DOE strategy it is necessary to make a preliminary assumption on the type of metamodel, which is to be fitted through the response

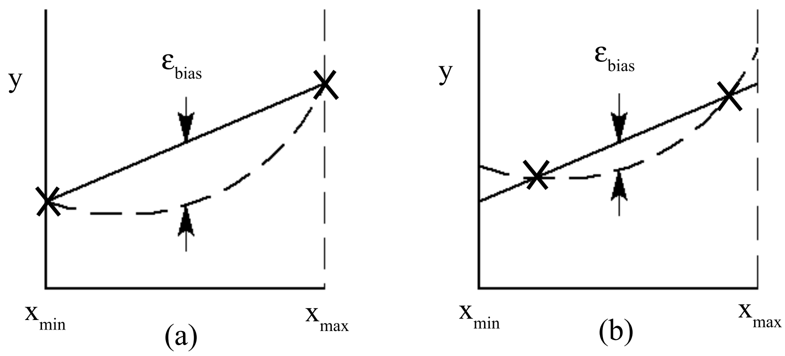


Figure 3.9: The bias error of a linear metamodel which approximates a quadratic response function constructed from (a) DOE points on the boundary of the design space and (b) from DOE points in the interior of the design space [11]

functions. The metamodels used in this optimisation strategy are based on two methods, referred to as Design and Analysis of Computer Experiments (DACE) and Response Surface Methodology (RSM). The metamodels based on DACE and RSM are further discussed in Section 3.5. In this section it is also concluded that DACE seems to be slightly more suitable for optimisation problems of metal forming processes than RSM. Therefore the applied DOE strategy is primary adjusted to a DACE based metamodel.

3.4.1 Properties of the DOE strategy

For DACE a good experimental design should:

- result in a good fit of the response;
- be cost-effective.

For a good fit of the response the design points should cleverly be chosen taking into account the shape of the metamodel. In the case of DACE the functions determining the metamodel are extremely flexible. Therefore, no assumption can be made for the final shape of the metamodel and it is important to gather information on the investigated phenomenon throughout the entire design space. Another interesting issue is the presence of a random error. This error is present in stochastic models. The metamodels based on RSM possess such an error (see Equation 3.1 in Section 3.5). The random error has an influence on the metamodel. However, for deterministic computer experiments no random error term is present. If an error is present, this is a bias error. A bias error is an error which evolves when the true response is of a different shape than the presumed metamodel. Generally, the bias error increases when DOE points are located on the boundaries of the design space. This is demonstrated in Figure 3.9. Noticing these facts gives reason to choose a DOE strategy for which the experimental design points are evenly spread over the (interior of the) entire design space [11]. Such a design is called a spacefilling design. Besides points in the interior of the design space, points on the boundary are also of interest in case of optimisation. For optimisation it is important that the metamodel gives accurate results in the neighbourhood of the optimum. Often, this optimum will be constrained and thus will be located at the boundary of the design space. Therefore, an accurate prediction is needed on the boundary, which implies performing measurements on the boundary.

The second desirable property of the DOE strategy, cost-effectiveness, is a controversial demand in case of applying DACE. Of course, it is necessary to keep the number of expensive computer simulation limited, but, unfortunately, the large flexibility of the DACE metamodels generally demands significantly more measurement data to fit a good metamodel. Despite this fact, measures may be taken to minimise the number of expensive computer simulations. On the one hand, this can be accomplished by eliminating replicate points within the design. Since computer experiments are deterministic, replicate runs which are time expensive are useless and should be prevented. On the other hand, use can be made of sequential experimentation. Applying such a sequential strategy, one starts with an initial number of simulations. Schonlau reports that specialists advise to start with 10 experiments per design variable, which demonstrates the fairly large number of measurements used for fitting a good DACE metamodel [28]. Subsequently, the metamodel is fitted and validated. If the quality of the metamodel is not satisfying, one or more design points are added, the metamodel is fitted and validated again, and so on until some quality criterion is met. This so-called sequential improvement procedure is applied in the optimisations strategy and is treated in Section 3.8.

Concluding, the DOE strategy should satisfy the following properties:

- be spacefilling;
- have design points on the boundaries of the design space;
- eliminate replicate points.

3.4.2 The DOE strategy

Bonte conducted a study to different DOE strategies and concluded that a Latin Hybercube Design (LHD) is a suitable strategy for the applied optimisation algorithm for metal forming processes [11]. Such a design excludes duplicate points and is easily made spacefilling.

Applying LHD comprises dividing the design space into a number of cubes and choosing one measurement randomly in one cube. In Figure 3.10a this is presented for a two-dimensional case; each column or row of the design comprises only one measurement. Since spacefillingness is essential the LHD is modified with the *maximin criterion*, which maximises the minimum distance between the design points. The maximin criterion ensures that no two design points are too close to each other. The combination results in a spacefilling Latin Hypercubes Design as shown in Figure 3.10b, which is a very powerful DOE strategy for DACE [26]. The maximin criterion is also applied when new design points are added by the user for improving the accuracy of the metamodels.

An LHD will generally provide design points in the interior of the design space and hardly any on the boundary. To compensate for this lack of points on the boundary, the LHD is combined with a full factorial design, which places DOE points right in the corners of the design space in a 2-dimensional case (see Figure 3.10c).

An advantage of LHD is that it provides excellent projection properties. Suppose one of two design variables appear to be not significant, then all measurements in that dimension can be

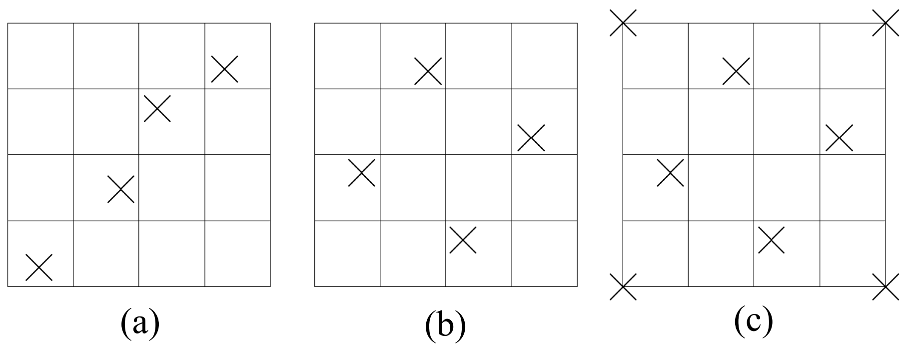


Figure 3.10: Latin Hybercube Designs: (a) a normal LHD (b) a spacefilling LHD (c) a spacefilling LHD combined with a full factorial design [11]

skipped. Using LHD, the expensive simulations that have already been preformed by then, can be projected onto the other dimension resulting in a uniform distribution in this other dimension.

3.5 Fitting the metamodel

To approximate the response functions metamodelling is applied. A metamodel approximates the response function knowing only the response values for a number of points, namely the DOE points. Two types of metamodelling are applied: Response Surface Methodology (RSM) and Design and Analysis of Computer Experiments (DACE) [7].

3.5.1 RSM

Applying RSM, natural variables are first transformed to coded variables which are dimensionless with mean zero and with the same spread or standard deviation, i.e. the errors are assumed to be uncorrelated and to have no regularities between them. Next, regression coefficients are obtained by least square regression. The response function $\mathbf{y}$ can now be expressed by the design matrix $\mathbf{X}$ containing the coded experimental design points, the regression coefficients $\boldsymbol{\beta}$ and the random error term ε according to the following equation [21]:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{3.1}$$

The metamodel approximates the response function with:

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} \tag{3.2}$$

The metamodels can have different shapes. For the optimisation algorithm four metamodels with a different shape are defined. The response is approximated by the models mentioned below. The definition of the models are given in case of two variables.

- Linear: $\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2$
- Interaction: $\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2$
- Elliptic: $\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} {x_1}^2 + \beta_{22} {x_2}^2$
- Quadratic: $\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} {x_1}^2 + \beta_{22} {x_2}^2 + \beta_{12} x_1 x_2$

The linear model is a model including only the main effects between the variables. The interaction model includes interaction between variables. These first order models are suitable for a small region where there is only little curvature in the response function. Second order models are better suitable for regions with substantial curvature in the true response functions as is the case in the neighbourhood of an optimum. An advantage of RSM is that the regression coefficients $\boldsymbol{\beta}$ are determined easily by applying the method of least squares. An example of a linear RSM based metamodel is shown in Figure 3.11 [11].

When the model is fitted, the metamodel predicted variance of any point var($\hat{y}^*$) can be determined, and also the error variance σ^2 can be estimated by the mean squared error (MSE):

$$\mathrm{MSE} = \frac{\mathrm{SSE}}{n-p} = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-p} \tag{3.3}$$

where SSE is the sum of squared errors, which equals the square of the difference between the measured response points and the response points predicted by the polynomial metamodel. n is the total number of measurements, whereas p is the number of regression coefficients. The metamodel can subsequently be visualised by plotting these predicted response values and the error variance. This is only possible for two dimensions; other variables are then fixed at a constant value or the dependence of the response on these design variables is averaged.

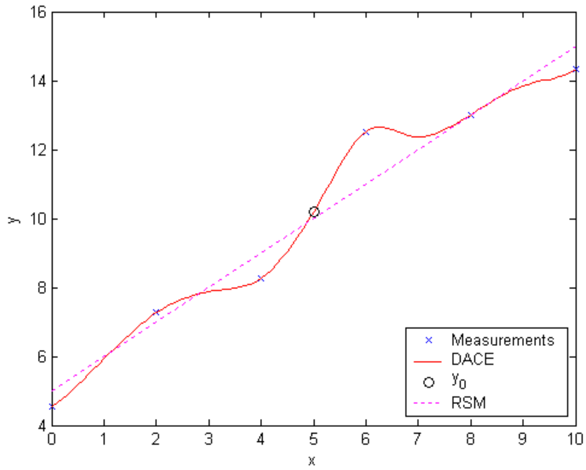


Figure 3.11: Fitting a RSM and a DACE metamodel [11]

3.5.2 DACE

A variant on the Response Surface Methodology is Design and Analysis of Computer Experiments (DACE), also called Kriging. Using Kriging, an interpolation model is fitted through all responses of the experimental points using a stochastic function Z. Whereas RSM is known to have restricted model flexibility and is primarily used for the global investigation of optima, Kriging models are known to capture local behaviour better. A Kriging metamodel is defined as:

$$\hat{\mathbf{y}} = \mathbf{X}\boldsymbol{\beta} + Z(\mathbf{x}) \tag{3.4}$$

The first part of the equation covers a global trend. The Gaussian stochastic process Z which accounts for the local deviation of the data from the linear regression metamodel has zero mean, variance σ_z^2, and a covariance which includes a Gaussian exponential correlation function $R(\theta, \mathbf{x})$. The unknown parameters β, σ_z and θ are estimated using Maximum Likelihood Estimation, which is generally considered to be the best way [26]. The influence of the magnitude of θ is visualised in Figure 3.12 [11].

For the optimisation strategy three types of Kriging models are applied. For the different types different polynomials are used as a regression model:

- Kriging with a 0^{th} order polynomial as a trend function
- Kriging with a 1^{st} order polynomial as a trend function
- Kriging with a 2^{nd} order polynomial as a trend function

The DACE based metamodel shown in Figure 3.12 for example makes use of a 0^{th} order polynomial.

After fitting the metamodel, it can predict the response value of an unknown point $y^*(\mathbf{x}^*)$ at the location defined by the design vector $\mathbf{x}^*$. Just as for RSM, the mean squared error (MSE) can be determined. The MSE at y^* is expressed by:

$$\text{MSE}\,(y^*) = \sigma_z^2 \left(1 - \begin{bmatrix} \mathbf{f}^{\mathrm{T}} & \mathbf{r}^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \mathbf{0} & \mathbf{F}^{\mathrm{T}} \\ \mathbf{F} & \mathbf{R} \end{bmatrix} \begin{bmatrix} \mathbf{f} \\ \mathbf{r} \end{bmatrix}\right) \tag{3.5}$$

This MSE equals 0 in case y^* equals a known measurement point. In case of applying DACE the predicted response values and the MSE are plotted to visualise the metamodels. Just as for RSM, this can only be done for two variables.

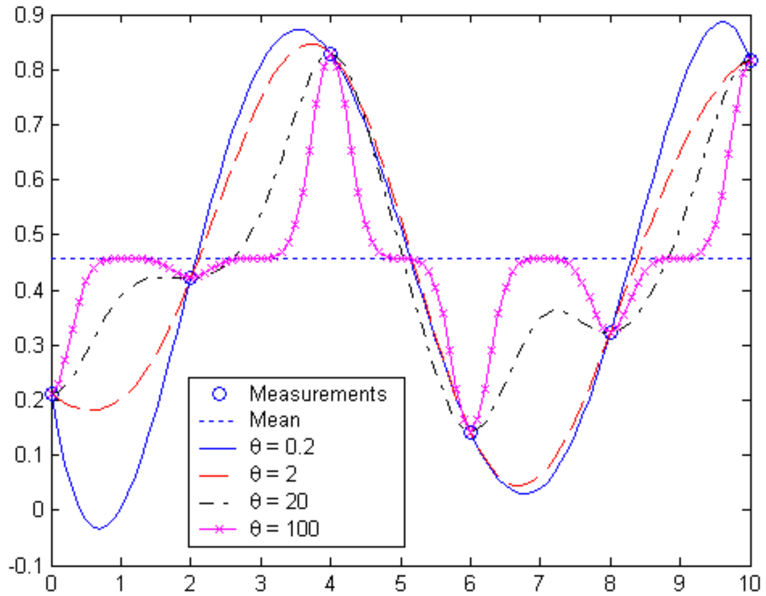


Figure 3.12: The influence of θ on the DACE metamodel [11]

3.5.3 RSM versus DACE

The different approach of RSM and DACE is shown in Figure 3.11. The limited complexity of shapes of RSM metamodels can be compensated by the flexibility of DACE (see Table 3.2). On the contrary, the sensitivity of DACE to numerical noise, which may be present, can be overcome by the smoothening behaviour of RSM. The major disadvantage of both RSM and DACE is the *curse of dimensionality* which limits the number of design variables that can be taken into account. To overcome this problem, a screening experiment as discussed in Section 3.3 is applied.

A final choice between RSM and DACE depends on the type and the shape of the response function as presented in Table 3.3. It is quite impossible to make a solid assumption on the shape of the responses. Therefore it is best to assume the worst case scenario: highly non-linear responses. In principle, the computer experiments are deterministic. Nevertheless, slightly different settings can result in significant different outcome. This can be a consequence of the physical phenomenon, in this case the assumption that computer experiments are deterministic still holds. But the differences can also be due to the adaptation of numerical parameters within the computer code (adaptive mesh refinement, automatic step size adjustment, etc). This difference could

	Advantages	**Disadvantages**
RSM	well-established transparent smooths numerical noise	up to 2^{nd} order $<$ 10 variables stochastic
DACE	flexible deterministic	$<$ 10 variables time consuming to fit sensitive to numerical noise complexity of the method

Table 3.2: Advantages and disadvantages of RSM and DACE [11]

		Response shape	
		Linear or parabolic	Highly non-linear
Response type	Stochastic	RSM	RSM/DACE
	Deterministic	DACE/RSM	DACE

Table 3.3: Criteria for using RSM or DACE

be compared to the random measurement noise assumed to be present when using RSM and is called numerical noise. The presence of noise indicates a stochastic problem which seems to be in favour of using RSM. However, it is questionable if numerical noise caused by FEM results in the zero mean random errors assumed for RSM. Hence, DACE appears to be better applicable with respect to the response type, too.

The preceding discussion gives us a small favour for applying DACE. Nonetheless, it is well possible to be dealing with particularly smooth response shapes for which numerical noise is present. In that case, using RSM would be more appropriate. Noticing this and the popularity of RSM as a metamodelling technique within the field of metal forming pleads for incorporating both RSM and DACE within the optimisation algorithm.

3.6 Metamodel validation and selection

After a metamodel is fitted it should be validated to test if the results are useful. Validation is done for two reasons:

- to test the accuracy of the metamodel;
- to test the assumptions.

In this section it will be discussed how validation is done for as well RSM as DACE on the basis of the plots which are made and the values which are calculated for this purpose. The validation plots and values are only briefly described here. For more information on the subject one is referred to the extensive elaboration of the metamodel based optimisation strategy by Bonte [11].

3.6.1 RSM

In Figure 3.13 an example of the plots made to validate the RSM model are given.

For the metamodel the sum of squares of the total variability of the measurements (SST), the variability of the metamodel or regression model (SSR) and the variability of the remaining error (SSE) can be defined:

$$\begin{aligned} \mathrm{SST} &= \sum_{i=1}^{n} (y_i - \bar{y})^2 \\ \mathrm{SSR} &= \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \\ \mathrm{SSE} &= \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \end{aligned} \tag{3.6}$$

The error SSE can be split up in a a Pure Error (PE) of the metamodel and a part that is caused by a Lack Of Fit (LOF) of the response points with respect to the real response:

$$\mathrm{SSE} = \mathrm{SSPE} + \mathrm{SSLOF} \tag{3.7}$$

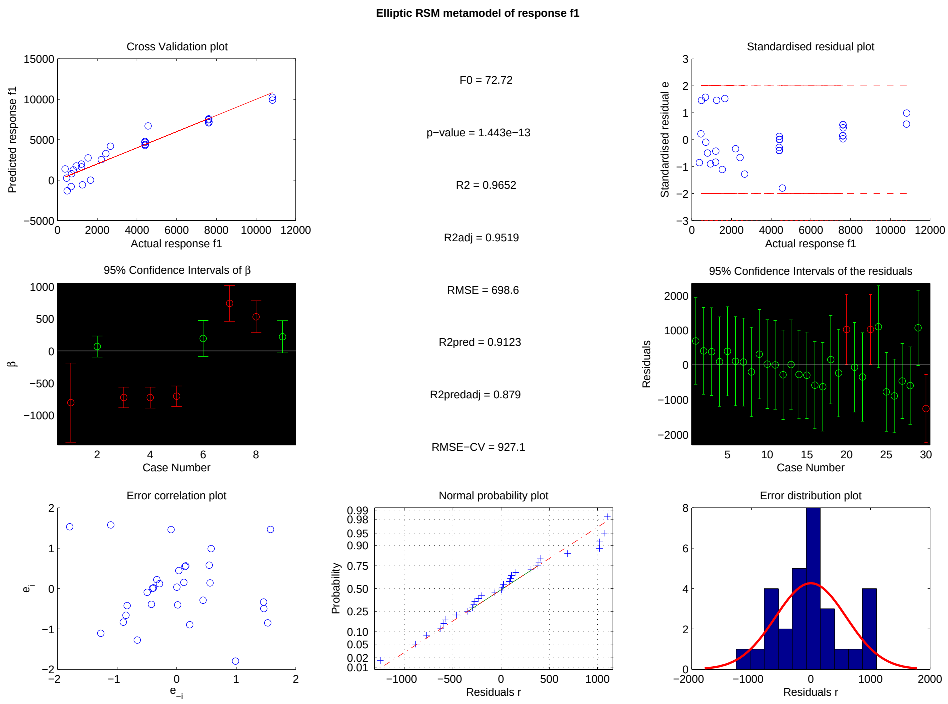


Figure 3.13: Validation plots for RSM

The R^2 value which is based on these sums of squares is one measure for how well the metamodel explains the total variability in the measurements:

$$R^2 = \frac{\text{SSR}}{\text{SST}} = 1 - \frac{\text{SSE}}{\text{SST}} \tag{3.8}$$

A larger R^2 indicates a better metamodel, whereas the metamodel is perfect when $R^2 = 1$. Since R^2 increases with every design variable, also non-significant design variables will cause an increase of the R^2 value. To investigate if a design variable is significant one can use the adjusted R^2 value. R^2_{adj} decreases when non-significant variables are added. R^2 values that approach a value of 1 indicate accurate metamodels.

The confidence intervals of the regression coefficients $\boldsymbol{\beta}$ can also provide useful validation information: if a confidence interval includes 0, the corresponding design variable is likely to be not significant with respect to the response. Hence, if all regression coefficients corresponding to the quadratic terms in an elliptic metamodel are not significant, a linear metamodel is sufficiently accurate to describe the response.

Another measure for accuracy is the Prediction Error Sum of Squares (PRESS) statistic, based on Cross Validation (CV). CV plots are also used for the validation of DACE based metamodels and are further discussed in Section 3.6.2. The PRESS statistic produces the predicted values of R^2, to know R^2_{pred} and R^2_{predadj}, which provide information on how well an unknown point is predicted. For models which fit the responses well these values are also close to 1.

To test the assumptions for RSM the residuals are analysed. To satisfy the assumptions there should be no trend in the errors and the mean should be 0. Besides, the errors should be normally distributed. The latter three plots in Figure 3.13 visualise these assumptions. The error correlation plot should reveal randomly spread errors with a mean 0. The normal probability plot and error

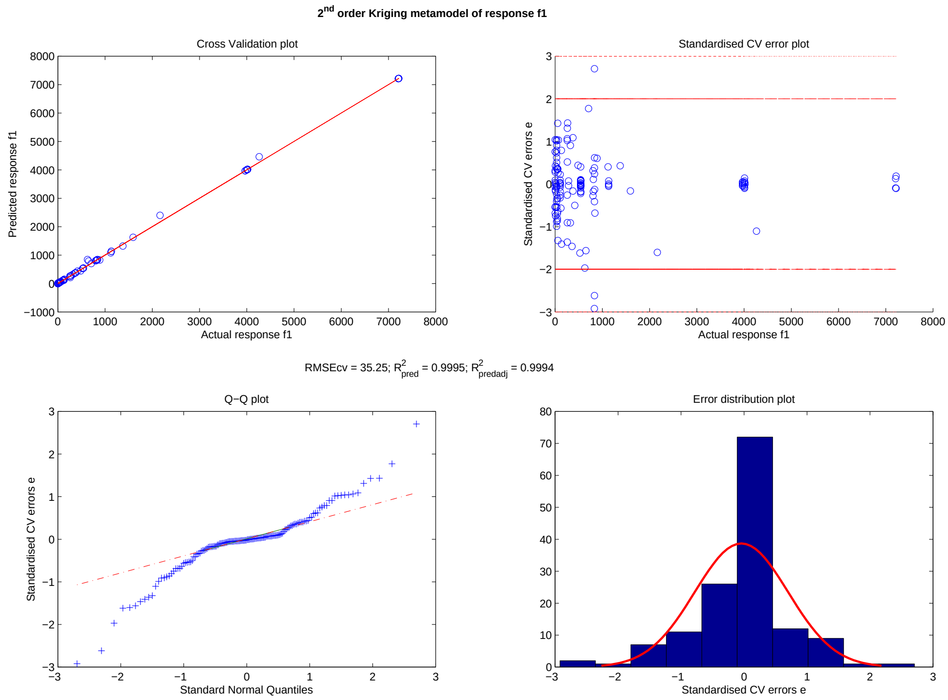


Figure 3.14: Validation plots for DACE

distribution plot should indicate a normal distribution. The latter is less reliable for only few measurements.

3.6.2 DACE

The accuracy of a DACE based metamodel is validated by Cross Validation (CV). Residual plots as for RSM are not meaningful for DACE since for interpolating the errors at the measurement points are by definition 0. Validation is done by visualising a CV plot as in the first plot of Figure 3.14. When the measurements are on a straight line this means that the model is accurate. The RMSEcv is a quantitative measure for how accurate the metamodel is. An RMSEvc of 0 indicates a perfect metamodel; in this case all measurements lie exactly on a straight line. However, if the error is not 0, a standardised CV error plot can tell us if the model is accurate: if the CV errors stay in the domain $[-2, 2]$ (or $[-3, 3]$ in case of many measurements), the metamodel is predicted accurate. Just as for RSM, it is possible to use the PRESS statistic to determine the values R^2_{pred} and R^2_{predadj}. Again, a value close to 1 indicates an accurate metamodel.

The assumptions in case of DACE which need to be tested are the following: the stochastic process Z should be normally distributed and have a zero mean. These assumptions can be validated by investigating if the standardised CV error plot reveals any trend and if it indicated a zero mean. The Q-Q plot and error distribution plot in Figure 3.14 should indicate normality.

3.6.3 Selection best metamodel

The validation plots are helpful tools for determining the best metamodel for each response. For the optimisation strategy the approach is to choose the best of the seven metamodels, even if the metamodel is not accurate, and optimise this metamodel. If the approximated optimum is not accurate enough, additional DOE points are selected and a more accurate metamodel can be obtained.

3.7 Metamodel optimisation

The next step is to optimise the metamodel. Optimising the mathematical metamodel is much more efficient than optimising the FE model directly. Any optimisation algorithm can be used for this depending on the shape of the objective function and possible constraints.

We assume non-linear responses and (thus) complex metamodels. Complex metamodels imply local optima and therefore it is preferable to use a global optimisation algorithm. Global optimisation algorithms try to find the absolute optimum and generally search a larger domain than local optimisation methods. This makes them slower, which may form a problem, especially for large optimisation problems with many variables. Nonetheless, a global approximation optimisation strategy is chosen to exclude the risk to result in an erroneously predicted global optimum which is actually a local optimum. Besides, compared to a FEM analysis such an optimisation strategy is still quite fast.

An iterative Sequential Quadratic Programming (SQP) algorithm is chosen, which is available in MATLAB's Optimization Toolbox [30]. Since iterative algorithms also may get stuck in local optima, a multi-start approach is implemented which means that the algorithm starts at more than one location. In the case of the UT-algorithm the optimisation will start in all DOE points.

The optimisation of the metamodel results in a set of optimal design variables. These settings can subsequently be used as an input for a final FEM calculation to evaluate if the optimum approximated using the metamodel is close enough to the real value of the objective function obtained by the FE model. If this is not the case, sequential improvement may be applied.

3.8 Sequential improvement

Depending on the shape of the unknown objective function, the approximation obtained by the metamodel may be quite coarse. To improve the approximation it is possible to perform another approximation in the neighbourhood of the optimum obtained after the first approximation. This process can be done over and over again until a well predicted optimum is found. This sequential improvement of the optimum is done until the metamodel is accurate enough and the approximate optimum is close enough to the real value of the objective function obtained by running a FEM calculation with the optimised design variable settings.

In [11] different sequential improvement strategies are considered. It is recommended to use either the method of Minimising the Merit Function (MMF) or the method of Maximum Expected Improvement (MEI). Both methods focus on finding the global optimum as soon as possible rather than obtaining accurate metamodels. In [11] some conclusions and rules of thumb on the MEI en MMF strategies were obtained:

- Fit an initial metamodel from about 10 times the number of design variables as initial points.
- When using MMF set the value of the weight factor to 1.

MEI and MMF are similar strategies. An advantage of MEI over MMF is that no choice needs to be made for the weight factor. A disadvantage is, however, that if metamodels are not indicating the right region of the global optimum, MEI is known to look exhaustively near the erroneously predicted global optimum before it finally decides to look at other locations. For the application described in this report a choice has been made for MMF.

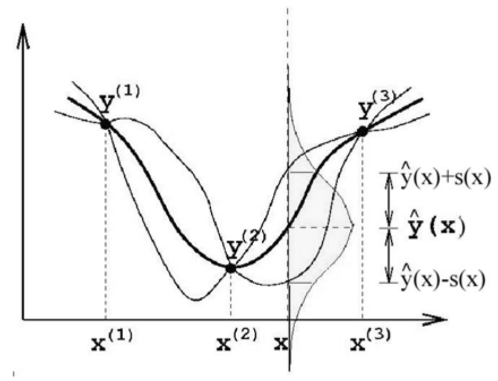


Figure 3.15: Normal distribution around an unknown point [11]

3.8.1 Minimising the Merit Function (MMF)

The method of Minimising the Merit Function (MMF) selects new DOE points in the neighbourhood of the global optimum or on locations where a high root mean squared error (RMSE) value is predicted by minimising a so-called merit function. The RMSE equals the standard deviation of the metamodel as depicted in Figure 3.15. The merit function is:

$$f_{min} = \hat{y} - w\text{RMSE}(\hat{y}) \tag{3.9}$$

where $\hat{y}$ is the metamodel and w is a weight function that can be used to tune the relative importance between the objective function values and the predicted errors. A value of $w = 0$ will only take into account the current global optimum of the objective function, whereas an increasing value of w will widen the new design space more and more.

The method is presented in Figure 3.16. After having obtained an initial metamodel from the calculations that resulted in the crosses, the merit function shown as the dotted line is minimised. This results in a number of DOE points shown as indicated with the circles. However, to avoid wasting expensive calculation time for locating local optima, only those DOE points are selected that are predicted to be underneath the level $\hat{y}^* + w\text{RMSE}(\hat{y}^*)$ as shown in Figure 3.16, in which $\hat{y}^*$ denoted the global optimum predicted by the metamodel of the objective function.

Now, the procedure applied to the UT-algorithm is as follows [11]:

1. Start with a DOE of approximately 10 times the number of design variables. Run the FEM simulations
2. Fit the seven metamodels
3. Optimise the best metamodels using the multistart SQP algorithm
4. Minimise the merit function using the multistart SQP algorithm
5. Select only the minima of the merit function that satisfy $\hat{y} + w\text{RMSE}(\hat{y}) \leq \hat{y}^* + w\text{RMSE}(\hat{y}^*)$ as the new DOE points

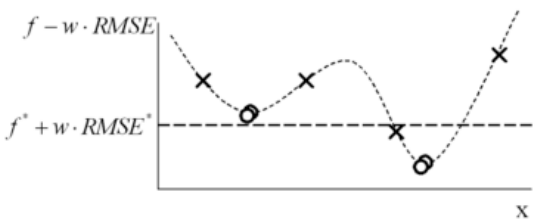


Figure 3.16: Metamodel improvement by Minimising a Merit Function [11]

6. Run the calculations and return to step 2

3.8.2 Applying MMF to the optimisation process of the bearing tube

In case of the optimisation for the application described in this report, the deep drawing process of the bearing tube, the sequential improvement strategy MMF has been chosen. However, an adjustment has been made.

For the MMF procedure as described in Section 3.8.1 the level which is displayed in Figure 3.16 is defined by $f_{\text{min}} = \hat{y}^* + w\text{RMSE}(\hat{y}^*)$. The definition of this level has been adjusted to make the procedure less dependent on the RMSE of the metamodel. The level f_{min} is now determined by the minimal feasible objective function value that is obtained at that moment. This is done by replacing step 5 of the MMF procedure by the following steps:

- First, it is determined which values of the objective function f are feasible. These values are no approximations, but real values obtained by FEM calculations. The f values that belong to a constraint value larger than 0 ($g \geq 0$) are infeasible values and are eliminated
- Now, f_{min} is determined by the minimum value of the feasible objective function values: $f_{\text{min}} = \min \mathbf{f}_{\text{feasible}}$
- Select only the minima of the merit function that satisfy $\hat{y} + w\text{RMSE}(\hat{y}) \leq f_{\text{min}}$ as the new DOE points

3.9 Conclusions

This chapter presented an overview of the optimisation strategy which can be applied to metal forming processes using time consuming FEM calculations.

The strategy contains a guideline how to model the mathematical optimisation model which consists of an objective function, implicit constraints and design variables. When this optimisation problem is complete, a screening methodology can be applied to reduce the number of design variables. For screening a few FEM calculations are run to determine the effect of the different variables on the responses. Pareto and effect plots support the choice to exclude variables with a small influence on the responses. After the reduced optimisation model is defined, it can be solved by applying an optimisation algorithm.

The optimisation algorithm selected for the optimisation strategy is a sequential approximate optimisation algorithm. This algorithm uses metamodelling techniques to approximate the responses. It starts with a Design Of Experiments for cleverly choosing experimental points for which FEM simulations need to be run to determine the response values. The applied DOE strategy is a spacefilling Latin Hybercube Design (LHD) combined with a full factorial design. Next, two types of metamodelling techniques are applied to the response values: Response Surface Methodology (RSM) and Design and Analysis of Computer Experiments (DACE). The RSM and DACE metamodels are subsequently validated by constructing a number of validation plots. With help of these validation plots the best metamodel can be selected and this metamodel is optimised by applying a Sequential Quadratic Programming (SQP) algorithm. This results in an approximate optimum. This approximate optimum may be validated by running a final FEM calculation using the optimised design variable settings. When the solution is not accurate enough more experimental points need to be added and the process starts over again until an optimum is found which is accurate enough. This process is called sequential improvement. For the optimisation strategy two sequential improvement strategies are recommended: the method of Minimising the Merit Function (MMF) or the method of Maximum Expected Improvement (MEI). For the application described in this report MMF is applied. The algorithm is implemented in MATLAB; this software is called OptformV2.0.

This strategy, which is developed at the University of Twente, is applied to the optimisation problem of the bearing tube in Chapters 4 and 5.

Chapter 4

Optimisation of the reference process of the bearing tube

The optimisation strategy which has been introduced in Chapter 3 is applied to the production process of the bearing tube with as primary goal to prevent the cracks from occurring. The first step in mathematical optimisation is creating the optimisation model. This is described in the first section of this chapter. In Section 4.2 it is observed that the number of variables is large for optimisation and therefore the variables are screened to reduce the number of variables. The final optimisation model is now defined and can be optimised. The results of the optimisation a presented in Section 4.3. In the last section of this chapter the findings are summarised.

4.1 Modelling

The modelling of the optimisation problem includes selecting the design variables, determining the optimisation objective and determining the set of constraints which define the process boundaries. In this section the structured procedure which is described in Section 3.2 to identify these matters is elaborated for the bearing tube. It comprises the following steps:

1. Determine the appropriate optimisation situation
2. Select only the necessary responses
3. Select one response as objective function, the others as implicit constraints
4. Quantify the objective function and implicit constraints
5. Select possible design variables
6. Define the ranges on the design variables
7. Identify explicit constraints

Step 1: Define the appropriate optimisation phase

The bearing tube is already in production which means that the appropriate Product Development phase is Production. For the application of optimisation to solve production problems, three stages are distinguished. In Figure 4.1 these stages are presented.

The first stage is problem identification. The production problem can be Design Parameter related or Functional Requirement related. If the problem is Design Parameter related, this means that the production process does not yield the product the part designer intended it to have. In case the problem is Functional Requirement related, the production process does not yield products that are able to satisfy the Functional Requirements. Defects like cracks, wrinkles and necks are deficiencies with respect to the intended part geometry. The production problem is such a defect - cracks occur - and hence the problem is Design Parameter related.

If the problem is identified, optimisation techniques can be applied to solve the production problem. This is done in one of the next two stages of Figure 4.1. To solve the problem the part can be redesigned or the process can be redesigned without changing the final geometry of the part. In case of the bearing tube, other parts in the A-arm depend on the geometry of the bearing tube.

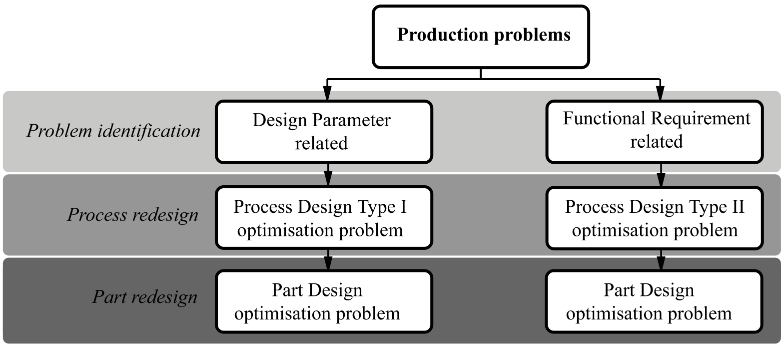


Figure 4.1: Application of optimisation to solve production problems

Applying part redesign implies that these parts also should be altered when the geometry of the bearing tube is changed. Therefore, the preference of VW is to keep the same geometry for the part, hence a first attempt would be to apply optimisation techniques to redesign the process. If after process redesign the problem persists, the final possibility to get rid of the problem can be to redesign the part. From Figure 4.1 it can be obtained that we now speak of a Process Design Type I optimisation problem.

The input-response model for FEM in a Process Design Type 1 situation is presented in Figure 4.2a. Comparing this figure with Figure 4.2b which was introduced in Section 3.1 makes it is easy to see which quantities to select as objective function, which as constraints and which design variables to take into account. Possible responses, the objective function and implicit constraints, are the Design Parameters, Defects and Costs. Possible design variables are the Process Variables.

Step 2: Select only the necessary responses

Design Parameters, Defects and Costs are possible responses. In this case we are interested in preventing the cracks from occurring. Also, other defects should be prevented. The cracks are responsible for a costly production process of the bearing tube, because a 100% crack control is applied. If the cracks are prevented from occurring there is no need for this control and this will reduce the costs. Other costs are of less importance, thus include Costs as response is of less interest in this optimisation problem. Design Parameters are of less interest too, because the geometry and material of the product are no critical issues.

In Chapter 2 it was observed that the cracking in the critical areas occurs in or after the last

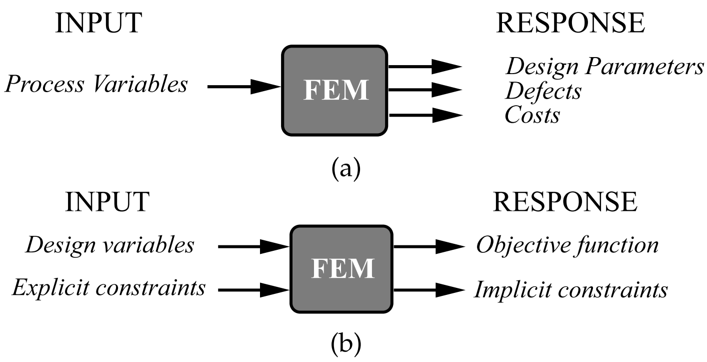


Figure 4.2: Input-response model for FEM (a) in a Process Type 1 situation and (b) in relation to optimisation

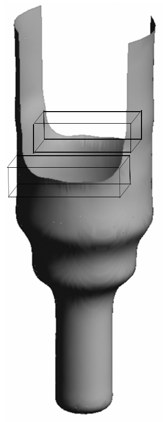


Figure 4.3: The definition of two critical areas

forming step when the flange is drawn in the direction of the bolt. It was concluded that large deformations and work hardening in the critical areas are responsible for the cracks. Hence cracking is quantified by the major and minor strains in the critical areas. The critical areas are shown in Figure 4.3.

Furthermore, the formability of the product should be guaranteed. This is done by preventing defects related to the formability of the product. A formability issue that should be prevented is necking. Necking is also related to the strains. Necking can be quantified by the distance of the strains to the Forming Limit Curve (FLC). Because of the multi-step deep drawing process it is not right to define the FLC at the end of the process by the same curve as the one that can be applied after the first deep drawing step. The FLC depends on the strain paths (the prestrains alter) and therefore changes during the deep drawing process (see Section 2.1.2). A non-linear strain path of one of the elements after the final deep drawing step in shown in Figure 4.4. For this element the FLC as defined by AutoForm cannot be applied. The strains as presented in this figure, are the strains after the last deep drawing step for the reference process. The reference process has proven to deliver a feasible product, thus indeed necking is not problem for this process although there are elements exceeding the initial FLC. An adjusted FLC is now defined as presented in Figure 4.4; this FLC is defined by the FLC which can be applied after the first forming step as defined by AutoForm for the DD13 material, but it also allows strains that emerge in a later stadium to exceed this FLC. The dashed line presents the new limit for elements with a minor strain between a value of -0.26 and 0.15. This line is based on the strains of the elements that exceed the FLC in the reference process but do not cause necking.

Other responses which may be considered are responses related to other defects, like wrinkling, earing and Cold Work Embrittlement (CWE). However, as concluded in Section 2.3.3, we expect CWE not to be problem. Earing is not present for the existing production process of the bearing tube. And, because the material stays the same and the process will not be subject to large changes, it is assumed that earing will not occur. The AutoForm simulation of the reference process does show some wrinkling in the middle of the product (in Figure 4.10 of Intermezzo 1 an example of a location of a wrinkle is indicated). However, in reality wrinkling is not a problem. It is assumed that wrinkling will also not become a problem and is excluded from the optimisation model.

Summarising, two responses are selected:

1. Major and minor strains to quantify cracking
2. The distance of the strains to the FLC to quantify necking

Type of response	No nodal/element value	Nodal/element value, critical	Nodal/element value, non-critical
Objective function, no target	$\min X$	$\min \max_N X$	$\min \frac{\sum_N(X)}{N}$
Objective function, target = X_0	$\min \lvert X - X_0 \rvert$	$\min \max_N \lvert X - X_0 \rvert$	$\min \frac{\lVert X-X_0 \rVert_2}{\sqrt{N}}$
Implicit constraint, USL	$X - USL \leq 0$	$\max_N(X - USL) \leq 0$	
Implicit constraint, LSL	$LSL - X \leq 0$	$\max_N(LSL - X) \leq 0$	

Table 4.1: Response quantification [10]

Step 3: Select objective functions and implicit constraints

The objective of optimisation is preventing the cracks from occurring, hence optimising the strains is selected for the objective function. Keeping the strains below the FLC to prevent necking is selected as an implicit constraint.

Step 4: Compose the final objective function and constraints

In this step mathematical formulations of both responses are formulated. Table 4.1 assists in selecting the formulations for the responses. In this case as well the objective function as the constraint are based on element values (strains). The large deformations in the critical areas correspond to a large major strain ε_1 and a small minor strain ε_2. Therefore the objective function is focussed on minimising the major strains while maximising the minor strains in the critical areas. The objective function is thus a critical value. The constraint is also, as always, a critical value.

The objective function is defined by the distance d_f of the strains to the origin in the Forming Limit Diagram (FLD) as shown in Figure 4.4. This distance is minimised. Minimising the objective function corresponds to the usage of objective functions in the software OptformV2.0.

In case of the constraint, it is not allowed that the strains of any element exceed the FLC. Therefore, the constraint is defined by the distance d_g of the strains of any element to the Forming Limit Curve (FLC) in the FLD as displayed in Figure 4.4. Because the constraint is a critical value, a positive or negative value of g needs to depend on the maximum strains: if the strains of one of the elements exceed the FLC g needs to return a positive value (in accordance with the software OptformV2.0).

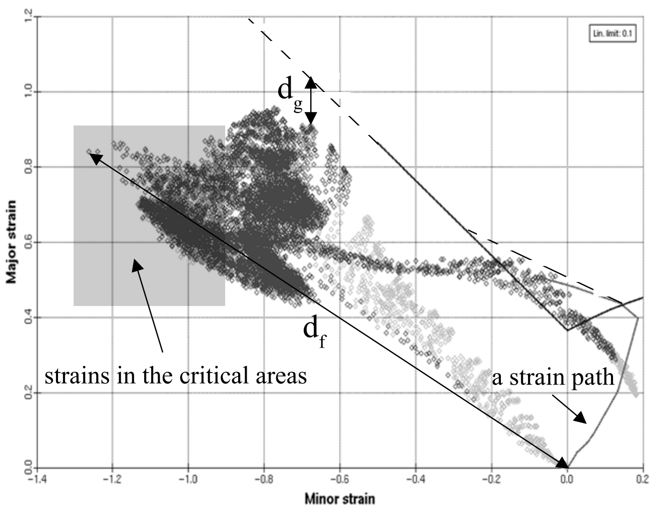


Figure 4.4: Definition of the objective function and the constraint

The mathematical definitions of the objective function and the constraint are now:

$$\begin{aligned} \min f(\varepsilon_1, \varepsilon_2) &= \min \max_{N_\text{crit}} \left(d_f(\varepsilon_1, \varepsilon_2)\right) = \min \max_{N_\text{crit}} \left(\sqrt{\varepsilon_1^2 + \varepsilon_2^2}\right) \\ g_\text{impl} &= \max_{N} \left(d_g(\varepsilon_1, \varepsilon_2)\right) = \max_{N} \left(\varepsilon_1(\varepsilon_2) - \varepsilon_1^\text{FLC}(\varepsilon_2)\right) \leq 0 \end{aligned} \tag{4.1}$$

where N_crit denotes the elements in the critical areas and N denotes all elements. ε_1 and ε_2 are the strains belonging to an element part of N_crit or N. ε_1^FLC is the major strain belonging to the FLC for ε_2 of the element.

The formulations of the responses in Equation 4.1 were subject to some changes before the final formulations were defined. This was done after a first optimisation. This first optimisation is discussed in Intermezzo 1. In this intermezzo the changes to the formulation of the responses are extensively described.

Intermezzo 1: Improvements to the optimisation model

A first optimisation of the responses in Equation 4.1 delivered unsatisfying results and some adjustments were made to the optimisation model to improve the optimisation results. This first optimisation and the adjustments which led to a final optimisation model are treated in this intermezzo. First, these first optimisation results are discussed. It appeared that the results did not converge to an optimum. Therefore, in the second part of this intermezzo the results are further analysed to identify the possible causes for this. Improving the optimisation model will result in better responses and therefore improve the optimisation results. The third part of this intermezzo is therefore focussed on improving the modelling of the optimisation problem. The intermezzo finalises with the results of the adjustments.

First optimisation results

For a first optimisation the responses are defined by Equation 4.1. Ten variables are chosen to be included in the optimisation model and upper and lower bounds have been defined for these variables (see later on in Section 4.1). In Table 4.2 these variables are presented with the values of their original settings of the reference process as introduced in Section 2.2. The optimisation model is screened to reduce the number of variables to take into account for optimisation. The Pareto plots after screening are shown in Figure 4.5. The lines indicate that almost 75% of the objective function value is determined by the variables x_{02}, x_{01}, x_{04} and x_{07} and almost 70% of the value of the constraint is determined by the variables x_{01}, x_{06} and x_{09}. These variables are chosen for a reduced optimisation model. Because the process already exists, the other four variables are set to their original values to minimise the number of adjustments to the process. The reduced optimisation model defined by only six variables now becomes:

$$\begin{aligned} \min f(\varepsilon_1, \varepsilon_2) &= \min \max_{N_\text{crit}} \left(\sqrt{\varepsilon_1^2 + \varepsilon_2^2}\right) \\ \text{s.t.} \quad g_\text{impl} &= \max_{N} \left(\varepsilon_1(\varepsilon_2) - \varepsilon_1^\text{FLC}(\varepsilon_2)\right) \leq 0 \\ 200 \text{ mm} &\leq \text{size} \leq 215 \text{ mm} \\ 2.8 \text{ mm} &\leq t(x_{02}) \leq 3.2 \text{ mm} \\ 10 \text{ mm} &\leq \rho_\text{punch02} \leq 15 \text{ mm} \\ 6 \text{ mm} &\leq \rho_\text{die04} \leq 10 \text{ mm} \\ 9 \text{ mm} &\leq \rho_\text{punch04} \leq 13 \text{ mm} \\ 0.08 &\leq \mu_\text{die/binder} \leq 0.15 \end{aligned} \tag{4.2}$$

x_{01}	size	204 mm	x_{06}	ρ_{die04}	8 mm
x_{02}	t	3.0 mm	x_{07}	ρ_{punch04}	11 mm
x_{03}	ρ_{die02}	9 mm	x_{08}	ρ_{die18}	10 mm
x_{04}	ρ_{punch02}	12.5 mm	x_{09}	$\mu_{\text{die/binder}}$	0.10
x_{05}	$F_{\text{blankholder02}}$	100000 N	x_{10}	μ_{punch}	0.12

Table 4.2: The ten variables for optimisation of the reference process with their original settings

To this reduced optimisation model a first optimisation has been applied by executing the optimisation algorithm that is treated in Chapter 3. The result after 169 FEM calculations was a reduced objective function value of 1.3618 (the objective function value of the reference process is 1.655). The plot of the objective function belonging to these results is displayed in Figure 4.6. The dot presents the lowest found objective function value that is feasible and has a constraint value lower than 0 (thus the strains of all elements stay below the FLC). The plot shows that after more calculations the algorithm finds more lower objective function values. However the thirteenth simulation is already a feasible one with a value of 1.3814 and thus, the improvement after many calculations is little. The objective function value does not show real convergence behaviour and also, metamodels are far from accurate. On the basis of these first results, it was decided to study the response in more detail by plotting this for only one variable.

Observation responses

The screening results indicated that the variables size (x_{01}) and thickness (x_{02}) have the most influence on the objective function value. Therefore the value for the objective function and the constraint were plotted for these two variables, while varying these variables within their variable range in 100 steps (the other variables were set to their original values for these calculations).

In the plots of Figure 4.7a it can be seen that there is not a clear trend for the objective function value when varying the size. For the constraint though, a trend is found. Figure 4.7b shows that varying only the thickness obtains a very noisy response for as well the objective function as for the constraint. In general, it is observed from these plots that the responses suffer from noise and unclear trends. A noisy response does not have to be a problem when there is a clear trend, however it is also possible that an existing trend line may not be discovered due to too much noise. Thus, goal is to reduce the noise and/or find a trend to improve the responses.

To be able to reduce noise or reveal a trend, we should know what influences these matters. The numerical noise is caused by the FEM simulation programme, however the formulation of the optimisation problem has effect on the influence of this noise on the responses. How the optimisation problem can influence the noise or the trend, is presented in Table 4.3. Noise and trend are further discussed.

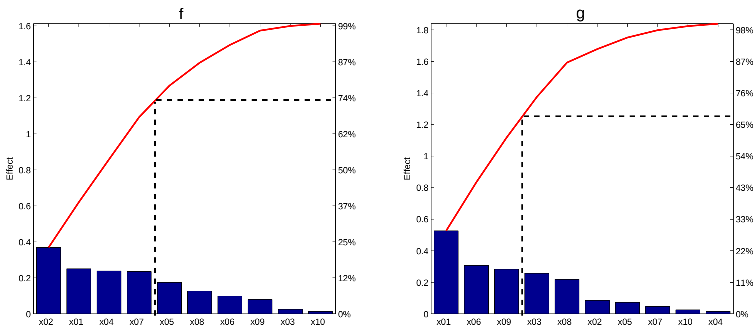


Figure 4.5: Pareto plots of the objective function and the constraint

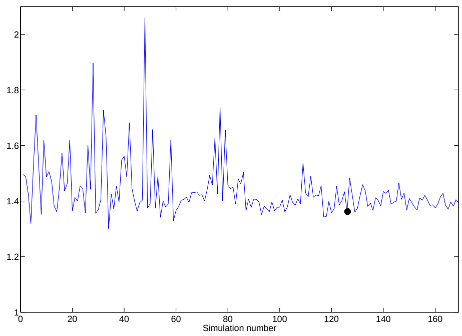


Figure 4.6: Plot of the objective function after running 169 simulations

Noise

The numerical noise is caused by the FEM programme and influenced by the definition of the optimisation model. The influence of numerical noise on the objective function value is quantified by looking at the values of the distance d_f for the elements surrounding the element belonging to the maximum distance $d_f^{\max}$. The average value of the differences between these distances is calculated by:

$$\overline{\Delta d_f} = \frac{\sum_{i=1}^{n} \left(d_f^{\max} - d_f^i\right)}{n} \tag{4.3}$$

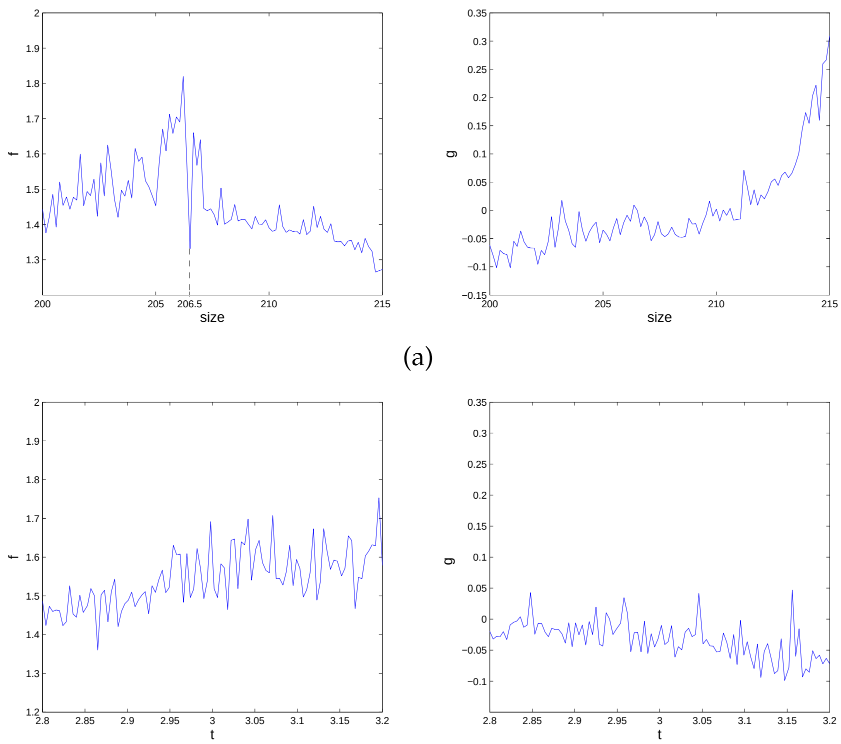


Figure 4.7: Plots of the objective function and the constraint varying (a) only the size or (b) only the thickness

Noise	**Trend**
Influence of the definition of the objective function	
Because the objective function is defined by a maximum value, this value is based on only one element. This influences the noise.	The objective function is defined in such a way that elements with the same distance to the origin are all equally judged regardless of the relationship between the major and minor strain. This relationship may be of importance.
Influence of the definition of the critical areas	
Due to a bad definition of the critical areas, a (small) shift of the location of the critical areas with respect to the product takes place in every calculation. Therefore, it can happen that for the calculation of the objective function elements on a certain location in the product are included in the critical areas for one FEM calculation, and excluded for another calculation.	It can be that large strains exist on different locations in the product due to different causes. The variation of the large strains on these locations can have different trends. The large critical areas can include elements that have large strains for different reasons. If this is the case, this will cause that different trends have influence on the response making every separate trend less clear or even invisible.

Table 4.3: Influences of the definition of the optimisation problem on the noise and the trend of the responses

n is the number of elements surrounding the element that belongs to the distance $d_f^{\max}$. An example of these n elements is visualised in Figure 4.8; the black element presents the element that belongs to $d_f^{\max}$. For five simulations the average value $\overline{\Delta d_f}$ is calculated. Four of these simulations were randomly chosen from the simulations that were executed to observe the responses. A fifth value was calculated for a more accurate simulation (simulation 5). The differences in the settings for this simulation are a smaller radius penetration and a smaller maximum element angle to define the mesh and a smaller maximum displacement to define the time steps. Table 4.4 presents the values for $\overline{\Delta d_f}$ as well as the values of the variables t and size for these five simulations. The value for simulation 5 shows that the more accurate simulation did not reduce numerical noise. Comparing the size of the values of $\overline{\Delta d_f}$ with the size of the variation in f defined by d_f (see Figure 4.7) shows that the noise caused by the simulation programme is a large part of the total variation of the objective function.

	t	size	$\overline{\Delta d_f}$
simulation 1	3.00 mm	204.0 mm	0.065
simulation 2	2.85 mm	204.0 mm	0.106
simulation 3	3.00 mm	207.5 mm	0.103
simulation 4	3.00 mm	210.0 mm	0.076
simulation 5 (acc)	3.00 mm	210.0 mm	0.101

Table 4.4: Quantification of numerical noise

As presented in Table 4.3 the definition of the optimisation problem influences the noise in two ways:

1. By the definition of the critical areas. This can cause that elements are compared to elements on another location which have strains of another magnitude. For example, the low value of f in the plot for the objective function in Figure 4.7a (for size = 206.5 mm) is caused by this. In this simulation the critical areas were shifted and only elements in the flanges were taken into account. A better definition of the critical areas can overcome this influence.
2. By the fact that the definitions of the responses are based on the strains of only one element. The relative difference between two elements can be larger than the differences between a group of elements. Therefore trying other response definitions based on more

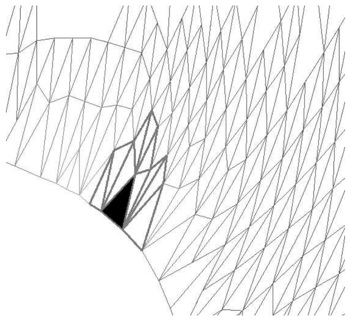


Figure 4.8: The elements surrounding the element belonging to the maximum objective function

elements can be a useful attempt to reduce the noise. This can be done by applying different norms.

Trend

A bad definition of the objective function or the constraint can cause that a trend which in theory is present, does not occur in the response. As presented in Table 4.3 the definition of the objective function and the critical areas influence the trend of the response.
Altering the definition of the objective function may result in a response with possibly a trend. Nevertheless, this is to be done carefully as there is need to keep the real objective in mind and the goal should not be to change the objective function in such a way that a trend is discovered. Because then, we might be optimising a problem which is of no interest. As noted before, the objective function for this optimisation problem is defined in such a way that elements with the same distance to the origin are all equally judged regardless of the relationship between the major and minor strain. In Figure 4.9 this is visualised; the elements on the circle all have an objective function value of 1.38. However, the relationship between ε_1 and ε_2 (thickness) may be of importance.
Furthermore, in case of the bearing tube, the large critical areas include elements in the middle of the product as well as elements in the corners where the cracks occur. It is observed that the elements in the middle of the product are subject to wrinkling and therefore also have large strains. Figure 4.10 presents the large strains on both locations. This causes that the objective function sometimes depends on large strains in the middle of the product and sometimes on large strains in the corners where the cracks occur. However, we are only interested in the large strains in the corners. A better definition of the critical areas should focus on only these elements and exclude the influence of wrinkling in the middle of the product on the objective

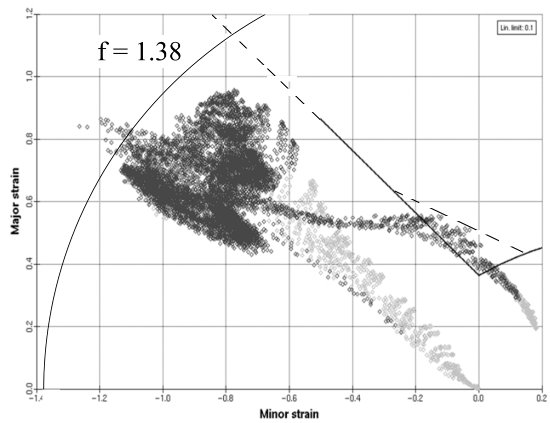


Figure 4.9: The circle definition of the objective function

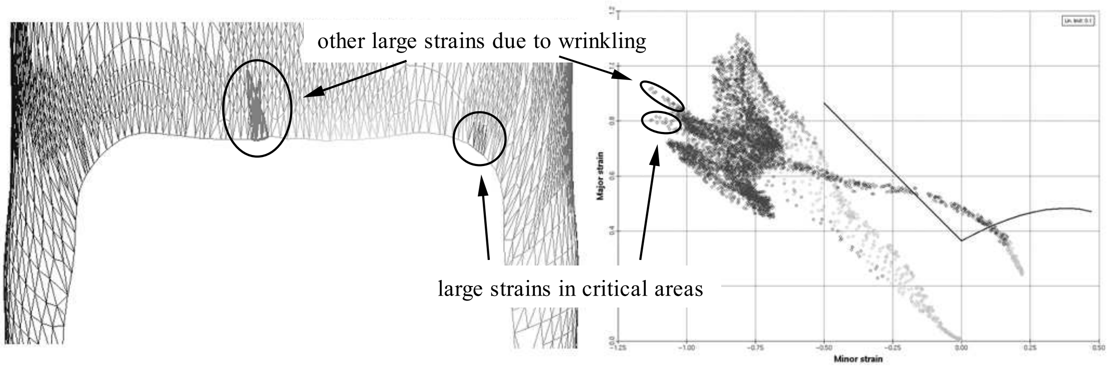


Figure 4.10: Large strains on different locations in the product

function.

Adjustments

Possible influences on noise and trend have been discussed. It was observed that a better definition of the critical areas can improve the response of the objective function by excluding elements in the middle of the product and by taking care that the same elements are included for every FEM calculation. Another improvement can be a better definition of the responses by applying norms and by a different definition of the objective function. These two adjustments, the definition of the critical areas and the definition of the responses, are applied and discussed in this section.

Critical areas

The areas have been adjusted in such a way that only the elements in the four corners where the cracks occur, are taken into account. Now, we exclude the influence of wrinkling in the middle of the product on the objective function.

Besides that, the location of the product differs in the different simulations (sometimes, it rotates a little). Therefore, the critical areas are better defined by using a local coordinate system for every FEM calculation. The location of the critical areas now depends on the location of the product. This results in more consistent critical areas in the separate simulations. The new critical areas are defined by the boxes as presented in Figure 4.11.

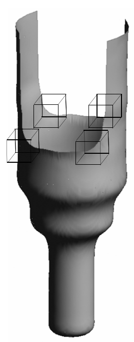


Figure 4.11: The definition of the four critical areas

Definition responses

It is tried to smoothen the responses by applying a p-norm as defined in Equation 4.4.

$$||x||_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}} \tag{4.4}$$

The size of the value of $||x||_p$ depends on the size of n. To make the norm-value independent of n the norm should be divided by $\sqrt[p]{n}$:

$$\overline{x_p} = \frac{||x||_p}{\sqrt[p]{n}} \tag{4.5}$$

When a p-norm is applied with $p = 1$ and this value is divided by n, this equals the average value of x_i. For $p \to \infty$ this norm is also called the maximum norm and equals the maximum value of x_i.
For the objective function Equation 4.5 is applied with $p = 2$. In this case, n stands for the number of elements and x is defined by the strains of these elements. This norm averages the values of x, but larger values have more effect on this average. In this way, on the one hand the influence of numerical noise on the response is reduced because the value not solely depends on one element, but on the other hand the objective is still focussed minimising the largest strains.
For the constraint it is of primary interest that the strains of none of the elements exceed the Forming Limit Curve (FLC). Therefore, a positive or negative value of g needs to depend on the maximum strains: if the strains of one of the elements exceed the FLC g needs to return a positive value (in accordance with the software OptformV2.0). If this is ensured, then a norm can also be applied to the definition of g to try to smoothen the response. If all the strains stay below the FLC, then the definition of g is based on all the elements. Because only the elements with the largest strains indicate if the strains are close to the critical value (the FLC) a norm with $p = 20$ is applied. The value that this norm returns is much more influenced by the maximum value than by the minimum value. If only some strains exceed the FLC, then the definition of g is based on these elements that have strains that exceed the FLC. Because the strains belonging to these elements are all large, a norm with $p = 2$ is applied.
Besides applying norms to the definition of the responses, it is also tried to improve the response of the objective function by adjusting the definition of the objective function. As noted before, the objective function was defined in such a way that elements with the same distance to the origin are all equally judged. However, the relationship between the major and minor strain may also be of importance. It was analysed by INPRO that thickening has a negative

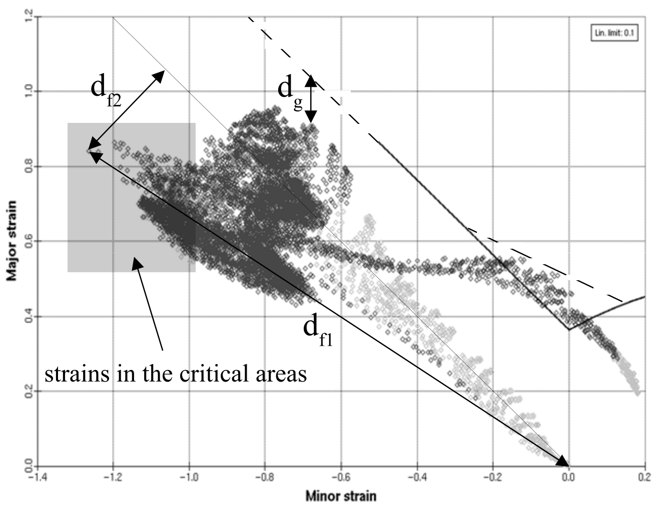


Figure 4.12: New definition of the objective function and the constraint

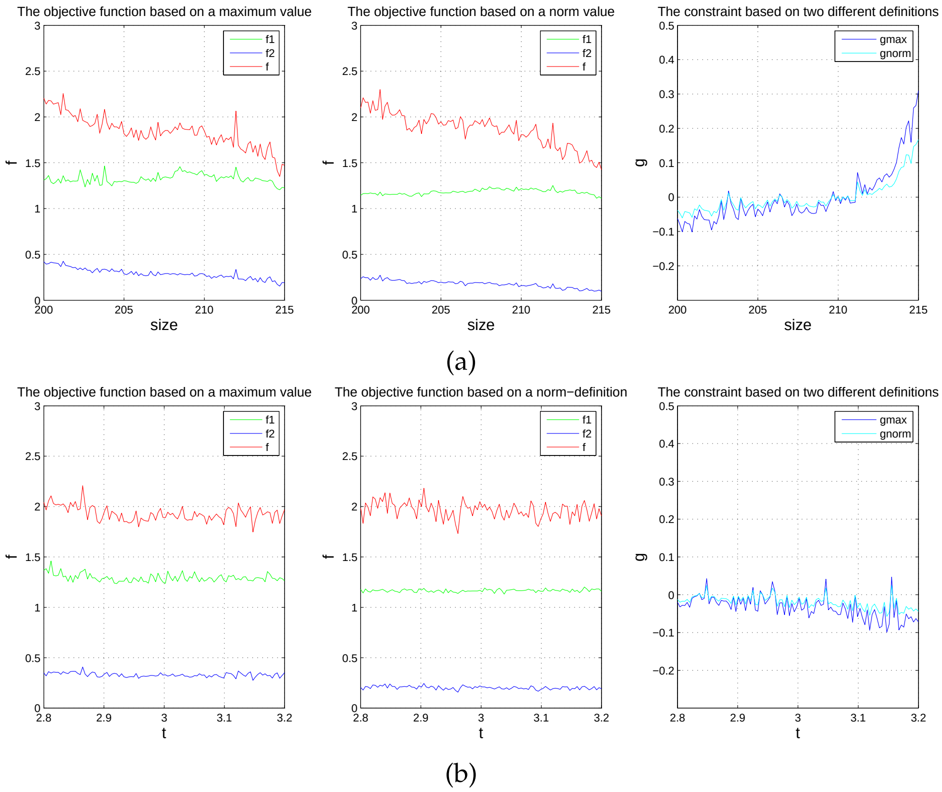


Figure 4.13: Response plots for (a) the variables size and (b) the variable thickness based on new definitions of the responses

influence on the crack occurrence (see Appendix A and it was also shown that for the modified process, which improved the process significantly, the thickness was reduced with respect to the reference process (see Figure 2.14. Therefore the definition of the objective function has been altered by including a part that judges the distances of the strains of the elements in the critical areas to the line of equal thickness. The new definition of the objective function is now based on the distance d_{f1} which is equal to the distance d_f defined in Equation 4.10 and by the distance d_{f2} of (ε_1,ε_2) of any element in the critical areas to the line of uniform thickness as displayed in Figure 4.12. The mathematical formulation of distances are presented in Equation 4.6.

$$\begin{aligned} d_{f_1} &= \sqrt{\varepsilon_1^2 + \varepsilon_2^2} \\ d_{f_2} &= \frac{|\varepsilon_1 + \varepsilon_2|}{\sqrt{2}} \\ d_g &= \varepsilon_1(\varepsilon_2) - \varepsilon_1^{\mathrm{FLC}}(\varepsilon_2) \end{aligned} \tag{4.6}$$

Both parts of the objective function are combined to one objective. This is done by making the parts proportional to each other by dividing the distances by the distances d_{ref_1} and d_{ref_2} belonging to the reference process with the original variable settings as presented in Table 4.2.

Two new definitions of the objective function and the constraint, based on a maximum value and on a norm value, now become:

$$
\begin{aligned}
\min f(\varepsilon_1, \varepsilon_2) &= \min \left(\max_{N_{\text{crit}}} \left(\frac{d_{f_1}}{d_{\text{ref}_1}} \right) + \max_{N_{\text{crit}}} \left(\frac{d_{f_2}}{d_{\text{ref}_2}} \right) \right) \\
\min f(\varepsilon_1, \varepsilon_2) &= \min \left(\frac{\frac{1}{\sqrt{n}} \|d_{f_1}\|_2}{\frac{1}{\sqrt{n_{\text{ref}}}} \|d_{\text{ref}_1}\|_2} + \frac{\frac{1}{\sqrt{n}} \|d_{f_2}\|_2}{\frac{1}{\sqrt{n_{\text{ref}}}} \|d_{\text{ref}_2}\|_2} \right)_{N_{\text{crit}}} \\
g_{\text{impl}} &= \max_N d_g \\
g_{\text{impl}} &= \begin{cases} \frac{1}{\sqrt[2]{n}} \left(\|d_g\|_2 \right)_{N_{\text{above FLC}}} & \text{if } \max_N d_g > 0 \\ \frac{1}{\sqrt[20]{n}} \left(-\|d_g\|_{20} \right)_N & \text{if } \max_N d_g \leq 0 \end{cases}
\end{aligned}
\tag{4.7}
$$

where N_{crit} denotes elements in the critical areas, N_{aboveFLC} the elements with strains above the FLC and N all elements. d_{f1} and d_{f2} are the distances in the FLD as presented in Figure 4.12 and d_{ref_1} and d_{ref_1} are these distances belonging to the reference process. n and n_{ref} are the number of the elements.
With these definitions the objective function value for the reference process becomes 2.

Results

The two new definitions for the objective function based on the strains of the elements in the new critical areas are tested for the variables size (x_{01}) and thickness (x_{02}). Also, the second definition for the constraint, based on norm values, is tried with the goal to reduce the noise in the response of the constraint. The results are shown in Figure 4.13. Besides the combined objective function, the two separate parts of the objective function defined by d_{f1} and d_{f2} are presented in this figure. The plots show that the adjustments to the optimisation problem are successful, because the response of f now reveals a trend for the variable size. Also, the norm definitions of the responses increase the trend in relation to the noise, therefore these definitions are chosen for a new optimisation model.

As described in Intermezzo 1, three adjustments have been made to the optimisation model. Besides a better definition of the critical ares, the mathematical formulation of the responses has been altered in two ways:

1. Alternative definition of the objective function
2. Application of norms to the mathematical formulation of the responses

The objective function is defined in such a way that one part of the objective function is still focussed on minimising the major strains while maximising the minor strains in the critical areas. However, another part of the objective function is now focussed on minimising the distance of the strains of the elements in the critical areas to the line of equal thickness. The objective function is defined by the distances d_{f1} and d_{f2} as displayed in Figure 4.12. Both parts of the objective function are combined to one objective. This is done by making the parts proportional to each other by dividing the distances by the distances d_{ref_1} and d_{ref_2} belonging to the reference process.

The second adjustment makes the objective function a non-critical value, because it is based on the strains of all the elements in the critical areas by applying a norm. This norm averages the distances of all elements in the critical areas, but larger values have more effect on the average. In this way, the influence of numerical noise on the response is reduced, but larger values still have more influence on the objective function than smaller values. The constraint is in the first place still a critical value, but norms are also applied to the mathematical formulation of the constraint to reduce the influence of noise on the response.

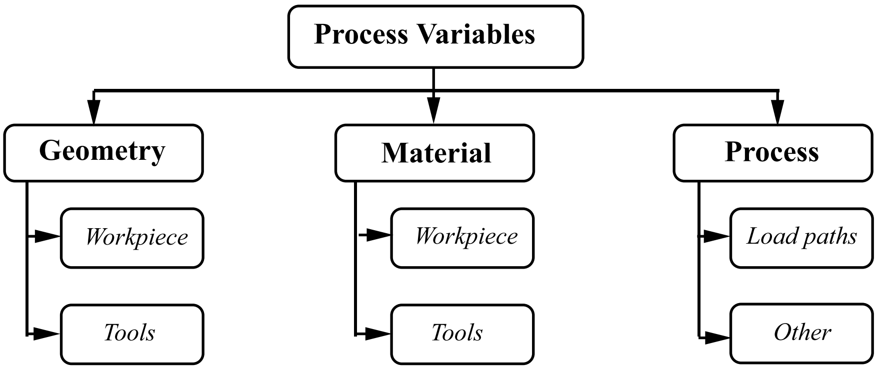


Figure 4.14: Process Variables in metal forming processes [11]

The new definitions of the objective function and the constraint are:

$$\begin{aligned} \min f(\varepsilon_1, \varepsilon_2) &= \min \left(\frac{\frac{1}{\sqrt{n}}\|d_{f_1}\|_2}{\frac{1}{\sqrt{n_{\text{ref}}}}\|d_{\text{ref}_1}\|_2} + \frac{\frac{1}{\sqrt{n}}\|d_{f_2}\|_2}{\frac{1}{\sqrt{n_{\text{ref}}}}\|d_{\text{ref}_2}\|_2} \right)_{N_{\text{crit}}} \\ g_{\text{impl}} &= \begin{cases} \frac{1}{\sqrt[2]{n}} \left(\|d_g\|_2\right)_{N_{\text{above FLC}}} & \text{if } \max_N d_g > 0 \\ \frac{1}{\sqrt[20]{n}} \left(-\|d_g\|_{20}\right)_N & \text{if } \max_N d_g \leq 0 \end{cases} \end{aligned} \tag{4.8}$$

where N_{crit} denotes elements in the critical areas, N_{aboveFLC} the elements with strains above the FLC and N all elements. d_{f1} and d_{f2} are the distances in the FLD as presented in Figure 4.12 and defined by Equation 4.6. d_{ref_1} and d_{ref_1} are these distances belonging to the reference process. n and n_{ref} are the number of the elements.

Step 5: Select only the necessary design variables

In step 1 it was identified that we are dealing with a Process Design Type I situation. Possible design variables in a Process Design Type I situation, are the Process Variables (PVs). The Process Variables are subdivided in material, process and geometry variables as shown in the top-down structure of Figure 4.14. As concluded in Chapter 2 the material used for the process is well chosen and therefore changing material variables does not seem to be a solution in this case. Process and design variables, however, can have a big influence on the objective of minimising the strains. The process variables in this case include friction coefficients μ's, the deep drawing depth d in the different steps and a blank holder force F. The design variables include blank size, shape and thickness t and geometry of the tools (radii r's and roundings ρ's of punches and dies). The Process Variables for process step OP02 are shown in Figure 4.15.

The production process is a complex process which consists of ten deep drawing operations. The operations do not all have a large influence on the strains in the critical areas and therefore the

Operations step	**Process Variables**
general	blank shape, t
OP02	$F_{\text{blankholder}}$, d, r_{punch}, r_{die}, ρ_{punch}, ρ_{die}, μ_{punch}, $\mu_{\text{die/binder}}$
OP04	d, r_{punch}, r_{die}, ρ_{punch}, ρ_{die}, μ_{punch}, μ_{die}
OP05 t/m OP16	no variations
OP18	ρ_{punch}, ρ_{die}, μ_{punch}, μ_{die}

Table 4.5: Process Variables in the deep drawing process of the bearing tube

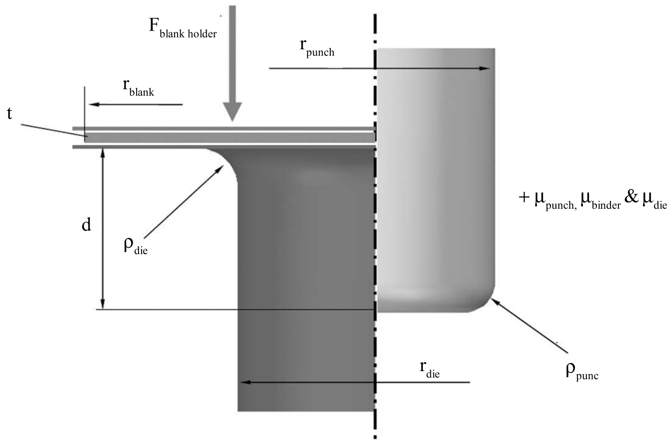


Figure 4.15: Process Variables in deep drawing step OP02 of the bearing tube

variables to take into account are limited to only a few operations which influence the strains the most, namely the first drawing operations OP02 and OP04 and the flanging operation OP18. Also, the trim line in OP11 and the geometry of the product after OP18 may not be changed, because in step 1 it was defined that the focus is on process redesign and not part redesign. An overview of the interesting process variables in the different steps is given in Table 4.5.

The different radii r and the deep drawing depth d which are Process Variables for the process of the bearing tube are excluded from optimisation, because these variables are difficult to change in AutoForm. Furthermore, the rounding of the punch ρ_{punch} in deep drawing step OP18 is excluded from the optimisation problem, because the punch in this step does not influence the process. Apart from the flanges no geometry is changed in step OP18 and the punch only keeps the product in place. The remaining variables are included for optimisation. Only two coefficients of friction are defined: one for all the punches and one for all the dies and the binder of deep drawing step OP02. These 10 variables with their original settings for the reference process as introduced in Section 2.2 are presented in Table 4.6.

Step 6: Select the ranges of the design variables

Keeping in mind that the final geometry of the product should stay the same, upper and lower bounds have been defined on all 10 design variables. These are presented in Table 4.7.

Step 7: Identify explicit constraints

For the optimisation problem of the bearing tube no relevant explicit constraints that describe impossible combinations of the design variables are present.

x_{01}	size	204 mm	x_{06}	ρ_{die04}	8 mm
x_{02}	t	3.0 mm	x_{07}	ρ_{punch04}	11 mm
x_{03}	ρ_{die02}	9 mm	x_{08}	ρ_{die18}	10 mm
x_{04}	ρ_{punch02}	12.5 mm	x_{09}	$\mu_{\text{die/binder}}$	0.1
x_{05}	$F_{\text{blankholder02}}$	100000 N	x_{10}	μ_{punch}	0.12

Table 4.6: The ten variables for optimisation of the reference process with their original settings

x_{01}	200 mm	215 mm	x_{06}	6 mm	10 mm
x_{02}	2.8 mm	3.2 mm	x_{07}	9 mm	13 mm
x_{03}	7.5 mm	12.5 mm	x_{08}	8 mm	12 mm
x_{04}	10 mm	15 mm	x_{09}	0.08	0.12
x_{05}	80000 N	120000 N	x_{10}	0.08	0.12

Table 4.7: The upper and lower bounds for the ten variables

The seven step methodology now yields the following mathematically formulated optimisation model:

$$\min f(\varepsilon_1, \varepsilon_2) = \min \left(\frac{\frac{1}{\sqrt{n}} \|d_{f_1}\|_2}{\frac{1}{\sqrt{n_{\text{ref}}}} \|d_{\text{ref}_1}\|_2} + \frac{\frac{1}{\sqrt{n}} \|d_{f_2}\|_2}{\frac{1}{\sqrt{n_{\text{ref}}}} \|d_{\text{ref}_2}\|_2} \right)_{N_{\text{crit}}}$$

$$\text{s.t.} \quad g_{\text{impl}} = \begin{cases} \frac{1}{\sqrt[2]{n}} \left(\|d_g\|_2 \right)_{N_{\text{above FLC}}} & \text{if } \max_N d_g > 0 \\ \frac{1}{\sqrt[20]{n}} \left(-\|d_g\|_{20} \right)_N & \text{if } \max_N d_g \leq 0 \end{cases} \leq 0$$

$$\begin{aligned} 200 \text{ mm} &\leq \text{size} \leq 215 \text{ mm} \\ 2.8 \text{ mm} &\leq t \leq 3.2 \text{ mm} \\ 7.5 \text{ mm} &\leq \rho_{\text{die02}} \leq 12.5 \text{ mm} \\ 10 \text{ mm} &\leq \rho_{\text{punch02}} \leq 15 \text{ mm} \\ 80000 \text{ N} &\leq F_{\text{blankholder02}} \leq 120000 \text{ N} \\ 6 \text{ mm} &\leq \rho_{\text{die04}} \leq 10 \text{ mm} \\ 9 \text{ mm} &\leq \rho_{\text{punch04}} \leq 13 \text{ mm} \\ 8 \text{ mm} &\leq \rho_{\text{die18}} \leq 12 \text{ mm} \\ 0.08 &\leq \mu_{\text{die/binder}} \leq 0.15 \\ 0.08 &\leq \mu_{\text{punch}} \leq 0.15 \end{aligned} \tag{4.9}$$

where t is the thickness of the blank, F the blank holder force, ρ the rounding of the tool and μ the coefficient of friction. d_{f1}, d_{f2} and d_g are the distances as displayed in Figure 4.12 and as defined by Equation 4.6.

By solving Equation 4.10 it will be tried to improve the reference process. These results will be compared to the modified process as introduced in 2.2. In Table 4.8 the values of the objective function and the constraint for the reference process and the modified process are presented. The negative constraints values denote that the strains stay below the FLC and both processes are feasible.

	f	g
Response values of the reference process	2.0000	-0.0171
Response values of the modified process	1.4823	-0.0459

Table 4.8: The response values for the reference and the modified process

4.2 Screening

The screening methodology which has been introduced in Section 3.3, is used to reduce the number of variables. A $2_{\text{III}}^{(10-6)}$ fractional factorial design is used to set up 16 FEM calculations with

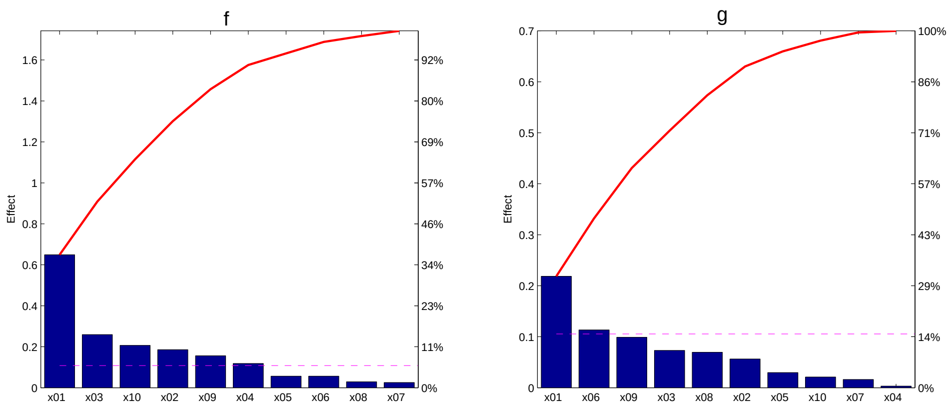


Figure 4.16: Pareto plots for the objective function and for the constraint

different variable settings. Subsequently an ANalysis Of VAriance (ANOVA) and Pareto plots are used to screen the 10 variables of the optimisation model.

The Pareto plots for the the objective function and the constraints are presented in Figure 4.16. The dashed line in the plots presents the estimated error of the metamodel. This error can be split in a pure error and a part that is caused by a lack of fit of the metamodel (see Equation 3.7). In case of computer experiments the pure error can be explained by the presence of numerical noise. The question now arises to what extent the error in Figure 4.16 presents a pure error and to what extent it presents a lack of fit. An analysis of the responses which is added in Intermezzo 2, showed that the responses of this optimisation problem are subject to a large percentage of numerical noise and it also showed that the assumption that interaction effects are of less interest than the main effects is true, meaning that the error caused by a lack of fit is not that large. The error thus mainly presents a pure error. This means that effects of the variables which have an effect smaller than the error can primary be caused by numerical noise. Including these variables would then result in making the responses more complex, without resulting in significantly better results. Therefore, we focus on the effects which are larger than this estimated error. The variables belonging to the effect larger than the estimated error, are shown in Table 4.9 (the real variables that belong to the coded variables x_{01} through x_{10} are presented in Table 4.6). In intermezzo 2, the responses of these variables are further analysed before selecting variables for a reduced optimisation problem.

Response	**Variables**
f	x01, x02, x03, x09, x10
g	x01, x03, x06, x09

Table 4.9: Variables with an effect larger than the estimated error

Intermezzo 2: Analysis of the responses for variable reduction

The responses of the variables with an effect larger than the estimated error are further analysed in this intermezzo. The responses for these variables are tested in the same way as was done for the variables x_{01} and x_{02} by running 100 FEM calculations: for every analysis the variable was varied from its minimum to its maximum value in 100 steps. The plots of this analysis are shown in Figure 4.18.

From a first sight it can be seen that none of these responses show a very clear trend. For example, although the effect for x_{10} on the objective function value is larger than the error, a trend cannot be found in the response. For x_{03} a trend can be seen, but the effect is small compared to the effect of x_{01}. It should be noted that for these response plots only one variable is varied and thus no interaction effects between variables are included. Nevertheless, the response plots give a good indication of the effect of the variables on the responses for the optimisation problem with 10 variables.

However, the question arises if the relatively large effect of variable x_{10} on the objective function that the Pareto plots show, is the effect of x_{10} on the response due to numerical noise which is shown Figure 4.18a or that this effect is the interaction effect of x_{01} an x_{02}. For screening a resolution III design is used that assumes that interaction effects are of less importance than main effects and in this case the main effect for x_{10} is aliased with the two-factor interaction effect of the variables x_{01} an x_{02}. To verify this assumption, a full factorial design for the variables x_{01} and x_{02} as described in Section 3.3 is used to execute four FEM calculations. This is a resolution V design in which none of the main effects or two-factor interactions are aliased with each other. The size of the interaction effect of x_{01} and x_{02} can be determined by this design. The Pareto plot of the effect of x_{01} and x_{02} on the objective function that follows after running the four FEM calculations belonging to the full factorial design for these two variables, is presented in Figure 4.17. Figure 4.19 shows the responses of the objective function and the constraint after running 100 FEM calculations in which the variables x_{01} and x_{02} are both varied in 10 steps. These plots show that the main effect of x_{02} is larger than the interaction effect of x_{01} and x_{02}. Thus, the effect of x_{10} that is shown in Figure 4.16 and which is larger than the effect of x_{02}, is expected to be indeed the effect of x_{10} and not the interaction effect of x_{01} and x_{02}.

Analysis of all response plots, makes us decide to take into account for optimisation only the variable x_{01} which seems to have the most influence on both responses. Also, variable x_{03}

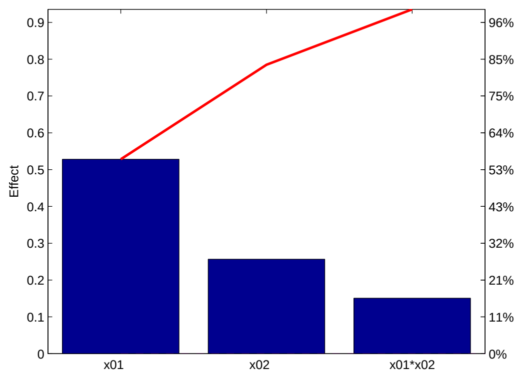


Figure 4.17: The Pareto for the objective function based on a full factorial design of the variables x_{01} and x_{02}

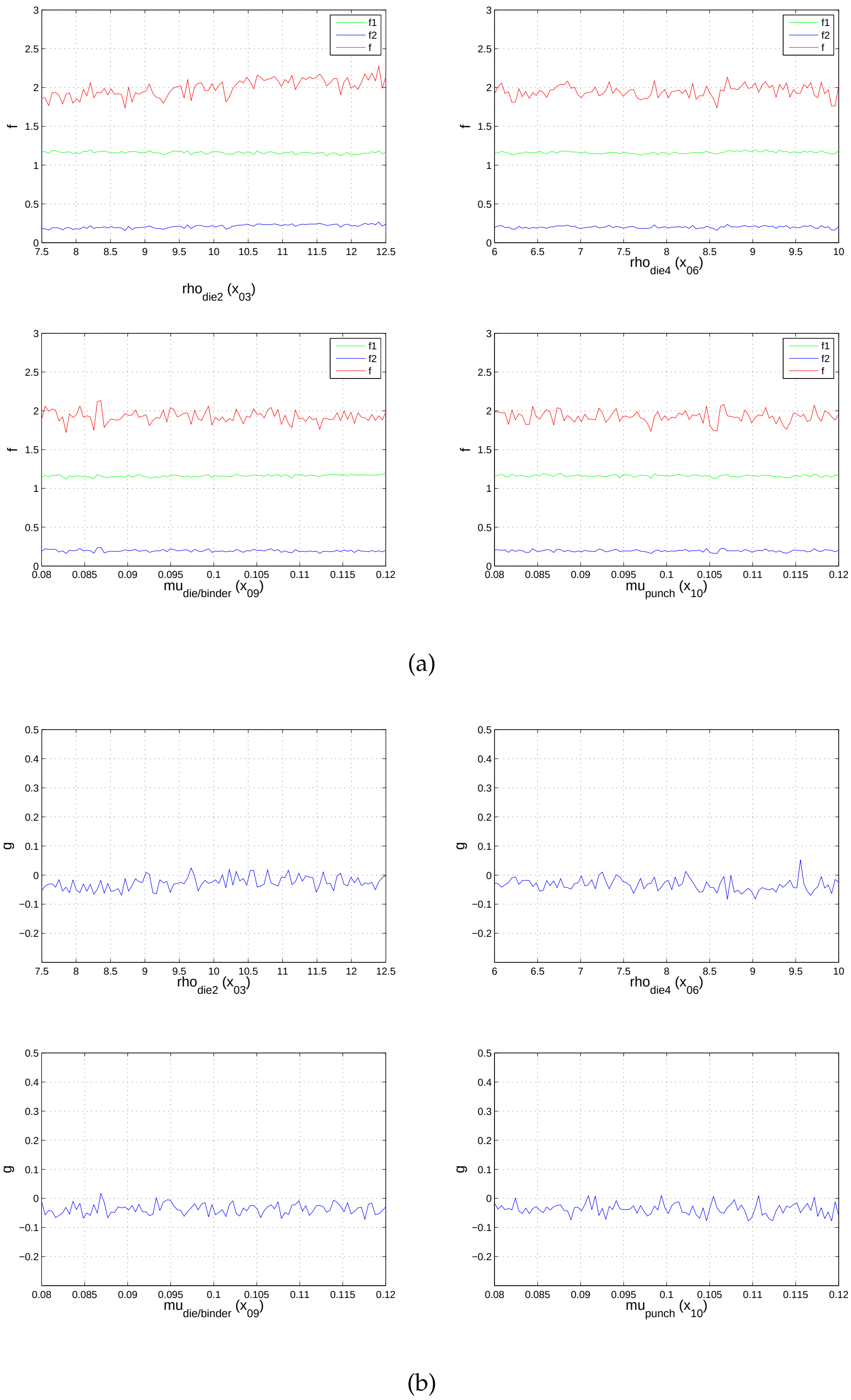


Figure 4.18: Figure (a) presents the objective function based on the norm definition for the variables x_{03}, x_{06}, x_{09} and x_{10} and (b) presents plots for the constraint for these variables

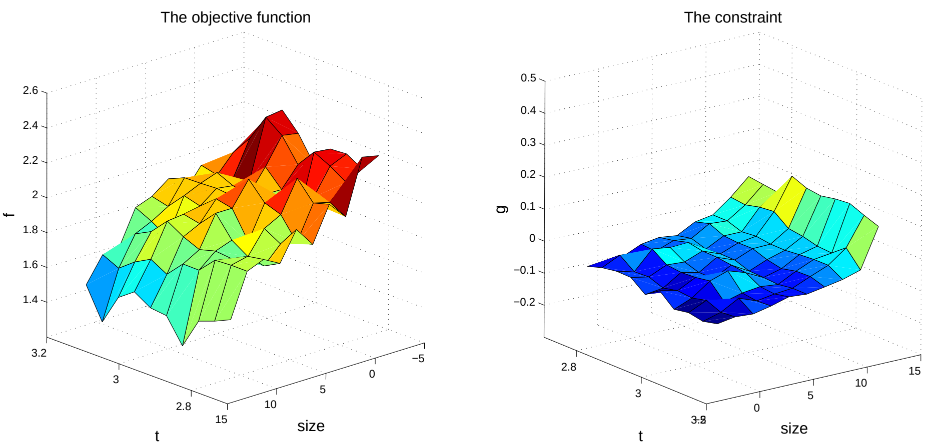


Figure 4.19: The responses for the objective function and the constraint for the variables x_{01} and x_{02}

shows a small trend in the response of the objective function. But the effect of x_{03} is only small and moreover, x_{03} has no large influence on the constraint. The expectation is therefore that when including x_{03} in the optimisation problem, an optimum will be found that is constrained by the boundary of this variable. This supports the possibility to set x_{03} to its minimum value to further minimise the objective function, but to exclude it from the optimisation problem, because it will make the problem more complex. However, for optimisation x_{03} is treated in the same way as the other variables that are excluded from the optimisation model. The variables x_{06}, x_{09} and x_{10} do not show any trend at all and will only cause more complex responses. Therefore, they are also excluded from the optimisation problem.

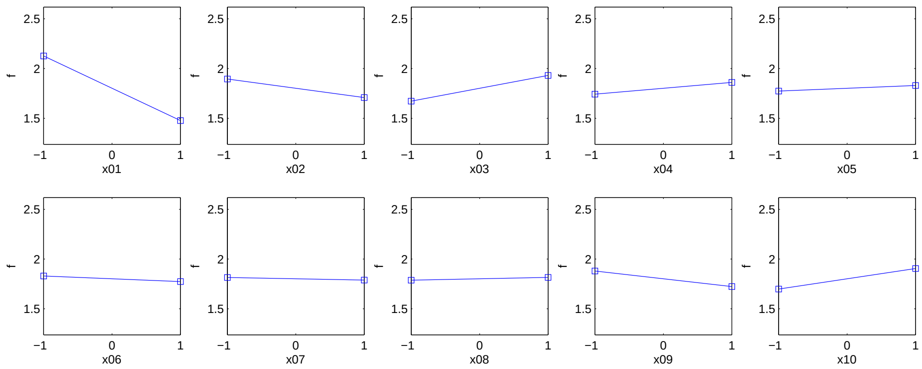
The observations discussed in Intermezzo 2 showed that the variable size (x_{01}) is the only variable with a significant effect on the objective function. The same observations led to the conclusion that for the constraint the variable size is the only variable of interest. Therefore, after screening it is decided to only take variable size into account for optimisation. The new optimisation model now becomes:

$$
\begin{aligned}
&\min f\left(\varepsilon_1, \varepsilon_2\right) = \min \left(\frac{\frac{1}{\sqrt{n}} \|d_{f_1}\|_2}{\frac{1}{\sqrt{n_{\mathrm{ref}}}} \|d_{\mathrm{ref}_1}\|_2} + \frac{\frac{1}{\sqrt{n}} \|d_{f_2}\|_2}{\frac{1}{\sqrt{n_{\mathrm{ref}}}} \|d_{\mathrm{ref}_2}\|_2} \right)_{N_{\mathrm{crit}}} \\
&\text{s.t.} \quad g_{\mathrm{impl}} = \begin{cases} \frac{1}{\sqrt[2]{n}} \left(\|d_g\|_2 \right)_{N_{\text{above FLC}}} & \text{if } \max_N d_g > 0 \\ \frac{1}{\sqrt[20]{n}} \left(-\|d_g\|_{20} \right)_N & \text{if } \max_N d_g \leq 0 \end{cases} \leq 0 \qquad (4.10) \\
&200 \text{ mm} \quad \leq \quad \text{size} \leq 215 \text{ mm}
\end{aligned}
$$

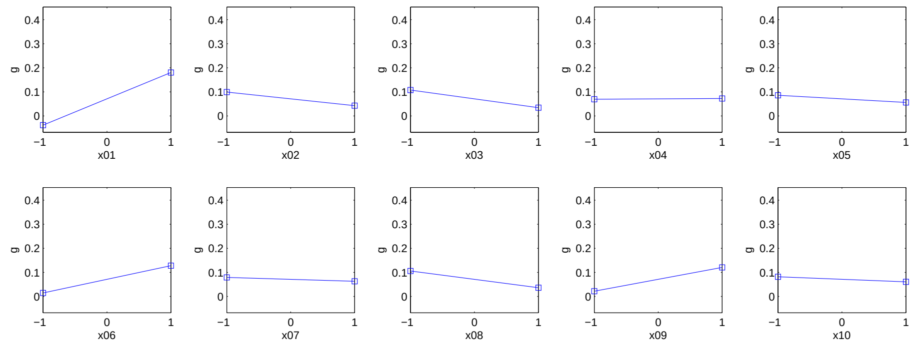
A question remaining is what to do with the other design variables. There are two options:

1. Set them to their original values
2. Set them to the optimal values which can be determined when using the effect plots

Because the process already exists, one optimisation is performed with the original settings of the variables to minimise the number of adjustments to the process. However, for comparison a second optimisation has been done with the optimal settings for the variables. These optimal



(a)



(b)

Figure 4.20: The effect of the ten variables on (a) the objective function and (b) the constraint

settings are determined using the effect plots which are presented in Figure 4.20a. In Table 4.10 presents these two sets of variable settings.

In Figure 4.20b the effects of the variables on the constraint are displayed. These plots show that the variable x_{01} has the largest effect on the constraint. It can also be seen that this is effect is opposite to the effect of x_{01} on the objective function: a larger blank size will result in a smaller objective function and reduce the occurring of the cracks, however if the blank is too large the constraint will get a positive value which means that necking will occur. This means that the optimisation will result in a constrained optimum.

	x_{01}	x_{02}	x_{03}	x_{04}	x_{05}	x_{06}	x_{07}	x_{08}	x_{09}	x_{10}
original variable settings	204	3.0	9	12.5	100000	8	11	10	0.10	0.12
lower bounds	200	2.8	7.5	10	80000	6	9	8	0.08	0.08
upper bounds	215	3.2	12.5	15	120000	10	13	12	0.15	0.15
optimal variable settings	-	3.2	7.5	10	80000	10	13	8	0.15	0.08

Table 4.10: Variables settings for optimisation

4.3 Optimisation

After having modelled the optimisation problem, it is now solved by the optimisation algorithm presented in Chapter 3. This is done with both the set of original variables settings and the set of optimal variable settings. Both optimisation are treated in this section.

4.3.1 Optimisation with original variables settings

For this optimisation the variables x_{02} through x_{10} were set to their original variables values as presented in Table 4.10.

A spacefilling Latin Hypercubes Design (LHD) combined with a full factorial Design Of Experiments (DOE) strategy as introduced in Section 3.4 was applied to generate 10 initial settings for the variable x_{01}. With the original variable settings and these settings for x_{01} 10 calculations were performed with the FEM code AutoForm. Subsequently, the four Response Surface and the three Kriging metamodels were fitted for the objective function f and the implicit constraint g.

Because the observations presented in Intermezzo 1 showed that numerical noise has a significant influence on the responses, the question arises which metamodel is best suited for this optimisation model: a RSM or a Kriging metamodel? RSM metamodels are less flexible than Kriging metamodels, whereas the disadvantage of Kriging metamodels is their sensitivity to numerical noise. To compare the applicability of both metamodels, two optimisations have been carried out using the original variable settings, using RSM metamodels for the first and using Kriging metamodels for the second. Thus, two optimisations are performed:

1. Optimisation using RSM metamodels and with the original variables settings
2. Optimisation using Kriging metamodels and with the original variable settings

The metamodels were validated after each batch. The validation plots after the last batches are presented in Figure 4.21. The metamodels were subsequently optimised using the multistart SQP-algorithm described in Section 3.7. For RSM a second batch of five experiments was designed by applying a maximin spacefilling LHD, which equals sequential improvement by applying Minimising the Merit Function with a weight factor $w \to \infty$. For the optimisation with Kriging metamodels sequential improvement is applied by MMF as introduced in Section 3.8 with a weight factor $w = 1$. The optimisation included three batches and in total 14 FEM calculations were performed.

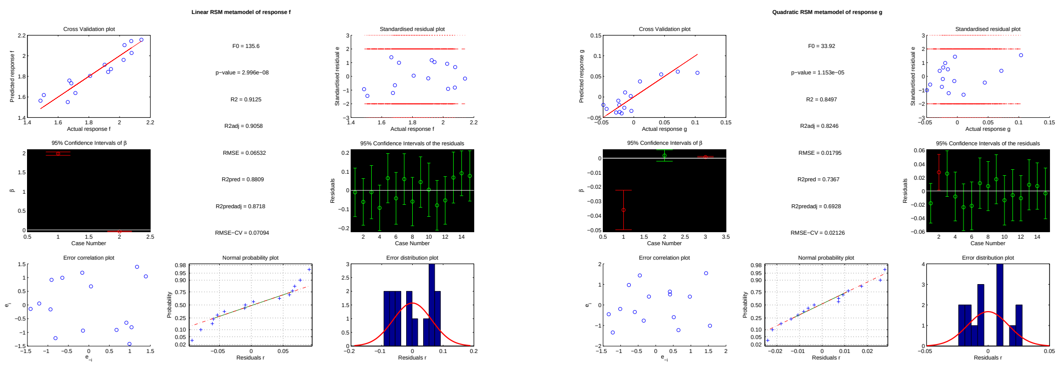
	R^2_{pred}	R^2_{predadj}	RMSE
Values of the RSM metamodel of f used for optimisation	0.8809	0.8718	0.06532
Values of the RSM metamodel of g used for optimisation	0.7367	0.6928	0.01795
Values of the RSM metamodel of f based on the 100 calculations	0.8008	0.7967	0.08097
Values of the RSM metamodel of g based on the 100 calculations	0.7492	0.7441	0.02184
Values of the Kriging metamodel of f used for optimisation	0.8519	0.7778	0.07666
Values of the Kriging metamodel of g used for optimisation	0.4849	0.2274	0.02675
Values of the Kriging metamodel of f based on the 100 calculations	0.8293	0.8202	0.07616
Values of the Kriging metamodel of g based on the 100 calculations	0.8987	0.8933	0.01431

Table 4.11: The values for validating the metamodels

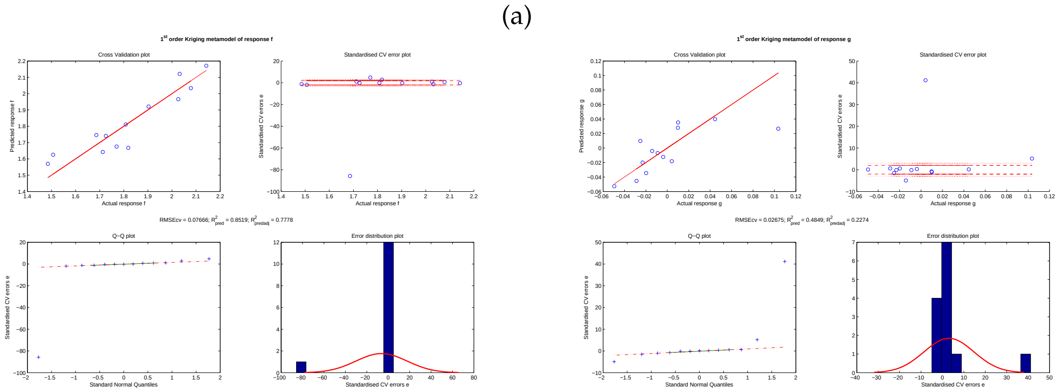
For comparison, 100 FEM were run in which the variable x_{01} was equally varied in 100 steps. In Figure 4.22 the metamodels for as well the optimisations as for these 100 response points are visualised. The dashed lines indicate metamodels of the 100 calculations. The R^2_{pred}, R^2_{predadj} and RMSE values belonging to the validation plots of the metamodels used for optimisation as well as of the metamodels based on these 100 calculations are presented in Table 4.11. The high R^2_{pred} and R^2_{predadj} values show that the metamodels are approximating the responses well. Only the Kriging metamodel for g is not as accurate as the other metamodels. The values for the RSM metamodel of f based on the 100 calculations shows that the RSM metamodel will not become much more accurate after many calculations. This observation gives reason to conclude that running more than 15 calculations for optimisation using RSM metamodels will not improve the optimisation results much. The Kriging metamodel of the constraint for the 100 calculations is more accurate than the RSM metamodel and much more accurate than the Kriging metamodel after 14 calculations. Thus it can be concluded that adding more experimental points will finally result in a more accurate Kriging metamodel and subsequently a better optimum can be found. However, more batches and subsequently more time consuming computer simulations are needed.

The optima found after optimising the metamodels after the last batches are validated by running a FEM calculation with these optimum values for x_{01}. It turned out that the constraint value for the optimum found for the last Kriging metamodel is positive; this denotes that this is not a feasible optimum. This is not completely unexpected, because the metamodel for the constraint was not very accurate. More calculations would be needed to improve the metamodel of g and find a feasible optimum. Nevertheless, the best found feasible objective function value resulting from the 14 calculations, the constraint value and the variable value are displayed in Table 4.12. Also, for RSM the optimum of the last metamodel is not the best value (although it is feasible): the 11^{th} calculation which is the optimum found after the first batch, delivers the best (feasible) objective function value. This value is also presented in Table 4.12.

For this optimisation problem using RSM metamodels showed to result in a slightly better objective function value, but the difference is smaller than the influence of numerical noise and cannot be assigned to a better metamodel. After running more than 15 calculations RSM is not expected to deliver much more accurate metamodels. Kriging will probably finally result in a slightly better optimum when more design points are added, but the consideration between running more expensive FEM calculations and the expected small improvement it will yield, makes us decide to stop optimisation. The difference between the objective function values of both optimisation

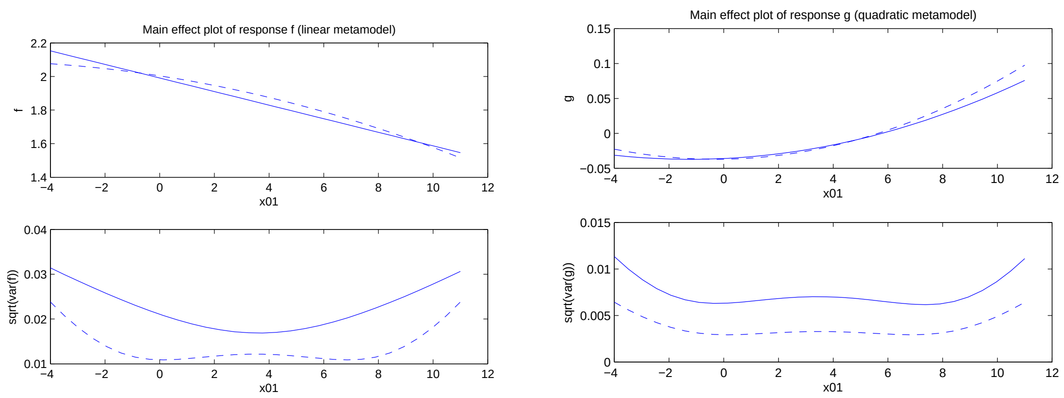


(a)

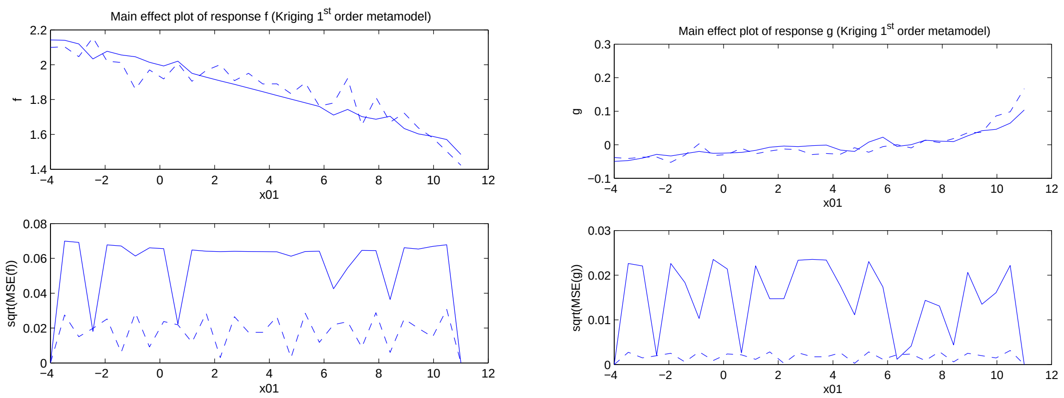


(b)

Figure 4.21: Validation plots of (a) the RSM metamodels and (b) the Kriging metamodels after the last batches; left belonging to the objective function and right belonging to the constraint



(a)



(b)

Figure 4.22: (a) RSM metamodels and (b) Kriging metamodels of the responses after the last batches; the dashed lines present the metamodels of the 100 calculations

is smaller than the influence of numerical noise (see Table 4.4 and thus both approaches resulted in an equally improved objective function value (after the same number of calculations). Compared to the reference process these optimised values improved the process much, although the modified process still results in a better objective function value.

The Forming Limit Diagrams (FLDs) with the strains of all the elements are presented in Figures 4.23a and 4.24a for the reference process, the modified process and the optimised reference process. These figures show that all the strains stay below the FLC and the processes fulfill the constraint. In Figures 4.23b and 4.24b only strains belonging to elements in the critical areas are displayed and in Figures 4.23c and 4.24c the thickness is showed.

	f	g	x_{01}
Values of the optimisation using RSM metamodels	1.6732	-0.0049	209.90 mm
Values of the optimisation using Kriging metamodels	1.6861	-0.0141	210.50 mm
Best value after running 100 calculations	1.7402	-0.0063	210.00 mm
Values of the reference process	2.0000	-0.0171	204.00 mm
Values of the modified process	1.4823	-0.0459	204.00 mm

Table 4.12: The optimisation results of the optimisation with the original variable settings

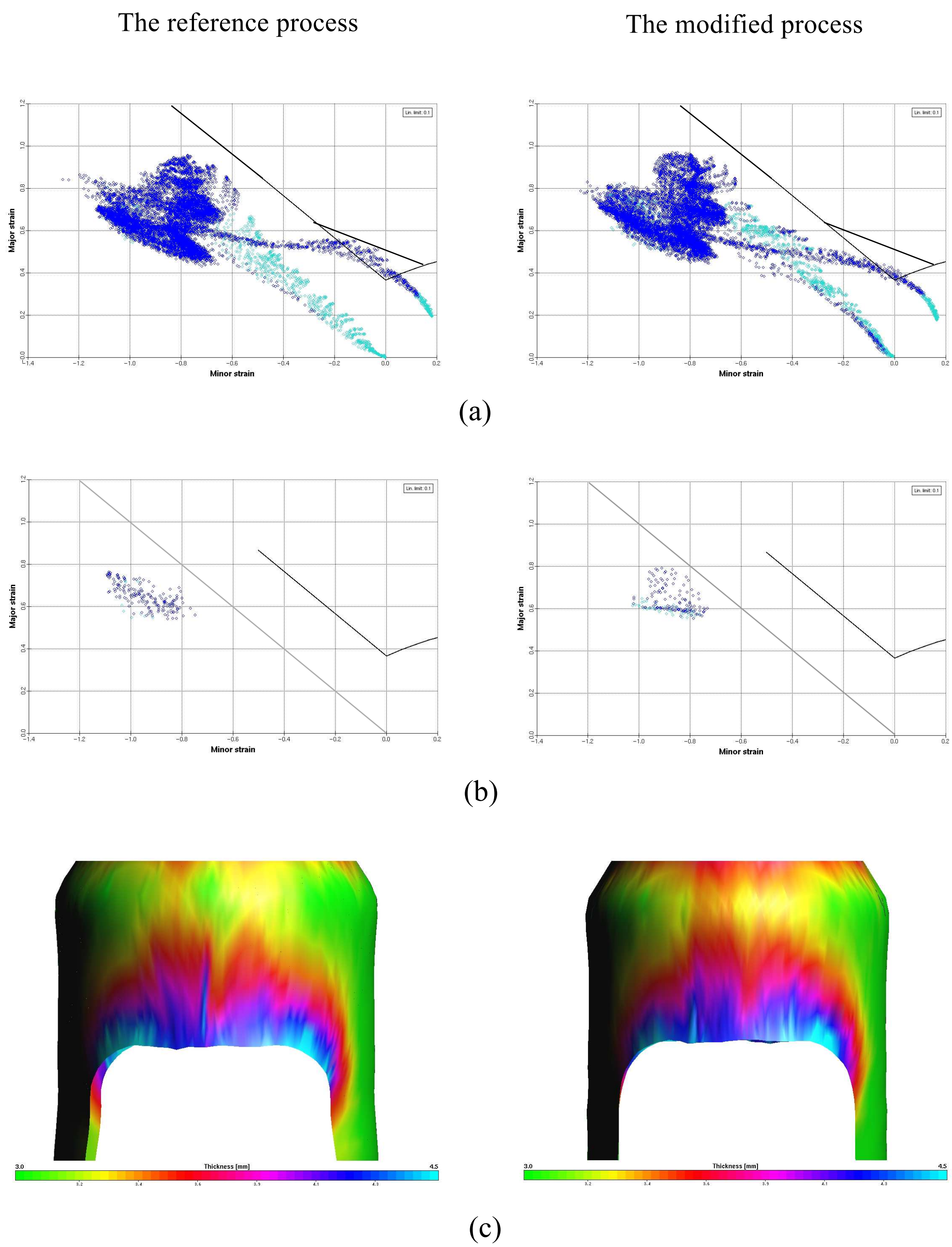


Figure 4.23: (a) The FLD with strains of all the elements (b) the FLD with strains of elements that are located in the critical areas and (c) the thickness for the reference and the modified process

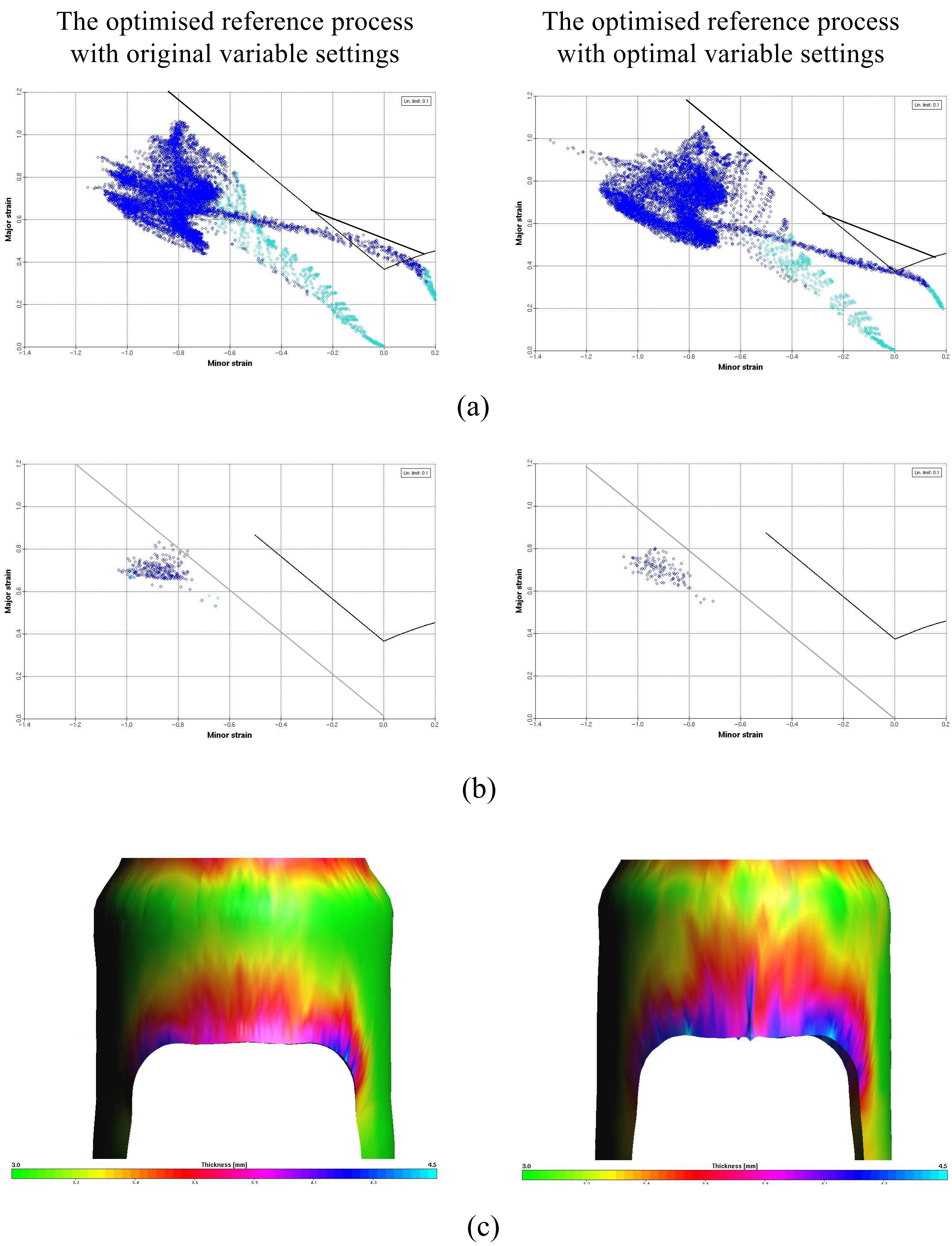


Figure 4.24: (a) The FLD with strains of all the elements (b) the FLD with strains of elements that are located in the critical areas and (c) the thickness for the optimised reference process with original and optimal variables settings

Comparing the reference process to the modified process, shows that the minor strains in the critical areas get less extreme values and major strains increase somewhat. For the optimal process after optimisation of the reference process this is also the case (the strains shift a little to the line of equal thickness), but the effect is less. In the modified process the thickness in the corners reduce a little with respect to the reference process: the maximum thickness shift a little to the middle of the product. Also in the optimised reference process the thickness reduces clearly, however, the largest thickness still can be found in the corners.

4.3.2 Optimisation with optimal variable settings

For this optimisation the variables x_{02} through x_{10} were set to their optimal variables values as presented in Table 4.10.

The responses of the optimisation with the optimal variable settings are expected to look like the responses in case of the optimisation with original variable settings. For the latter, RSM and Kriging metamodels proved to result in comparable optima that are accurate enough for this optimisation problem after the same number of calculations. In other words, the difference between using Kriging or RSM metamodels is small for this optimisation. The optimisation with optimal variable settings is performed using only RSM metamodels.

The same DOE strategy is applied to generate the settings for the variable x_{01} to run 10 FEM calculations with, this time, the optimal variable settings. The next step is to fit metamodels through the responses of these calculations. The validation plots of the RSM metamodels are shown in Figure 4.25. For f the validation plot shows R^2_{pred} and R^2_{predadj} values of respectively 0.8878 and 0.8557 and the value of the RMSE is 0.08003. These values are for g respectively 0.8804, 0.8463 and 0.03653. These metamodels are accurate enough regarding the influence of numerical noise on the responses. Thus, no extra batch is added and the optimum found by optimising the metamodel which is fitted through the 10 response points, is validated. The metamodels for f and g are visualised in Figure 4.26. The optimisation results are presented in Table 4.13.

The negative value for the implicit constraint denotes that the formability of the product is guaranteed. The optimised objective function value outperforms the objective function value obtained by the optimisation with the original variable settings. This demonstrates an easy improvement of the optimum by choosing optimal settings for the unimportant variables based on the screening results.

For the optimised reference process with the optimal variable settings the Forming Limit Diagram with the strains of all the elements, with strains belonging to the elements in the critical areas and the thickness are again presented (see Figure 4.24). Figure 4.24a shows that all the strains

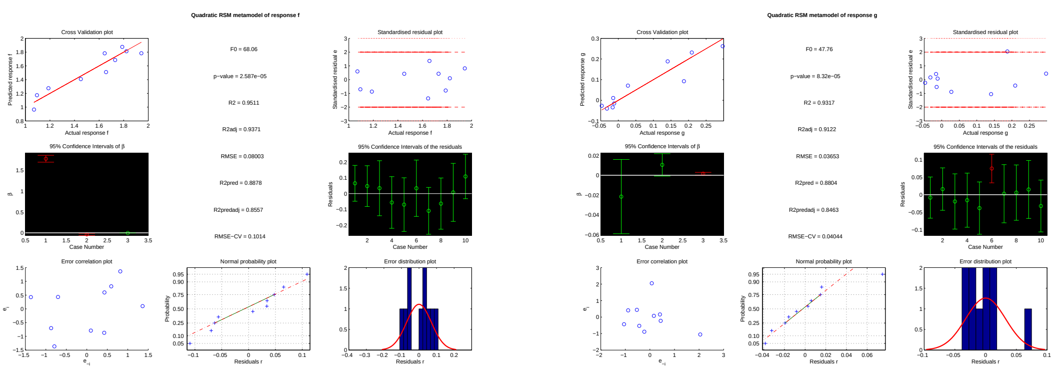


Figure 4.25: Validation plots of the RSM metamodels; left belonging to the objective function and right belonging to the constraint

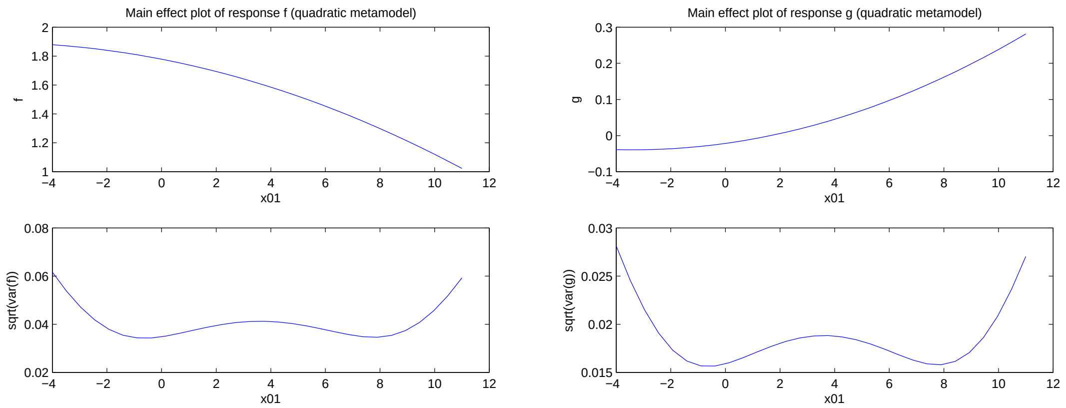


Figure 4.26: RSM metamodels of the objective function and the constraint

stay below the FLC and the process fulfills the constraint. In Figure 4.24b it can be seen that the maximum strains which were present in the reference process have reduced and the strains again shifted a little to the line of equal thickness. Comparing the optimised process with the original variable settings with the optimised process with the optimal variables settings shows that the maximum minor strains not increase further, but the maximum thickness in the corner reduces a little and the maximum thickness shifts from the corner to the middle of the product. The latter observation was also the case in the modified process compared to the reference process.

In Figure 4.27 the location of other large strains that are caused by wrinkling in the middle of the product is indicated. It was assumed that wrinkling will not form a problem in the final product, however it should be verified in reality if no problems will exist at the locations where the wrinkling seems to appear.

	f	g	x_{01}
Values of the optimisation using original variable settings	1.6732	-0.0049	209.90 mm
Values of the optimisation using optimal variable settings	1.6457	-0.0118	204.61 mm

Table 4.13: The optimisation results of the optimisation with the optimal variable settings

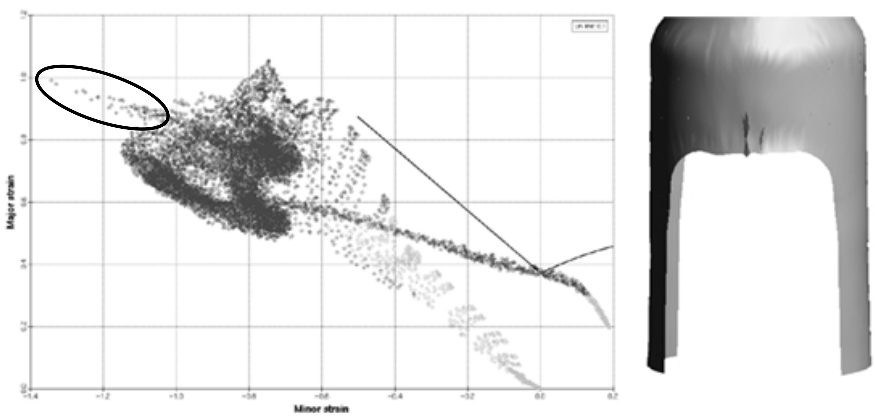


Figure 4.27: Large strains in the middle of the product due to wrinkling

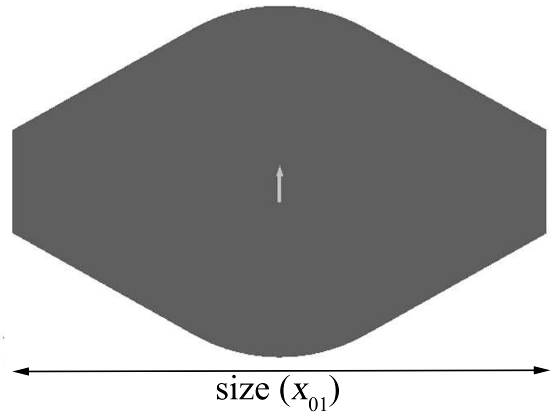


Figure 4.28: The optimised variable blank size

4.4 Conclusions

The optimisation strategy as described in Chapter 3 has been been applied to the reference process of the bearing tube. A first optimisation delivered unsatisfying results: after running 169 calculations the objective function value did not show real convergence behaviour and the metamodels were far from accurate. Therefore the responses were analysed and it was concluded that the responses suffer from a large percentage of numerical noise and that a trend could not be revealed. Three adjustments have been made to improve the optimisation model:

1. The critical areas in which the cracks occur are better defined
2. An alternative mathematical formulation of the objective function is defined
3. Norms are applied to the mathematical formulation of the responses

The improved optimisation model was screened and this resulted in a reduced optimisation model with only one variable. To this reduced optimisation model two optimisations are performed:

1. One with the original variables settings for the design variables that are excluded from the optimisation model

	Reference process	Modified process	Optimised ref. process with original settings	Optimised ref. process with optimal setting
f	2.000	1.4823	1.6732	1.6457
g	-0.0171	-0.0459	-0.0049	-0.0118
x_{01}	204 mm	204 mm	209.90 mm	204.61 mm
x_{02}	3.0 mm	3.0 mm	3.0 mm	3.2 mm
x_{03}	9 mm	9 mm	9 mm	7.5 mm
x_{04}	12.5 mm	12.5 mm	12.5 mm	10 mm
x_{05}	100000 N	100000 N	100000 N	80000 N
x_{06}	8 mm	8 mm	8 mm	10 mm
x_{07}	11 mm	11 mm	11 mm	13 mm
x_{08}	10 mm	10 mm	10 mm	8 mm
x_{09}	0.1	0.1	0.1	0.15
x_{10}	0.12	0.12	0.12	0.08

Table 4.14: The optimisation results of the reference process

2. One with the optimal variables settings for the design variables that are excluded from the optimisation model

For the first optimisation both RSM and DACE metamodels were applied. It turned out that both metamodelling techniques delivered more or less equal optimised objective function values and that both techniques are suitable for the optimisation problem as defined in this chapter.

With the optimisations it was attempted to improve the reference process. This was successful: the optimised objective function values for as well the optimised reference process with the original as with the optimal variable settings are improved with respect to the objective function value of the original reference process. Furthermore, the optimisation with the optimal variable settings showed an improvement with respect to the optimisation with the original variable settings. Comparison of the results to the modified process, showed that the modified process still outperforms the optimised reference process. Figure 4.28 presents the optimised variable size (x_{01}). The optimised value of this variable, the other variable settings and the response values for the different processes are presented in Table 4.14. The strains and thickness of these processes are presented in Figures 4.23 and 4.24.

Chapter 5

Optimisation of the modified process of the bearing tube

The optimisation strategy is also applied to the modified process with an extra blank holder in OP16 as described in Section 2.2 in order to improve the current situation. This is done analogous to Chapter 4. In the first three section the modelling, screening and optimisation are described and the last section summarises some conclusions.

5.1 Modelling

Again the seven step procedure is applied to describe the mathematical optimisation problem. The responses, the objective function and implicit constraint are equal to those for the optimisation problem of the reference process, hence for steps 1 to 4 we refer to Section 4.1. The process, however, is slightly different to the process discussed in the former chapter. Consequently, the process variables are not exactly the same. Steps 5 to 7 are therefore executed again.

Step 5: Select only the necessary design variables

An overview of the interesting process variables for the modified process is given in Table 5.1. In the modified process only the process step OP16 is altered compared the reference process. In this process step a blank with a blank holder force is added, the other variables are equal to process step OP16 of the reference process. The large deformations in the critical areas now appear during the deep drawing step OP16. In the last drawing step, OP18, the strains in these areas do not change and therefore the process variables of this step are of no interest for optimisation.

Operations step	**Possible variations**
general	blank size, blank shape, t
OP02	$F_{\text{blankholder}}$, d, r_{punch}, r_{die}, ρ_{punch}, ρ_{die}, μ_{punch}, $\mu_{\text{die/binder}}$
OP04	d, r_{punch}, r_{die}, ρ_{punch}, ρ_{die}, μ_{punch}, μ_{die}
OP05 t/m OP11	no variations
OP16 + extra step	$F_{\text{blankholder}}$, d, ρ_{punch}, ρ_{die}, μ_{punch}, $\mu_{\text{die/binder}}$
OP18 + extra step	no variations

Table 5.1: Variables in the modified deep drawing process of the bearing tube

In the same way as in Section 4.1 a set of variables to take into account for optimisation is chosen. The difference between this set of variables and the set of variables for the optimisation problem in Chapter 4 is the extra blank holder force we include and the radius of the die of process step OP16 instead of the one of process step OP18. However, these dies have the same geometry. We now have selected 11 variables which are presented in Table 5.2.

Step 6 and 7: Select the ranges of the design variables and identify explicit constraints

The same upper and lower bounds are selected for the variables as for the variables in the reference process. The boundaries for the extra blank holder force are chosen to be of the same size as the boundaries for the blank holder of process step OP02. Table 5.3 present the ranges of the variables. As for the reference process, no relevant explicit constraints that describe impossible combinations of the design variables are present for this optimisation problem.

The model now becomes:

$$\min f(\varepsilon_1, \varepsilon_2) = \min \left(\frac{\frac{1}{\sqrt{n}} \|d_{f_1}\|_2}{\frac{1}{\sqrt{n_{\mathrm{ref}}}} \|d_{\mathrm{ref}_1}\|_2} + \frac{\frac{1}{\sqrt{n}} \|d_{f_2}\|_2}{\frac{1}{\sqrt{n_{\mathrm{ref}}}} \|d_{\mathrm{ref}_2}\|_2} \right)_{N_{\mathrm{crit}}} \tag{5.1}$$

$$\text{s.t.} \quad g_{\mathrm{impl}} = \begin{cases} \frac{1}{\sqrt[2]{n}} \left(\|d_g\|_2\right)_{N_{\text{above FLC}}} & \text{if } \max_N d_g > 0 \\ \frac{1}{\sqrt[20]{n}} \left(-\|d_g\|_{20}\right)_N & \text{if } \max_N d_g \leq 0 \end{cases} \leq 0$$

$$\begin{aligned} 200 \text{ mm} &\leq \text{size} \leq 215 \text{ mm} \\ 2.8 \text{ mm} &\leq t \leq 3.2 \text{ mm} \\ 7.5 \text{ mm} &\leq \rho_{\mathrm{die02}} \leq 12.5 \text{ mm} \\ 10 \text{ mm} &\leq \rho_{\mathrm{punch02}} \leq 15 \text{ mm} \\ 80000 \text{ N} &\leq F_{\mathrm{blankholder02}} \leq 120000 \text{ N} \\ 6 \text{ mm} &\leq \rho_{\mathrm{die04}} \leq 10 \text{ mm} \\ 9 \text{ mm} &\leq \rho_{\mathrm{punch04}} \leq 13 \text{ mm} \\ 8 \text{ mm} &\leq \rho_{\mathrm{die16}} \leq 12 \text{ mm} \\ 80000 \text{ N} &\leq F_{\mathrm{blankholder16}} \leq 120000 \text{ N} \\ 0.08 &\leq \mu_{\mathrm{die/binder}} \leq 0.15 \\ 0.08 &\leq \mu_{\mathrm{punch}} \leq 0.15 \end{aligned} \tag{5.2}$$

where N_{crit} denotes the elements in the critical areas, $N_{\text{above FLC}}$ the elements that exceed the FLC and N all elements. n is the number of the elements. d_{f1}, d_{f2} and d_g are the distances as displayed

x_{01}	size	204 mm	x_{07}	ρ_{punch04}	11 mm
x_{02}	t	3.0 mm	x_{08}	ρ_{die16}	10 mm
x_{03}	ρ_{die02}	9 mm	x_{09}	$F_{\mathrm{blankholder16}}$	120000 N
x_{04}	ρ_{punch02}	12.5 mm	x_{10}	$\mu_{\mathrm{die/binder}}$	0.1
x_{05}	$F_{\mathrm{blankholder02}}$	100000 N	x_{11}	μ_{punch}	0.12
x_{06}	ρ_{die04}	8mm			

Table 5.2: The eleven variables for optimisation of the modified process with their original settings

x_{01}	200 mm	215 mm	x_{07}	9 mm	13 mm
x_{02}	2.8 mm	3.2 mm	x_{08}	8 mm	12 mm
x_{03}	7.5 mm	12.5 mm	x_{09}	80000 N	120000 N
x_{04}	10 mm	15 mm	x_{10}	0.08	0.15
x_{05}	80000 N	120000 N	x_{11}	0.08	0.15
x_{06}	6 mm	10 mm			

Table 5.3: The upper and lower bounds for the eleven variables

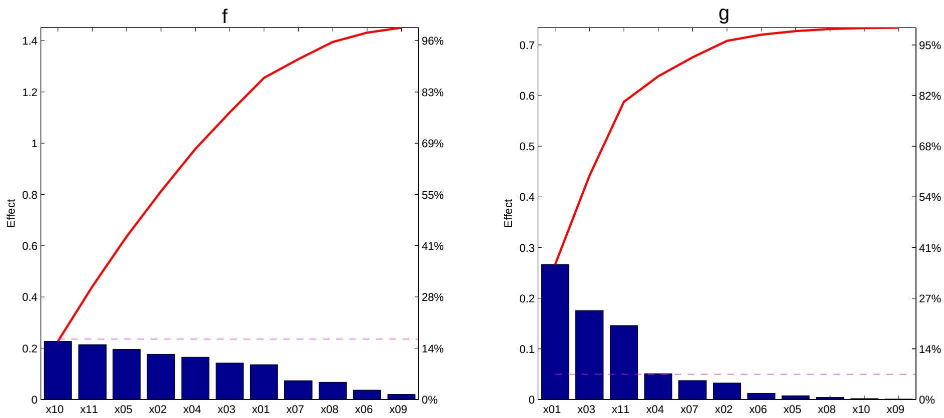


Figure 5.1: Pareto plots for the objective function and for the constraint

in Figure 4.12 and as defined by Equation 4.6. d_{ref_1} and d_{ref_2} are the distances d_{f1} and d_{f2} for the reference process. t is the thickness of the blank, F the blank holder force, ρ the rounding of the tool and μ the coefficient of friction.

By solving Equation 5.1 it will be tried to improve the modified process as introduced in Section 2.2. In Table 5.4 the values of the objective function and the constraint for the reference process and the modified process are presented. The negative constraints values denote that the strains stay below the FLC and both processes are feasible.

	f	g
Response values of the reference process	2.0000	-0.0171
Response values of the modified process	1.4823	-0.0459

Table 5.4: The response values for the reference and the modified process

5.2 Screening

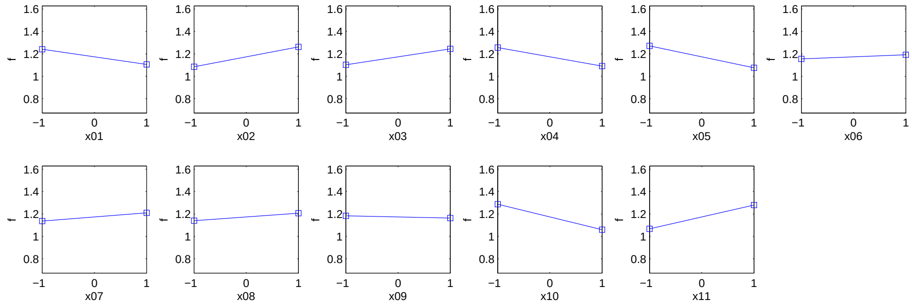
The model is screened using a $2^{(11-7)}_{\mathrm{III}}$ fractional factorial design. The Pareto and effect plots for f and g are shown in Figures 5.1 and 5.2. From the Pareto plot of f it can be seen that there are no variables that have a dominating influence on the process.

For the constraints there are three variables with a significant effect on the response. Choosing these variables well can be important to guarantee the formability of the product.

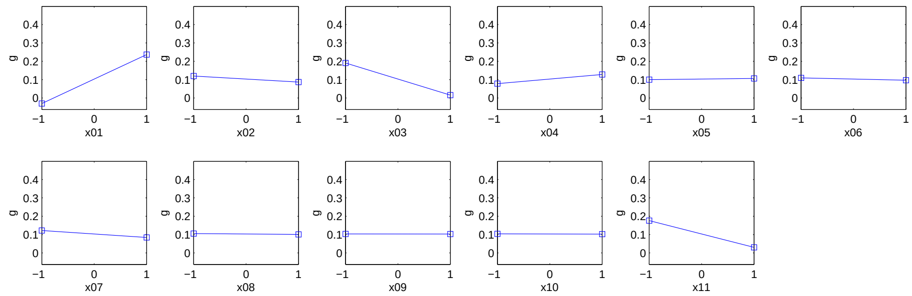
There are no variables with a dominant influence on the response, one might decide to optimise the optimisation problem with the most, if not all, of these variables. In this case it is decided not to define a reduced optimisation problem, but to perform the optimisation with all the variables.

5.3 Optimisation

Because the responses are subject to a large percentage of numerical noise, including all the variables will make the responses very complex. Therefore, it is decided to only execute a very simple optimisation based on the screening results. The variables settings for the calculations performed



(a)



(b)

Figure 5.2: The effect of the eleven variables on (a) the objective function and (b) the constraint

for screening and the effects of the variables on the objective function and the constraint following from screening are analysed. With this information a few variable settings are cleverly chosen using only maximum and maximum values of the variables. Then, only a few FEM calculations need to be executed to calculate the objective function and constraint values for the process with these variable settings. With this optimisation it will be tried to improve the modified process and furthermore, it will show what the improvement can be when using optimal variable settings for variables that seem to have only a small influence on the objective function (after all, none of the variables seems to have a large effect on the objective function).

In Section 5.3.1 cleverly choosing a few variable settings is described. Subsequently, in Section 5.3.2 the results of this optimisation are discussed.

5.3.1 Choosing optimal variable settings

The effect plots in Figure 5.2 are used to determine the optimal variable settings. The settings for the variables for the best f and for the best g value are displayed in Table **??**. In the effect plots -1 denotes a minimum variable value and 1 denotes a maximum variable value. The objective function is minimised and the constraint return a value ¡ 0, therefore for both f and g a lower value is better. The effect plot now show that for almost all variables the direction of the effect on the objective function is opposite to the direction of the effect on the constraint. This means that if we choose the best values for the variables with respect to the objective function, probably a low objective function value will result, but a high constraint value. This may result in a infeasible process.

Besides the effect plots, we can look at the response values of the 16 screening calculations. Seven of these calculations deliver feasible products ($g < 0$). For the variable settings of all these seven calculations x_{01} is set on its minimum value. This corresponds to the Pareto plot of g: this plot showed that x_{01} has a large influence on the constraint.

With this information four different variable settings are defined, which are displayed in Table 5.5. The first set includes all settings that will result in a low objective function following from the effect plots. A process with these variables is not likely to be feasible, because these settings have a negative influence on the constraint. Therefore, three other sets of variable settings are defined where the variable x_{01} is set on its minimum value, because it appeared that the value of this variable is critical for the constraint. For these three sets of variables different combinations of x_{03} and x_{11} are tried, because these variables have (except for x_{01}) the largest influence on the constraint as the Pareto plot for the constraint in Figure 5.1 showed. The real minimum and maximum values of the variables were presented in Table 5.3. The four FEM calculations with these variable settings are run.

	$\boldsymbol{x_{01}}$	x_{02}	$\boldsymbol{x_{03}}$	x_{04}	x_{05}	x_{06}	x_{07}	x_{08}	x_{09}	x_{10}	$\boldsymbol{x_{11}}$
variable set 1	**1**	-1	**-1**	1	1	-1	-1	-1	1	1	**-1**
variable set 2	**-1**	-1	**-1**	1	1	-1	-1	-1	1	1	**-1**
variable set 3	**-1**	-1	**-1**	1	1	-1	-1	-1	1	1	**1**
variable set 4	**-1**	-1	**1**	1	1	-1	-1	-1	1	1	**1**

Table 5.5: Four sets of variable settings for optimisation

5.3.2 Optimisation results

After having run the four calculations, the objective function values and the implicit constraint values are calculated. Table 5.6 presents the results. For comparison, the response values of the original modified process are also presented in this table.

The calculation with the best variable settings for the objective function (variable set 1) turns out to be not feasible at all ($g > 0$). However, the second and fourth calculations with slightly different variable settings result in a feasible product as denoted by the negative constraint value. The third calculation is also not feasible, but the constraint value is much smaller than the constraint value for calculation with variable set 1. The second calculation delivers the best results and the optimised variable settings belonging to this calculation are displayed in Table 5.7.

The improvement of the optimised objective function compared to the objective function value of the original modified process, demonstrates that it can be very efficient to optimise an optimi-

	f	g
Values of the optimisation using variable set 1	4.9610	0.6986
Values of the optimisation using variable set 2	0.9097	-0.0048
Values of the optimisation using variable set 3	0.8939	0.0025
Values of the optimisation using variable set 4	0.9147	-0.0181
Values of the original modified process	1.4823	-0.0459

Table 5.6: Optimisation results of the modified process

x_{01}	size	200 mm	x_{07}	ρ_{punch04}	9 mm
x_{02}	t	2.8 mm	x_{08}	ρ_{die16}	8 mm
x_{03}	ρ_{die02}	7.5 mm	x_{09}	$F_{\text{blankholder16}}$	120000 N
x_{04}	ρ_{punch02}	15 mm	x_{10}	$\mu_{\text{die/binder}}$	0.15
x_{05}	$F_{\text{blankholder02}}$	120000 N	x_{11}	μ_{punch}	0.08
x_{06}	ρ_{die04}	6mm			

Table 5.7: Optimised variable settings of the modified process

The modified process

The optimised modified process with variable set 2

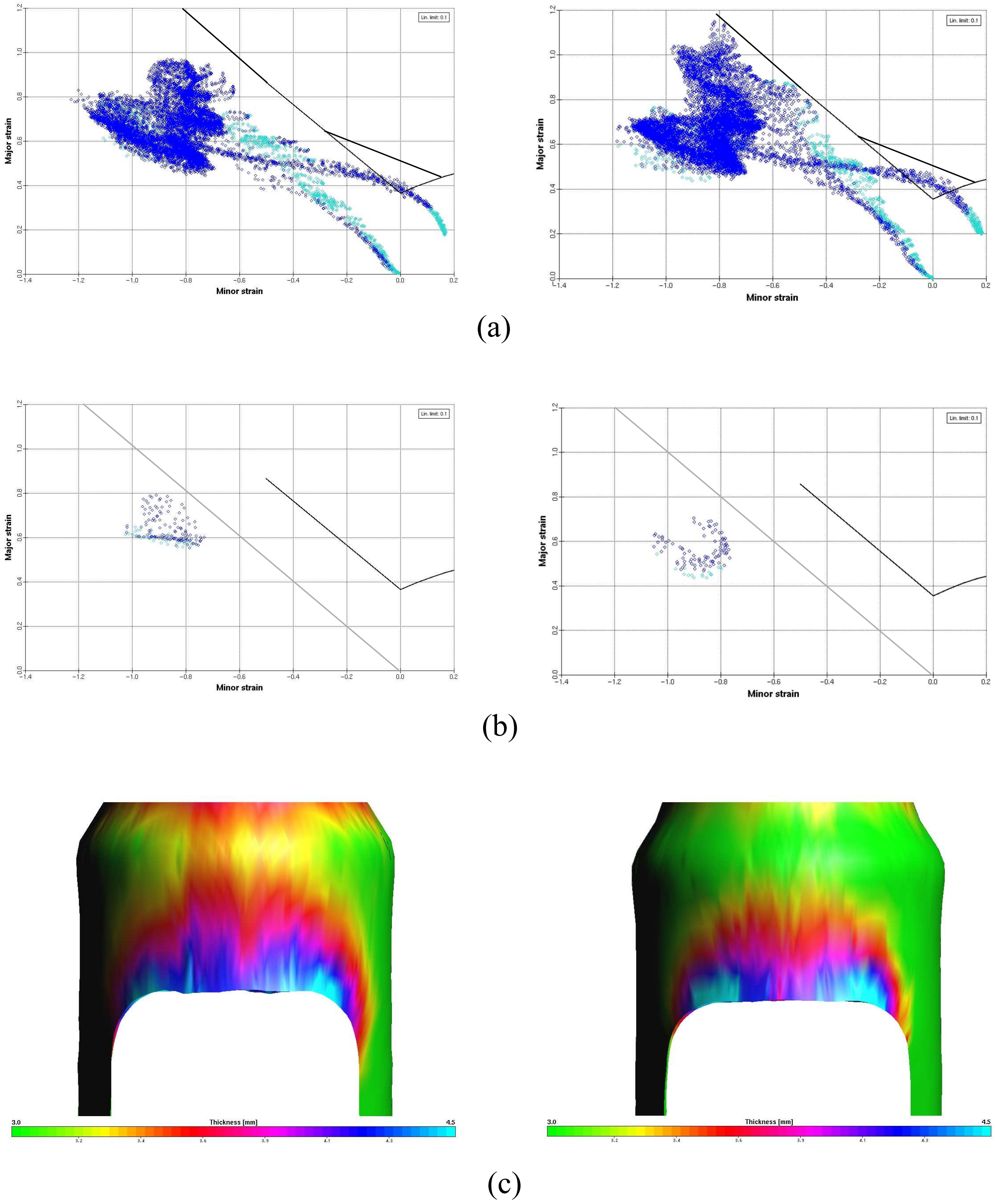
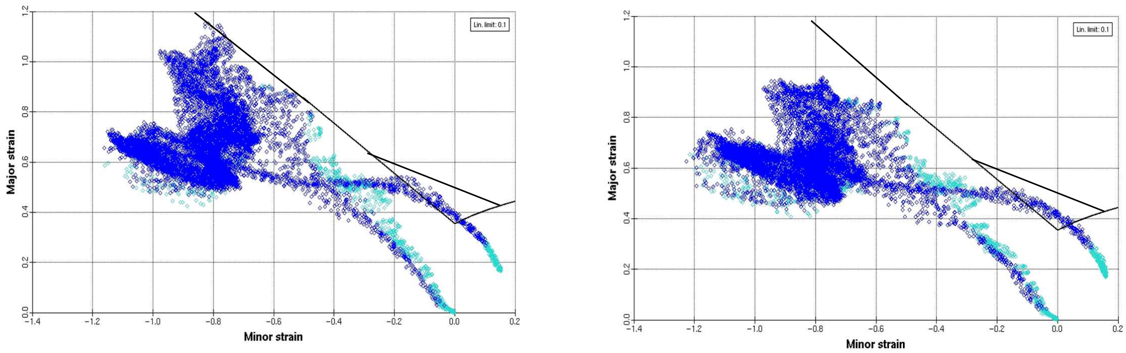


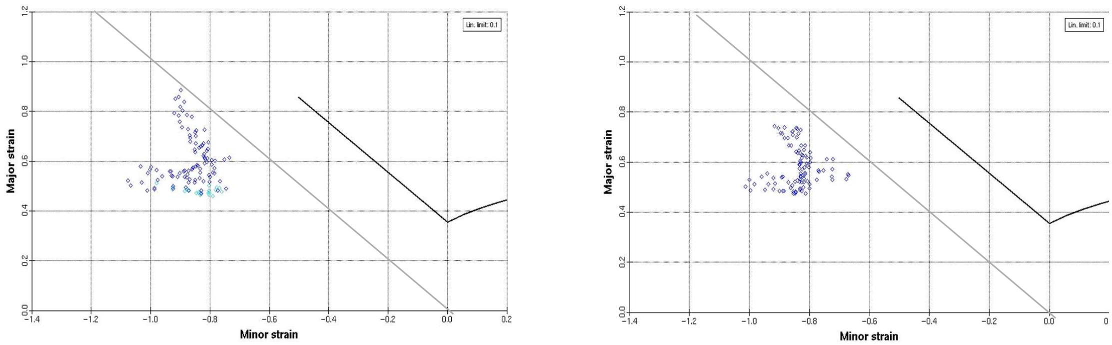
Figure 5.3: (a) The FLD with strains of all the elements (b) the FLD with strains of elements that are located in the critical areas and (c) the thickness for the original modified process and the optimised modified process with variable set 2

The optimised modified process with variable set 3

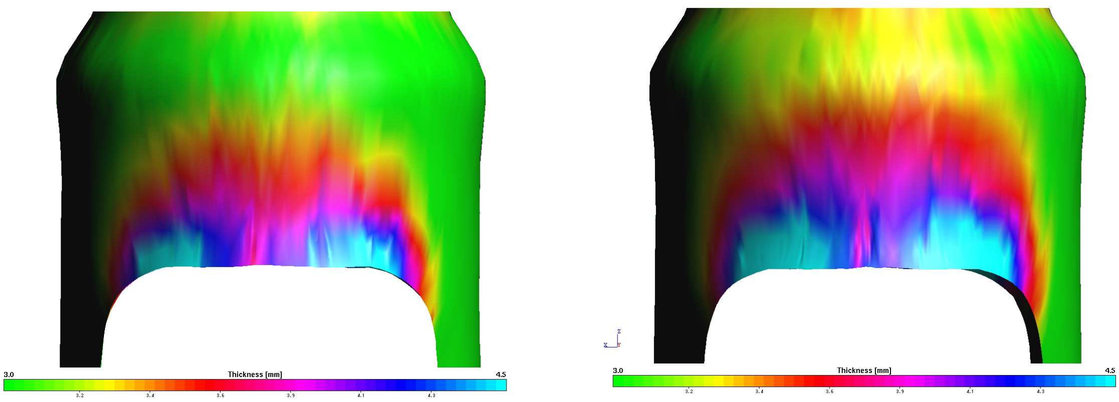
The optimised modified process with variable set 4



(a)



(b)



(c)

Figure 5.4: (a) The FLD with strains of all the elements (b) the FLD with strains of elements that are located in the critical areas and (c) the thickness for the optimised modified process with variable set 3 and 4

sation problem by using optimal variables settings based on the effect plots after screening. Even if these variables not seem to have a large influence on the objective function.

Just as for the reference process the Forming Limit Diagram (FLD) with the strains of all the elements, with strains belonging to the elements in the critical areas and the thickness are presented (see Figures 5.3 and 5.4). These plots are presented for the optimised processes with the variable sets 2, 3 and 4. For comparison, these plots are also presented for the original modified process in Figure 5.3.

Figures 5.3a and 5.4a show that for the original modified process and the optimised process with variable set 2 and 4 all strains stay below the FLC and these processes are thus feasible. There are some strains that exceed the FLC for the optimised process with variable set 3, this corresponds to the positive constraint value g. In Figure 5.3b it can be seen that the major strains in the critical area are reduced for all optimised processes. Minor strains have also reduced, but the smallest minor strains still exist. Furthermore, the plots show that for the optimised process with variable set 2 the major strains belonging to elements with the most extreme minor strains are larger than for the optimised process with variable set 4. Thus, however variable set 2 delivers a lower objective function value (in which the values of all strains in the critical areas are included), it may be that the effect on reducing the occurrence of the cracks is better for variable set 4 than for variable set 2. The thickness in Figures 5.3c and 5.4c shows that after optimisation the thickness is more concentrated and again shifted a bit more from the corners to the middle of the product.

5.4 Conclusions

In this chapter the optimisation strategy was applied to the modified process of the bearing tube. The optimisation model does not differ much from the optimisation model for the reference process as defined in Chapter 4: the mathematical definitions of the responses are equal, only some variables are different. Screening this optimisation model did not result in a reduced optimisation model because none of the variables has a dominating influence on the objective function. After screening, the optimisation strategy was not further used for optimisation.

Optimisation is performed by cleverly choosing a few sets of variable settings based on the screening results. FEM calculations are run for these variable settings. This resulted in an (feasible) objective function value that improved much with respect to the objective function value of the original modified process. This showed that changing the settings of all the variables can

	Modified process	Optimised process with variable set 2	Optimised process with variable set 3	Optimised process with variable set 4
f	1.4823	0.9097	0.8939	0.9147
g	-0.0459	-0.0048	0.0025	-0.0181
x_{01}	204 mm	200 mm	200 mm	200 mm
x_{02}	3.0 mm	2.8 mm	2.8 mm	2.8 mm
x_{03}	9 mm	7.5 mm	7.5 mm	12.5 mm
x_{04}	12.5 mm	15 mm	15 mm	15 mm
x_{05}	100000 N	120000 N	120000 N	120000 N
x_{06}	8 mm	6 mm	6 mm	6 mm
x_{07}	11 mm	9 mm	9 mm	9 mm
x_{08}	10 mm	8 mm	8 mm	8 mm
x_{09}	120000 N	120000	120000	120000
x_{10}	0.1	0.15	0.15	0.15
x_{11}	0.12	0.08	0.15	0.15

Table 5.8: The optimisation results of the modified process

have a large effect on the optimisation results, although separately they not seem to have a large influence on the objective function.

The values of the variables for the modified process with the different variable settings and the response values for these processes are presented in Table 5.8. The strains and thickness of these processes are presented in Figures 5.3 and 5.4.

Chapter 6

Conclusions and recommendations

In this chapter some conclusions are drawn and recommendations for further work are formulated.

6.1 Conclusions

A generally applicable optimisation strategy for metal forming processes has been applied to the deep drawing process of a bearing tube. This strategy includes modelling, screening and solving the optimisation problem. The optimisation strategy as applied in this project is explained and subsequently applied to:

- optimise the reference process to compare these results with the modified process;
- optimise the modified process to further reduce scrap.

Before applying the strategy the cause of the cracks was identified. The assumption was made that the product is subject to critical strains due to large deformation in the last forming step and that these strains give rise to crack occurrence. Also, other failures mechanisms during deep drawing are discussed. This investigation gives insight into which quantities are involved and can subsequently be used for modelling the optimisation problem.

For modelling the optimisation problem a seven step procedure based on the product life cycle was introduced and applied to the deep drawing process of the bearing tube. It was shown that all formerly identified quantities for the deep drawing of the bearing tube are selected using this procedure. This demonstrates the applicability of the procedure to model mathematical optimisation problems of metal forming processes. The procedure also showed to be a straightforward approach to choose which quantities to select as objective function, which as implicit constraint and which as design variables. Defining the objective function and the implicit constraint are important steps, as the application of the optimisation strategy to the reference process showed us that a noisy response without a trend that stands out, is difficult to optimise. The response should be formulated in such a way that it reduces noise as much as possible and that it reveals a trend. Applying a norm to the definition of the responses can lower the response to noise ratio and improve the responses.

Furthermore, the optimisation of the reference process learned us that having knowledge on the shape of the responses is important to model the optimisation problem well. A noisy response function needs to be simplified as much as possible for an efficient optimisation. In case of a noisy response function, screening is a very important step. It should be pointed out which variables can be excluded from the optimisation problem. When numerical noise is present, variables that show to have a small effect on the response, can make the response only more complex without having an effect on the trend of the response.

After having modelled the optimisation problem for the reference and the modified process, both processes have been optimised by applying the optimisation algorithm. This resulted in some conclusions regarding the mathematical optimisation as well as in some conclusions regarding process of the bearing tube.

Optimisation results

- The mathematical optimisation was successful. The optimisations resulted in both cases in improved objective function values. The optimised objective function value for the reference process is much closer to the objective function value of the modified process that reduced occurrence of the cracks than the objective function value of the reference process is. Also, the objective function value of the optimised modified process improved much with respect to the original modified process.
- The responses of the optimisations showed to be subject to numerical noise. Optimisation with RSM and Kriging metamodels delivered more or less equally improved objective function values. Kriging can finally lead to a more accurate value, but much more calculations are needed. Besides that, the question remains if this value is a better value than the optimum obtained by RSM or by Kriging after just a few calculations, because the difference between these values might not or nearly be larger than the error caused by numerical noise. Thus, although Kriging metamodels seem to result in generally more accurate optima, when numerical noise is present, RSM is also an efficient way for optimisation.

- For as well the reference as the modified process, optimisation is applied by using optimal variables settings which are extracted from the effect plots yielded from screening. This proved to be an easy way to improve the optimisation results. This application of cleverly choosing variables settings for the modified process based on screening results showed that in case of complex processes, where many variables are influencing the process but none are dominant, this is an easy and successful way to optimise. With these optimal variable settings the process was significantly improved.

Results for the process of the bearing tube

For the optimisations the FEM programme AutoForm is used. Based on these simulations and the identified cause for the cracks, the optimisations delivered good results. However, the AutoForm simulations did not coincide with reality and with the simulations made in INDEED, so comparison to the modified process as simulated in INDEED and to reality is difficult. Also, the cause of the cracks is still uncertain. But, although the results should be validated, the results showed that there are possible improvements for the process of the bearing tube.

- Applying the optimisation algorithm to the final optimisation problem of the reference process did improve the process. However, the process is not as much improved as the improvement made by the modified process that INPRO proposed. The reference process is improved by reducing the size of the blank (x_{01}) and setting the other variables to their optimal settings as defined in Table 4.14. The effect of the optimisation on the strains and the thickness are displayed in Figures 4.23 and 4.24.

- A second optimisation of the modified process that includes an extra blank holder in the critical deep drawing step, easily reduced the objective function value significantly. This demonstrates how optimisation techniques are useful for solving production problems of already existing metal forming processes. The results of the optimisation are presented in Table 5.8 and in Figures 5.3 and 5.4.

6.2 Recommendations

Three recommendation are made with respect to the process of the bearing tube:

1. Because the AutoForm simulations did not coincide with the simulations in INDEED and with reality and because the cause of the cracks is still uncertain, the results should be validated to check if the optimised process indeed reduces the cracks in reality. It should better be validated if the results of the AutoForm simulations coincide with reality or not and if the definition of the objective function is representative. For the latter, it needs to be validated if reducing the strains and the thickness are related to a reduction of the occurrence of the cracks.
2. To improve the reference process of the bearing tube it is recommended to reduce the blank size (x_{01}) and to set the other variables to their optimal settings as defined in Table 4.14. If the preference is to keep the variables to their original settings, to adjust the process a less a possible, it is recommended to only set the rounding of the die of deep drawing step OP02 variable x_{03} to its minimum value of 7.5 mm (besides of course reducing the blank size). After analysis of the responses, this is the only variable, next to the size, that was shown to have effect on the response.
3. It is recommended to change the variable settings for the modified process that now exist, to the the optimised values as presented in Table 5.8.

Based on the AutoForm simulations these adjustments will reduce the large strains and thickness in the critical areas. It is expected that this will further reduce the occurrence of the cracks. However, this needs to be validated in reality.

Also, on the basis of the conclusions, four recommendations are made to the optimisation strategy that is applied to the optimisation of the bearing tube:

1. It is useful to get more insight in the shapes and the complexity of response functions. Identifying the presence and magnitude of numerical noise is also of importance. Obtaining this knowledge for different optimisation problems, is an important step for further improving the optimisation strategy for metal forming processes.
2. If more knowledge on the responses is obtained, a subsequent investigation can prove which metamodels are best to be applied in which cases: RSM or DACE metamodels.
3. It can be helpful, when more knowledge on the responses is gained, to convert this knowledge to a more extended approach for formulating the mathematical definition of the responses. In this approach, the importance of good definitions and possibilities for formulating the objective function and implicit constraints to reduce noise can be discussed.
4. It is recommended to extend the screening step of the optimisation strategy. Taking too many variables into account can make the responses unnecessarily complex. The screening step should be elaborated in such a way that after screening it is clear which variables are suitable to be included in the optimisation problem. For this purpose, the importance of the error displayed in the screening plots can be further investigated. Also, the option to use optimal variables settings following from screening should be included in this step.

Appendix A

Analysis by INPRO

INPRO simulated the deep drawing process of the bearing tube. The simulation showed how the thickness increases in the critical area where the cracks appear and that in this area the major and minor strains have some extreme values after the last deep drawing step OP18 (see Figure A.1 and Figure A.2).

On the basis of this simulation INPRO concluded that when the thickness could be reduced

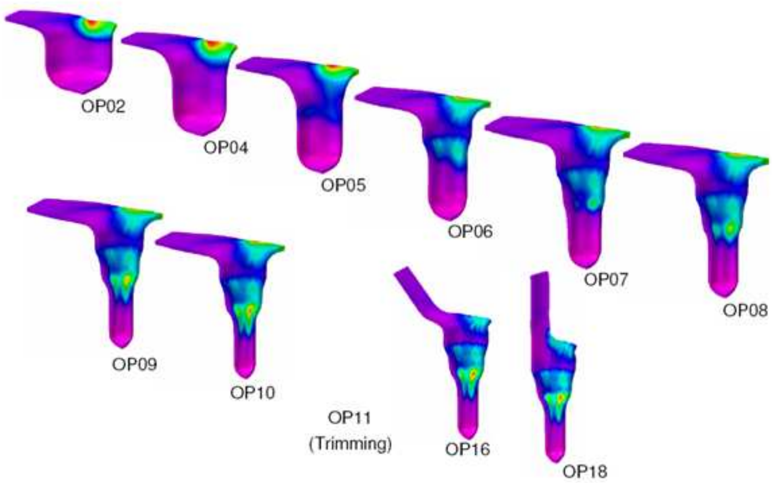


Figure A.1: The thickness simulated by INDEED

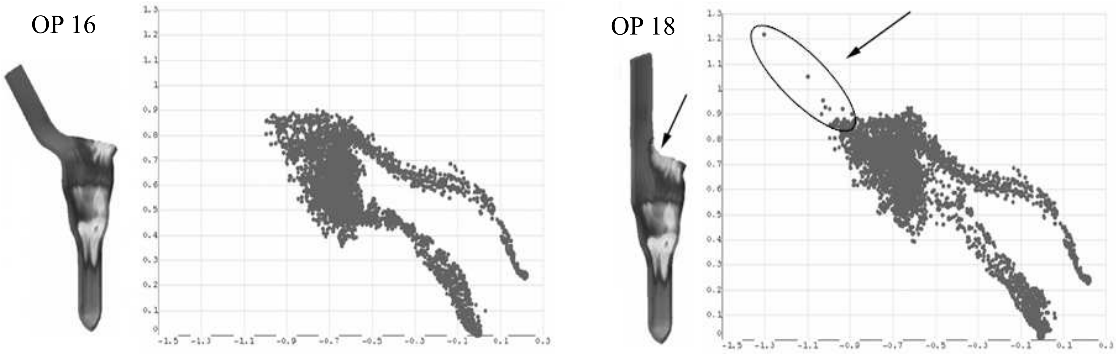


Figure A.2: Strains in the critical area the process would be more stable. Therefore INPRO thought of some adjustments which reduce the thickness and simulated the process with these adjustments to look how these adjustments reduce the minor and major strains. The adjustments made, were:

- different shapes of the blank;
- new trimming shapes;
- use of a conical blank holder in OP02 and in the following OPs;
- adding an extra deep drawing step with a blank holder before OP18.

The effect of these adjustments are shown in the Figures A.3 and A.4. The adjustment with the extra deep drawing step was most successful and is implemented in the process of the production of the parts in the factory. This adjustment reduced the scrap significantly.

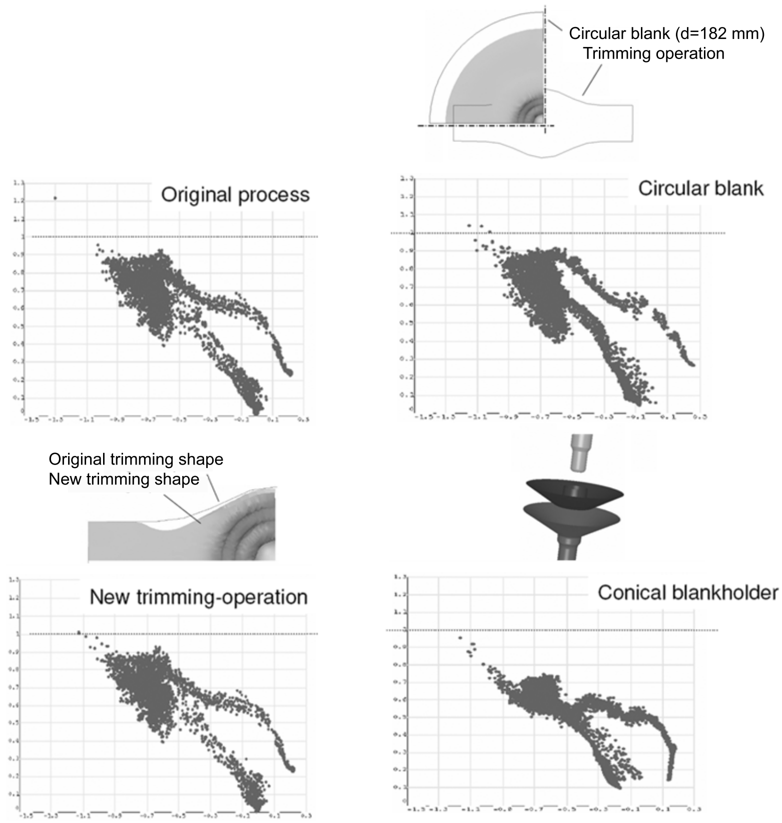


Figure A.3: The influence of the adjustments on the major and minor strains

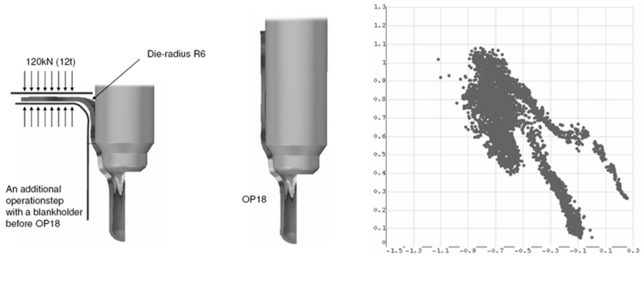


Figure A.4: The influence on the strains by adding an extra deep drawing step

Appendix B

Material properties

The material properties of the material which is used by Fischer & Kaufmann for the production of the bearing tubes can be found in Figure B.1.

Abnahmeprüfzeugnis 3.1 nach EN 10204
Inspection certificate 3.1 according to EN 10204

Nr. / No. 17390 L vom / of 28.03.06

Bestellung Nr., vom /your Order-No., dated	Auftrags-Nr. /our Order-No.	Abmessung /Dimension
03.09.03	48/173404/01/06	196,00 . 3,00

Besteller /Customer	Lieferzustand /Condition of Delivery
FISCHER & KAUFMANN GMBH &CO KG FISCHER U KAUFM FINN	WALZEN BEIZEN OELEN
	Erzeugnisform /Discrition Part WARMBAND
Bestellung Nr., vom /Order-No., dated 03.09.03	Werkstoff-Normbezeichnung /Standard Grade of Material DD 13
Lieferbedingung /Material Specifications	Material-Nr./Material No. 14179

Lieferschein Nr., vom / Delivery note no., dated	Transport-Nr. / No. of unit	Stück / Piece	Masse /Mass kg	Walzfolge Nr. / Sequence of Rolling	Schmelze Nr. / Heat No.
45053 28.03.06		2	3680	46482	068656

Chemische Zusammensetzung /Chemical Composition

C	Si	Mn	P	S	Al	Cr	Cu	V
0,028	0,007	0,220	,0100	,0020	0,060	0,016	0,012	0,001
N	Ni	Mo						
0032	0,019	0,002						

Prüfergebnisse /Test Results

Probe Nr. / Test Piece No.	Zugversuch / Tensile Test: Proben-lage 1) / Pos. of sample 1)	R_{eL} N/mm²	R_m N/mm²	A_5 %	$A_{L=80}$ %
5564	L	231	345	48,7	38,3

Figure B.1: Material properties

Bibliography

[1] *Dieptrekken (VM 110)*. FME, The Netherlands.

[2] *Metallurgist*. Personal communication by e-mail.

[3] Norm EN 10111:1998 - continuously hot-rolled low carbon steel sheet and strip for cold bending - technical delivery conditions.

[4] *Metallurgists*. Personal commnication in a meeting on March 20$^{\text{th}}$, IJmuiden, The Netherlands, 2006.

[5] AVITZUR, B. *Handbook of Metal-Forming Processes*. John Wiley & Sons, Inc., New York, USA, 1983. ISBN 0-471-03474-6.

[6] BONTE, M., DO, T., FOURMENT, L., VAN DEN BOOGAARD, A., HUÉTINK, J., AND HABBAL, A. A comparison of optimisation algorithms for metal forming processes. In *Proceedings of ESAFORM* (Glasgow, UK, 2006), pp. 883–886.

[7] BONTE, M., VAN DEN BOOGAARD, A., AND HUÉTINK, J. Metamodelling techniques for the optimisation of metal forming processes. In *Proceedings of ESAFORM* (Cluj-Napoca, Romania, 2005), pp. 155–158.

[8] BONTE, M., VAN DEN BOOGAARD, A., AND HUÉTINK, J. Solving optimisation problems in metal forming using finite element simulation and metamodelling techniques. In *Proceedings of APOMAT* (Morschach, Switzerland, 2005), pp. 242–251.

[9] BONTE, M., VAN DEN BOOGAARD, A., AND HUÉTINK, J. A metamodel based optimisation algorithm for metal forming processes. *Accepted for: International Journal of Forming Processes* (2006).

[10] BONTE, M., VAN DEN BOOGAARD, A., AND HUÉTINK, J. An optimisation strategy for industrial metal forming processes. *Submitted to: Structural and Multidisciplinary Optimization* (2006).

[11] BONTE, M. H. A. Solving optimisation problems in metal forming using FEM: A metamodel based optimisation algorithm. Internal report P05.1.059, Netherlands Institute for Metals Research (NIMR), Delft, The Netherlands, 2005.

[12] CALLISTER, W. *Materials Science and Engineering*, 4$^{\text{th}}$ ed. John Wiley & Sons, Inc., New York, USA, 1997. ISBN 0-471-13459-7.

[13] HOESCH HOHENLIMBURG GMBH. Material data DD13. Hagen, Germany.

[14] HOSFORD, W. F., AND CADDELL, R. M. *Metal forming: Mechanics and metallurgy*, 2$^{\text{th}}$ ed. Prentice-Hall, Inc., Englewood Cliffs, USA, 1993. ISBN 0-13-588526-4.

[15] HUÉTINK, J., AND MEINDERS, T. Solid mechanics II, Plasticity. Lecture notes University of Twente, Enschede, The Netherlands, 2005.

[16] M. H. A. BONTE. A comparison between optimisation algorithms for metal forming processes – with application to forging. Internal report CTW.05/TM-5513, University of Twente/Ecole Nationale Supérieure des Mines de Paris, Enschede, Netherlands, 2005.

[17] MASCHKE, Y. Personal communication by e-mail.

[18] MASCHKE, Y. Presentation held at INPRO on February 15$^{\text{th}}$. Berlin, Germany, 2006.

[19] MATHES, M., AND HECHT, U. Schadensfall wst24: Rissbildung an der stanzkante. Internal report, Access Materials + Processes and FIuKA GmbH, Finnentrop, Germany, 2006. In German.

[20] MUSSER, K. M., AND VERLOOP, W. C. Testing cold work embrittlement behaviour: A Round Robin. *Society of Automotive Engineers 021BECB-17* (2001).

[21] MYERS, R., AND MONTGOMERY, D. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, 2^{nd} ed. John Wiley & Sons, Inc., New York, USA, 2002. ISBN 0-471-41255-4.

[22] NITSCHE, M. Presentation held at INPRO on February 15^{th}. Berlin, Germany, 2006.

[23] PEARCE, R. *Sheet Metal Forming*. Adam Hilger, Bristol, England, 1991. ISBN 0-7503-0101-5.

[24] RAMAEKERS, J. A. H., AND PEETERS, P. B. G. *Omvormen van metalen*. Academic Service, Schoonhoven, The Netherlands, 1998. ISBN 90 395 0579 9 (In Dutch).

[25] REINHOLD, P. Querlenker PQ 35 prozess optimierung. Internal Report K-GQL-2/1, 1437/2, Volkswagen AG - Konzern-Qualitätssicherung, Germany, 2006. In German.

[26] SANTNER, T., WILLIAMS, B., AND NOTZ, W. *The Design and Analysis of Computer Experiments*. Springer-Verlag, New York, USA, 2003.

[27] SARRAK, V. I., AND FILIPPOV, G. A. Delayed fracture, hydrogen, and impurities. *Metal Science and Heat Treatment 19*, 8 (1977), 656–661.

[28] SCHONLAU, M. *Computer Experiments and Global Optimization*. PhD thesis, University of Waterloo, Waterloo ON, Canada, 1997.

[29] SCHRÖDER, M., AND MÜLLER, J. Personal communication. Finnentrop, Germany.

[30] THE MATHWORKS, INC. MATLAB Optimization Toolbox User Guide.