

# **Continuous Cast Width Prediction Using a Data Mining Approach**

*By*

*Petrus Gerhardus de Beer*

Submitted in partial fulfilment of the requirements for the degree of

**Master of Engineering (Mechanical)**  
in the  
**Faculty of Engineering, Built Environment and  
Information Technology**  
at the  
**University of Pretoria  
South Africa**

November 2006

## **Summary**

### **Continuous Cast Width Prediction Using a Data Mining Approach.**

**Compiled by** : Petrus Gerhardus de Beer

**Student number** : 2333715

**Study leader** : Prof. K. J. Craig

### **Abstract**

In modern times continuous casting is the preferred way to convert molten steel into solid forms to enable further processing. At Columbus Stainless the continuous casting machine cast slabs of constant thickness with varying width. One important aspect of the continuously cast strand that must be controlled, is the strand width. The strand width exiting from the casting machine, has a direct influence on the product yield which in turn influences the profitability of the company. In general, the strand width control on the austenitic and ferritic type steels achieved is excellent with the exception of the 12% chrome non stabilised ferritic steel. This steel type exhibited different strand width changes when a sequence of different heats was cast. The strand width changes corresponded to the different heats in the sequence. Each heat has a unique chemistry and a relationship between the austenite and ferrite fraction at high temperature and the resulting strand width change was explained by Siyasiya<sup>[27]</sup>. The relationship between the heat composition and width change has in the past resulted in the development of a model that enabled the prediction of the expected width change of a specific heat before it is cast to enable preventative action to be taken. This model has been implemented as an on-line prediction model in the production environment with very encouraging results. This study was initiated

because it was uncertain if the implemented model was the most accurate for this application. This study is concerned with the development of more models based on different techniques in an attempt to implement a more accurate model. The data mining techniques used include statistical regression, decision trees and fuzzy logic. The results indicated that the existing model was the most accurate and it could not be improved upon.

**Key words:** Continuous Casting, Stainless Steel, Strand width Control, Statistical Regression, Decision Trees, Fuzzy Logic, Rule Based Model, Width Change.

## **Acknowledgements**

1. I would like to thank the Technical Hot Products Manager of Columbus Stainless Mrs Alison Sutcliffe for allowing me the opportunity and freedom to start and implement this work in the production environment.
2. I would like to thank Mr Hans Kruger for converting the model to C programming language and implementing it on the control systems of the continuous casting machine.
3. A big thank you to the management at Columbus Stainless for allowing me the opportunity to present the results of the plant model at the European Continuous Casting Conference held in Nice, June 2005.
4. A word of thanks to Professor Ken Craig from the University of Pretoria for the guidance to make this project a success.
5. A very special word of thanks to my wife for all the moral support and encouragement during the duration of this project.
6. Lastly to my Heavenly Father for granting me the opportunity and ability to finish this project.

## Table of Contents

### **Chapter 1: Introduction**

1.1 Background.....	1
1.2 General Stainless Steelmaking Principles .....	3
1.2.1 Refining .....	4
1.2.2 Chemistry of 12% chrome ferritic non-stabilised stainless steel.....	6
1.2.3 Continuous casting of steel.....	7
1.2.3.1 Mould heat transfer.....	8
1.2.3.2 Secondary cooling .....	9
1.2.3.3 Mould set-up.....	11
1.3 Objectives of this project.....	11
1.4 Method of investigation.....	12
1.5 Outline of dissertation .....	12

### **Chapter 2: Literature Survey**

2.1 Continuous Cast Width Change .....	14
2.2 Fundamentals of the chemistry to cast width error relationship.....	16
2.2.1 Theoretical fundamentals .....	17
2.2.2 MEDUSA model. ....	22
2.3 Proposed mechanism for the chemistry to cast width relationship .....	23
2.4 Description of the model being used in the production environment .....	25
2.5 Data mining .....	27
2.5.1 Introduction .....	27
2.5.2 Statistical Analysis .....	32
2.5.2.1 Computationally simple descriptive statistics .....	33
2.5.2.1.1 Dot Diagrams.....	34
2.5.2.1.2 Stem-and-leaf plots.....	35
2.5.2.1.3 Frequency diagrams and histograms .....	35
2.5.2.1.4 Scatterplots and Run charts .....	36
2.5.2.2 Computationally intensive descriptive statistics .....	36
2.5.2.2.1 Fitting a straight line using least squares approximations....	37
2.5.2.2.2 Fitting surfaces using least squares approximations .....	39
2.5.3 Decisions Trees .....	40
2.5.3.1 Introduction .....	40
2.5.3.2 Advantages and disadvantages of decision trees.....	42
2.5.3.3 Design of decision trees.....	44
2.5.3.4 Performance measures for decision trees .....	46
2.5.4 Fuzzy logic .....	46
2.5.4.1 Introduction .....	46
2.5.4.1.1 Fuzzy Rule base.....	50
2.5.4.1.2 Fuzzy Inference Engine .....	51
2.5.4.1.3 Fuzzifiers .....	52
2.5.4.1.3.1 Singleton Fuzzifier .....	52
2.5.4.1.3.2 Gaussian Fuzzifier .....	52
2.5.4.1.3.3 Triangular Fuzzifier.....	52
2.5.4.1.4 Defuzzifiers .....	53
2.5.4.1.4.1 Centre of Gravity Defuzzifier.....	53
2.5.4.1.4.2 Centre Average Defuzzifier.....	53
2.5.4.1.4.3 Maximum Defuzzifier .....	54
2.5.4.2 Membership Functions .....	54
2.5.4.3 Fuzzy Set Operations.....	55

2.5.4.3.1 Union .....	55
2.5.4.3.2 Intersection .....	56
2.5.4.3.3 Complement .....	56
2.6 Summary of literature survey .....	56
<b>3. Data preparation and modelling</b>	
3.1 Gathering the data.....	58
3.2 Preparing the data for modelling .....	60
3.3 Data clean-up.....	61
3.4 Selection of significant variables.....	62
3.5 Visual representation of the data .....	63
3.6 Model evaluation criteria .....	68
3.7 Modelling using statistical regression .....	69
3.8 Modelling using a regression decision tree .....	74
3.9 Modelling using Fuzzy logic .....	78
3.9.1 Fuzzy logic model based on triangular membership functions .....	78
3.9.2 Fuzzy logic model based on polynomial membership functions .....	82
3.10 Summary of the results of the derived models .....	85
<b>4. Evaluation of derived models on a validation data set</b>	
4.1 Application of the derived statistical model on the validation data set .....	87
4.2 Application of the derived decision tree model on the validation data set.....	89
4.3 Application of the derived fuzzy logic model on the validation data set .....	91
4.4 Application of the model currently used on the validation data set .....	93
4.5 Summary of the results obtained by applying the models to the validation data .....	95
<b>5. Plant results from model best suited for this application</b>	
5.1 Introduction .....	97
5.2 Implementation of the model in the production environment .....	97
5.3 Plant results from the implemented model .....	99
5.4 Summary of plant results .....	104
<b>6. Conclusions and Recommendations</b>	
6.1 Conclusions .....	106
6.2 Recommendations for future work .....	107
<b>References</b> .....	109
<b>Appendix A</b> : Schematic overview of the mould set-up relationship to aim cast width on 12% Chrome non stabilised steel. ....	A
<b>Appendix B</b> : Relationship between the ferrite factor and the measured austenite start temperature $Ar_5$ for 3CR12 .....	B
<b>Appendix C</b> : Equations used for metallurgical parameters .....	C
<b>Appendix D</b> : Example of Aspen graph representing laser measurements.....	E
<b>Appendix E</b> : Response surface equations.....	F
<b>Appendix F</b> : Polynomials for fuzzy logic model .....	G
<b>Appendix G</b> : Raw laser data gathering and representation.....	H
<b>Appendix H</b> : Training and validation data set .....	J

## List of Tables

<b>Table 1:</b> Survey of data mining techniques .....	28
<b>Table 2:</b> Correlation coefficients between parameters and width error .....	62
<b>Table 3:</b> Results across the width error range from the different response surfaces .....	74
<b>Table 4:</b> Results across the width error range from the decision tree model as applied to the training data set.....	78
<b>Table 5:</b> Results across the width error range from the fuzzy logic model as applied to the training data set .....	82
<b>Table 6:</b> Correlation coefficients for the different degree of polynomials and input parameters.....	83
<b>Table 7:</b> Accuracy across the width range as obtained with the average of polynomials .....	85
<b>Table 8:</b> Summary table of the results obtained with the different models on the training set.....	86
<b>Table 9:</b> Summary of the accuracy obtained with the surface response models as applied to the validation set .....	89
<b>Table 10:</b> Accuracy obtained per width error group with the decision tree model .....	91
<b>Table 11:</b> Accuracy obtained per width error group with the pruned decision tree model.....	91
<b>Table 12:</b> Accuracy obtained per width error group with the fuzzy logic model based on triangular membership functions.....	92
<b>Table 13:</b> Accuracy obtained per width error group with the fuzzy logic model based on polynomials and applied to the validation data set.....	93
<b>Table 14:</b> Results of the current model as applied to the validation set .....	94
<b>Table 15:</b> Combined results of the different models as achieved on the validation data set .....	96
<b>Table 16:</b> Summary of reference population and model active population.....	101
<b>Table 17:</b> Summary of the populations with and without secondary cooling modifications .....	103

## List of Figures

<b>Figure 1:</b> Schematic representation of the Columbus Stainless Steelplant .....	4
<b>Figure 2:</b> Schematic representation of the continuous casting machine at Columbus Stainless .....	8
<b>Figure 3:</b> Typical CCM temperature profiles required for secondary cooling control at Columbus Stainless .....	10
<b>Figure 4a:</b> Fe-Cr Binary phase diagram .....	18
<b>Figure 4b:</b> Effect of carbon content on the austenite/ferrite ratio .....	19
<b>Figure 5:</b> CCT diagram for 12% chrome non-stabilised ferritic stainless steel .....	19
<b>Figure 6:</b> Slabs measured in plant to indicate width error mechanism .....	25
<b>Figure 7:</b> Box plot indicating outliers by using IQR (Interquartile range).....	30
<b>Figure 8:</b> Sample dot diagram .....	34
<b>Figure 9:</b> Stem-and-leaf plot of 100m sprint times .....	35
<b>Figure 10:</b> Sample histogram .....	36
<b>Figure 11:</b> Typical structure of a decision tree .....	41
<b>Figure 12:</b> Bivalent sets to characterize the strand width error .....	49
<b>Figure 13:</b> Triangular membership functions .....	50
<b>Figure 14:</b> Relationship of AC1 temperature to strand width error .....	64
<b>Figure 15:</b> Relationship of Amax to strand width error .....	64
<b>Figure 16:</b> Relationship of CR95 to strand width error.....	65
<b>Figure 17:</b> Relationship of Gamma max to strand width error.....	66
<b>Figure 18:</b> Relationship of ferrite factor to strand width error .....	67
<b>Figure 19:</b> Relationship of Carbon and Nitrogen to strand width error .....	68
<b>Figure 20:</b> Histograms of parameter values of the training data set.....	70
<b>Figure 21:</b> Training data set with predicted values using statistical regression model.....	72
<b>Figure 22:</b> Linear and interactive term surface versus actual values .....	72
<b>Figure 23:</b> Linear and quadratic term surface versus actual values .....	73
<b>Figure 24:</b> Full quadratic surface versus actual values.....	73
<b>Figure 25:</b> Un-pruned version of regression tree .....	75
<b>Figure 26:</b> Resubstitution and cross-validation error to estimate best tree size .....	76
<b>Figure 27:</b> Pruned version of decision tree.....	77



**Figure 28:** Actual width error and predicted width error using pruned decision tree ...77

**Figure 29:** Typical membership allocation using triangular membership functions .....79

**Figure 30:** Typical membership functions as defined across a parameter range.....80

**Figure 31:** Triangular membership functions of the six input parameters to the fuzzy logic model .....81

**Figure 32:** Predicted versus actual values as obtained from the polynomials .....84

**Figure 33:** Average of the predicted values from the six polynomials versus the actual width error .....84

**Figure 34:** Linear response surface model applied to the validation data set.....87

**Figure 35:** Response surface models applied to the validation data set .....88

**Figure 36:** Decision tree model results as applied to the validation data set.....90

**Figure 37:** Pruned decision tree model results as applied to the validation data set .....90

**Figure 38:** Predicted values from the fuzzy logic based model using polynomials versus standardized width error values .....92

**Figure 39:** Model decision making process as implemented in the plant .....98

**Figure 40:** Long-term width error trend of 12% Chrome non-stabilised ferritic stainless steel .....99

**Figure 41:** Normal distribution comparison of the three populations under study .....102

**Figure 42:** Normal distribution comparison around zero mean .....102

## Nomenclature

### **CCM**

- Continuous Casting Machine

### **Level 1**

- Control systems on a process consisting out of the measurement equipment and PLC's (Programmable Logic Circuit).

### **Level 2**

- Computer control systems on a process using level 1 system as input.

### **CCT**

- Continuous Cooling Transformation

### **ABS**

- Argon bubbling station

### **Width Error**

- The width error is defined as the difference between the actual cold width of a slab and the aim cold width.

$$\text{Width error} = \text{Actual cold width width} - \text{Aim cast cold width}$$

If the width error is positive it means the slab is wide from aim, conversely, if the width error is negative it means the slab is narrow from aim.

# Chapter 1

## Introduction

### 1.1 Background

With the rapid development of computer technologies during the last decades many industrial processes have benefited from ever-increasing automation of process control systems. Process modelling is a very important part of modern day process control that enables a “virtual” process view where scenarios can be run “off-line” reducing the time to develop new technology and also reducing the cost of implementing new technologies. A very important part of process modelling is prediction. Process models are used in a wide variety of industries as predictive tools in order to improve process control. Predictive process models are usually derived from theoretical engineering principles for example fluid mechanics and thermodynamics. Most industrial processes are however very complex and the process models usually range from being 100% based on theory to being 100% derived from empirical data. The latter usually arises when no theory is available to describe the process adequately or critical parameters needed as input for theoretical models are not available for the process.

Of particular interest to this study are the process models concerned with continuous casting of stainless steel and the effect on cast strand width control. Strand width control is important because it has a direct influence on the product yield which influences the profitability of the company directly. The desired cast width is a function of the ordered final width combined with the effects of all the processes between the continuous caster and final processing lines. The typical route for a cold rolled product in a stainless steel manufacturing plant for example, Columbus Stainless, is continuous casting, hot rolling, cold rolling, final annealing and final lines where the product is cut into ordered sizes.

The two processes that have the biggest direct influence on strand/strip width are the continuous caster and hot rolling mill. The CCM (Continuous Casting Machine) is the first process where the molten steel is shaped into a predetermined dimension and consequently plays a critical role in the resulting strand width. Instabilities in the cast strand width can also lead to instabilities during hot rolling of the slab. The hot rolling mill can influence the strip width by the edging strategy while rolling or by too high tensions that could lead to “necking” of the material that will result in the material being under width. At the final lines the cold rolled material is trimmed to ordered size. If the strip is too wide, too much material must be trimmed off which influences the product yield negatively. If the material is too narrow, it cannot be trimmed to ordered size and must be allocated to narrower ordered widths. When the material is then trimmed to width the result is again excessive trimming losses.

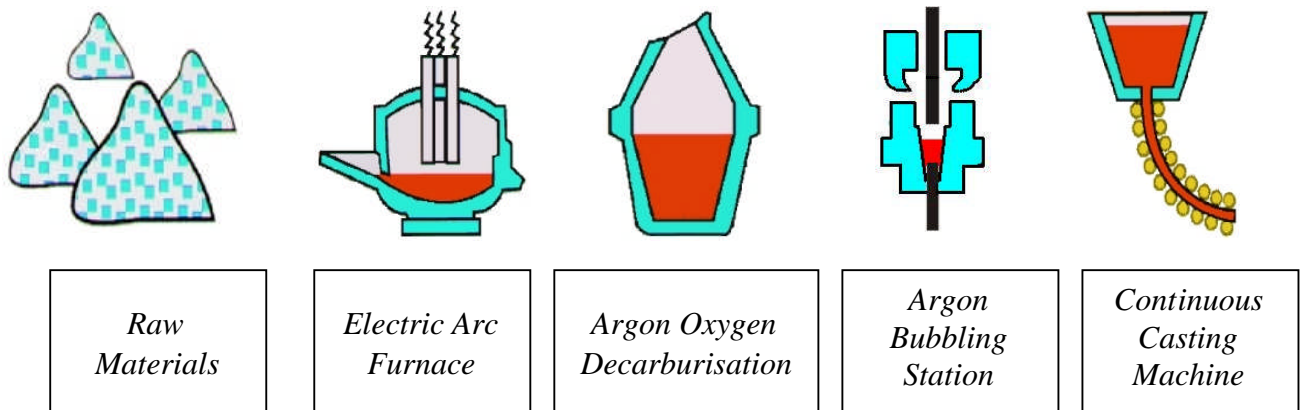
The CCM at Columbus Stainless utilizes a finite element model that calculates the cooling tempo required per strand position to maintain a specified strand temperature profile under varying operational conditions. A lot of work has been done at Columbus Stainless to minimise strand width changes from the aim cast width and in general most steel grades at Columbus Stainless has excellent cast strand width control and stability, except for the 12% Chrome ferritic non-stabilised material. This type indicated different width errors for different heats cast under constant conditions. An investigation revealed a strong relationship between the width error and the heat composition. More specifically a strong relationship between the ferrite and austenite fraction of a heat in the region 850°C to 1250°C and the resulting width error was found. The transformation behaviour and hot strength of this steel type during the continuous casting process has been studied in detail by Siyasiya et al<sup>[26,27,28]</sup>. They also came to the conclusion that the strand width

error can be influenced by the ratio between austenite and ferrite in the temperature range experienced in the continuous casting process. If the relationship can be determined accurately, it will make it possible to predict a certain width error before the heat is cast. If the predictions are accurate, counter measures can be put in place to compensate for the expected width error. A model is currently in use, using rules based on classic set theory incorporating “if-then” type statements. The success of this model serves as an indication that it is possible to predict the width change of 12% chrome, non-stabilised ferritic stainless steels before a heat is cast by using only the heat composition as input parameter base. The uncertainty whether the current model is the best for this specific application initiated this present study.

## **1.2 General Stainless Steelmaking Principles**

The continuous caster is usually the last process in a modern steel melting and refining facility. Three basic processing principles are part of a stainless steel melting and refining plant. The first is the melting of the raw materials, the second is the refining of the molten metal, by changing the chemistry of the raw material melt to the desired chemistries for the different types of stainless steels, and lastly the molten metal is cast into a specific shape using a continuous casting machine. At Columbus Stainless an arc furnace is used to melt the raw materials. The refining is done via an AOD (Argon-Oxygen Decarburization) converter and ABS (Argon Bubbling Station) and the continuous caster cast slabs of constant thickness (200mm) with width varying between 920mm to 1575mm. Figure 1 gives a schematic view of the Columbus Stainless steel plant.

***Figure 1: Schematic representation of the Columbus Stainless Steel plant***



### **1.2.1 Refining**

The refining basics were adapted from Meyn<sup>[15]</sup>. The part of the refining process that is of interest to the present study is the ABS. The ABS is the last station where changes to the chemistry can be made. The melt is tapped from the converter into a teeming (casting) ladle. The ladle is equipped with a slide gate bottom tap hole and a porous plug arrangement for argon purging.

At the rinse station, final adjustments in terms of chemistry and steel temperature are carried out. Major movements in the steel analysis are not desirable because of a risk of insufficient dissolution and mixing of added materials. This late in the process the alloys have to be of low carbon content, unless re-carburisation is required due to a lack of converter process control. The risk of introducing impurities into the clean steel in the form of oxides (corroded scrap) or moisture, needs to be managed. Many steel makers operate ladle furnaces to keep melts hot and even raise ladle temperatures and perform steel analysis adjustments within a much wider scope for production planning.

At Columbus Stainless there is no ladle furnace facility and all heats have to be tapped superheated. With the ladle refractory lining being kept hot at 1200 degrees Celsius when empty, there is a certain flow of energy from the steel into the lining. Through stirring with argon, temperature homogenisation is achieved resulting in an overall drop in metal temperature. For the casting process very stable temperature behaviour of the steel is essential (steel quality implication) to maintain casting parameters, such as casting speed at constant level.

The fully prepared teeming ladle has to be delivered to the casting deck, sometimes within a fairly narrow time window when sequences are being cast. When the required temperature-drop during the rinse treatment cannot be obtained by stirring only, the operator has the choice of adding (at final stainless steel specification) coolants in the form of granules.

The final rinse treatment period is reserved for a very gentle stir to float out non-metallics. These oxides from stirred-in slag or re-oxidation products will show up as inclusions in the final product. For their successful removal the slag has to be manipulated to have oxide-absorbing (scavenging) capabilities. Those non-metallics that cannot be floated out, can be modified by calcium-silicon wire injection treatment to solidify as globules (small round inclusions) finely dispersed in the steel. Through this technique, long clusters of oxides can be avoided that are harmful to mechanical properties, such as crack propagation etc.

Consequently the chemistry leaving the ABS can be assumed to be the same at the CCM since no chemistry changes are done at the CCM. This makes it possible to use the final

chemistry at the ABS as input chemistry to a model that can predict the width error of a heat before the heat is cast.

### 1.2.2 Chemistry of 12% Chrome ferritic non-stabilised stainless steel

The typical chemical composition of the stainless steel applicable to this study is the following:

% C	% Si	% Mn	% P	% S	% Cr	% Ni
0.030	1.00	2.00	0.040	0.030	10.50	1.50
max	max	max	max	max	12.50	max

The steel under review is not stabilised. This means that no stabilisation elements has been added. Sensitisation occurs when chromium carbides precipitates on the grain boundaries causing adjacent chrome depleted areas. In austenitic stainless steels the sensitisation temperature range is between 425°C to 815°C<sup>[42]</sup>. Above 815°C the chromium carbides are soluble and below 425°C the diffusion rate of carbon is too slow to form carbides. Ferritic stainless steel only sensitise after heating to more than 925°C<sup>[42]</sup> where the solubility of carbon and nitrogen in ferrite becomes significant. The low solubility of interstitials in ferrite causes ferritic stainless steels to sensitise more rapidly and at lower temperatures than austenitic stainless steels. The method to prevent sensitisation is to add elements with a great affinity for carbon in order to form stable carbides and doing so removing carbon from the matrix to prevent chromium carbides. The principle elements used for stabilisation is titanium and niobium<sup>[44]</sup>. The material under review in this study has no stabilisation elements.

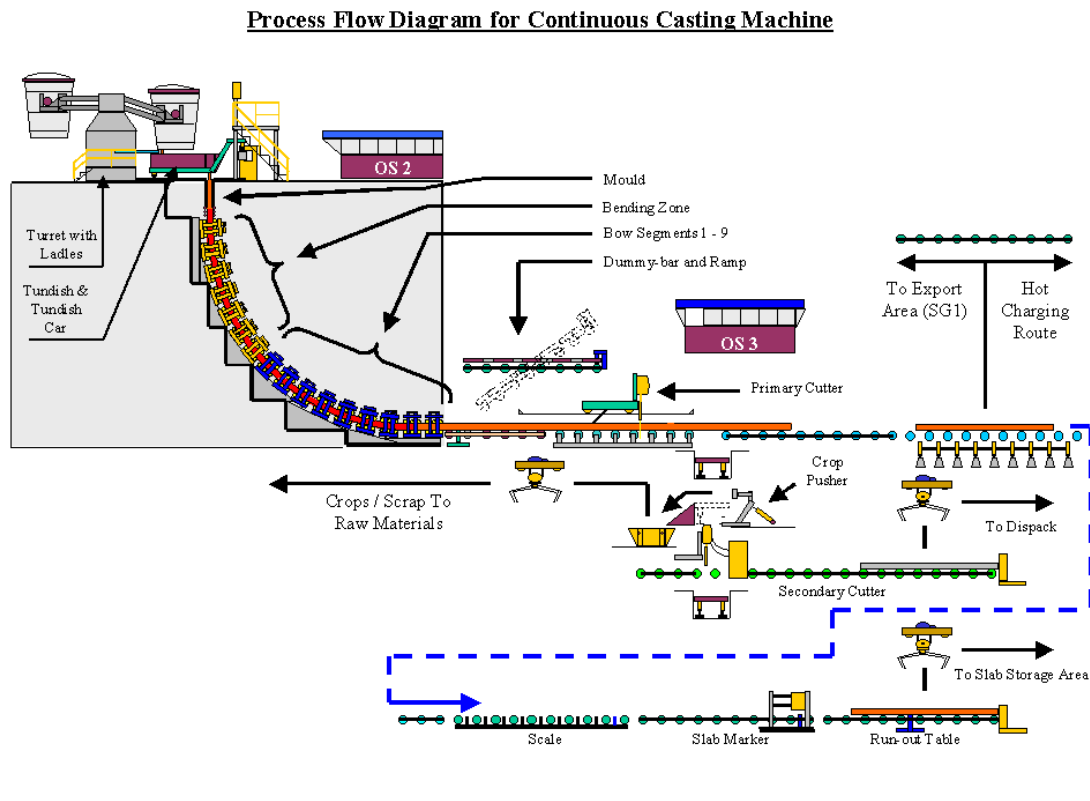


### 1.2.3 Continuous casting of steel

The majority of steel in modern times is produced via the continuous casting process. The basic principle is to continuously solidify molten steel in a water-cooled mould while withdrawing the solidified part at a constant rate. The solidified part is called a strand. A typical continuous casting machine would consist of a tundish that is fed with molten steel from a ladle. The molten steel flows from the tundish to the mould through a ceramic submerged entry nozzle (SEN). The withdrawing rate is also called the casting speed and is measured in meters per minute. The mould can take on different shapes, but slab and billet casting forms are very common. The molten metal that solidifies on the water-cooled copper mould plates forms a shell around a molten inner core. When the solidified shell exits the mould, the shell is usually strong enough to contend with the ferrostatic pressure exerted by the molten metal core. The strand that continuously exits the mould has certain dimensions. The caster at Columbus Stainless cast slabs of constant thickness (200mm) and varying widths (900mm – 1600mm). Casting powder is used to provide lubrication between the solidified shell and copper mould plates. The main constituents of casting powders are usually CaO and SiO<sub>2</sub>. The casting powders play a major role in the heat transfer between the solidified shell and copper plates. The heat transfer determines the shell thickness and if insufficient shell strength is present, a breakout usually follows (the shell rupturing with the molten metal being sprayed out into the casting machine). A breakout is a dangerous and very expensive event. When the strand exits the mould it enters the secondary cooling area of the casting machine. The strand is supported by rollers as it moves through different segments of the machine while being spray cooled. The spray cooling can be controlled from level 1 (Programmable Logic Computers (PLC's)) or from models on the level 2 computer systems (computer system responsible for the control of the process using input from the

level 1 systems). The objective of the secondary cooling practice is to solidify the molten inner core before the strand exits the machine.

***Figure 2: Schematic representation of the continuous casting machine at Columbus Stainless*** (Source: Columbus Stainless Intranet)



### 1.2.3.1 Mould heat transfer

The heat transfer in the mould is critical for the continuous casting process because the principle of casting continuously depends predominantly on the ability to extract heat sufficiently from the shell to solidify it. Jenkins<sup>[8]</sup> studied the heat transfer in the mould by looking at mould design, casting speed, steel grade and mould flux (casting powder), and concluded that for a given mould design and steel composition the mould flux was the principal factor governing the heat transfer. Jenkins<sup>[8]</sup> indicated that the contact resistance is influenced by the flux composition (basicity). Mills et al<sup>[16]</sup> highlighted the

importance of the break temperature of a flux and the effect it has on the heat transfer capability of the flux. The varying effect of mould flux on the heat transfer can be minimised by adequate compositional control from the manufacturer and using only one type of flux in the plant per steel grade. 12% Chrome non-stabilized stainless steel at Columbus is cast with only one type of flux. Heat transfer in the mould is important for slab width control because the heat transfer influences the shell thickness. It is also desirable to obtain uniform heat transfer in order to obtain uniform shell growth.

Another defect associated with a thin shell is slab bulging. Slab bulging occurs on the narrow sides of the slabs because there is no roller support (except strand guides beneath the mould) to ensure constant dimensions. The bulging can be due to insufficient shell thickness that cannot contend with the ferrostatic pressure, or due to creep of the hot shell under constant ferrostatic pressure.

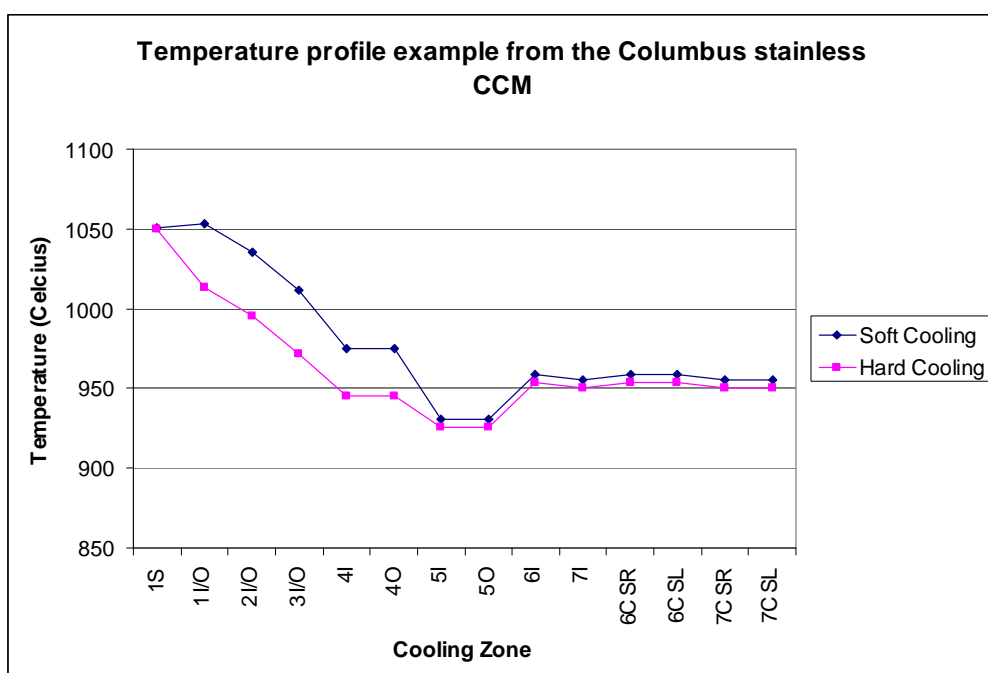
### **1.2.3.2 Secondary Cooling**

Secondary cooling in the continuous casting machine has the primary function of solidifying the strand after the mould and before it exits the machine. If the strand is not fully solidified at the exit in a vertical continuous casting machine, the result can be catastrophic. The ferrostatic pressure can cause the shell to rupture at exit and with the high ferrostatic pressure molten steel, can be sprayed out into a wide perimeter.

Secondary cooling control and optimisation are of utmost importance because defects such as slab cracking, strand width problems, bulging, central segregation and central delamination can be caused by incorrect secondary cooling practices. Different models or modes of secondary cooling exist. The two most important parameters that usually

determine the secondary cooling flow rates (for a specific steel grade), are the casting speed (m/min of strand extraction from the mould) and the superheat, of the molten metal. Intuitively, it is clear that a high casting speed with a high superheat, requires higher cooling rates than a low casting speed, and low superheat to achieve the same strand characteristics. At Columbus Stainless the secondary cooling control is done with a finite element model<sup>[18]</sup> that dynamically controls the secondary cooling per zone, to maintain a specified strand temperature in a particular zone. The process metallurgist specifies a cooling curve per material type, and the model uses the cooling curve as setpoints to adjust the cooling tempo accordingly in the different zones. Figure 3 serves as an example to indicate two typical cooling profiles. One profile would be used if a “soft cooling” (i.e. low water flow rates) is required, for example on certain crack sensitive grades. The other profile would serve for a steel grade where more aggressive cooling is required, due to various reasons. To change from a “soft cooling” to an “aggressive” cooling, the prescribed cooling profile is changed.

**Figure 3: Typical CCM temperature profiles required for secondary cooling control at Columbus Stainless**



### **1.2.3.3 Mould Set-up**

Before a cast is started in the continuous casting machine, the mould is prepared in terms of dimensions and prepared for the start-of-cast procedure. The mould dimensions determine the initial width of the strand. The term “strand width” is therefore used to describe the dimensional width of the continuously cast strand exiting the mould. To start the casting process, a dummy bar is inserted from the bottom of the mould or from the top depending on the design. The function of the dummy bar, is to provide an initial “plug” in the mould, in order to start the continuous process. When the cast is started, the dummy bar starts moving with the strand out of the mould and is later removed so that the strand can exit continuously. Before every cast, the mould is set up according to specified dimensions on the top and bottom of the mould. The bottom of the mould has smaller dimensions than the top. This provides a slope in the mould that is called the taper setting. The function of the taper is to compensate for the shell shrinkage as it solidifies in the mould. Typical taper settings are in the order of one percent. Correct taper setting is critical for heat transfer and can therefore influence the width of the strand. In practice, the mould dimensions are rarely changed and can be assumed to be constant. The mould dimensions are also measured continuously during a cast, and can therefore easily be checked for any movement. The continuous caster at Columbus Stainless is equipped with a mould that can adjust its dimensions while casting.

## **1.3 Objectives of this project**

The objective of this project is to improve the cast strand width control of 12% Chrome ferritic, non-stabilised stainless steels, cast at the continuous casting machine at Columbus Stainless. The heat composition to strand width change relationship will be used as basis for this study. The objective of this study is not to describe the fundamental

principles governing the strand width error but to use the theoretical principles in a practical way to decrease the strand width change variation.

## **1.4 Method of investigation**

The relationship of the chemistry to the cast width error on 12% Chrome non stabilised steel has been studied in detail by Siyasiya<sup>[26,27,28]</sup> and co-workers. They concluded that there is a relationship between the heat composition and the associated width change of the heat. As part of this study a relationship between the chemistry parameters that have an influence on the ferrite to austenite ratio at elevated temperatures and the width error was determined. Historical plant data from the identified parameters were used as input for different data mining techniques (i.e. empirical approach). The different data mining techniques together with the model currently used in the plant were applied to the same input data set and compared in terms of accuracy of prediction. The following data mining techniques were used:

- Statistical regression
- Decision Tree
- Fuzzy logic

## **1.5 Outline of dissertation**

Chapter one serves as an introduction to this dissertation where the background, objective and method of investigation are outlined. Chapter two gives a summary of the literature survey and starts with the background of the theoretical foundation for the heat chemistry to cast width relationship. Data mining techniques including statistical analysis, decision trees and fuzzy logic are also described. Chapter three details the data preparation used for the modelling process that includes the gathering, preparation, clean-up and selection

of significant variables as well as describing the derivation of the different models using the training data set. In chapter four the different derived models together with the model currently used in the plant are applied to the same validation data set to compare the accuracy of prediction. Chapter five outlines the results obtained by using the implemented model in the plant. Chapter six deals with the conclusions and recommendations from this particular study.

## Chapter 2

### Literature Survey

#### 2.1 Continuous Cast width Change

A limited amount of literature is available on the width change phenomenon at continuous casters. All literature agrees that cast width change can be very problematic for downstream processes and sometimes the real effect of poor dimensional control is hidden<sup>[38]</sup>. Most of the literature deals with techniques that were successfully implemented to improve the cast width change. The techniques range from simple monitoring techniques to theoretical prediction models. Evans et al<sup>[42]</sup> achieved improved width change results by better slab width verification techniques. The mould face positions were continuously monitored and the hot and cold width were matched. Unfortunately it is not evident how the matching between the cold and hot width was done. Nakamura<sup>[41]</sup> et al implemented a slab width control model that controls the width of the strand by changing the mould dimensions using a mould that can continuously adjust its dimensions. The model is based on the steel grade, casting speed and compression force that is influenced by the braking force during compression casting. The model is essentially based on the concept to correct the width during transient conditions. The model divided a cast into four portions. Portion one is the start-up where a narrow slab width is usually the result due to low ferrostatic pressure and overcooling. The mould dimensions are changed to start wide during portion one. Portion two is the section after start-up where the speed is slowly increased. The mould dimensions would also then be slowly adjusted to a narrower setpoint. Portion three is the “steady state” portion where the width change can be explained by changes in the casting speed. Portion four is the last section and represents “end of cast” conditions where the mould width is



increased again because the strand will be narrow due to low ferrostatic pressure and overcooling. Assar et al<sup>[39]</sup> attribute the widening of the slabs during casting to the ferrostatic pressure and shell malleability. They found the widening of the slabs to be directly correlated to the casting speed and depend on both the strand width and steel grade. They developed a prediction model based upon casting speed, mould width and steel grade. Their results indicated that wider slabs expand more and will therefore undergo a bigger width change. The prediction model was going to be implemented to continuously predict the cold width of the strand. The mould width would then be adjusted should the need arise to correct the strand width.

Kocatulum et al<sup>[40]</sup> studied the causes of width change using physical devices to measure slab width and also developed a statistical slab width prediction model. They found that the slab resident time in the upper cooling sections of the caster has a huge influence on the slab width change. The resident time in the upper sections of the secondary cooling correlates directly with the casting speed. They also concluded that the mould width plays a very big role in the resulting slab width. A slab width prediction model was developed incorporating the mould width and the resident time in the upper sections of the secondary cooling. They achieved very accurate results with the model. Mostert and Brockhoff<sup>[38]</sup> developed a slab width prediction model based on theoretical principles. They determined an equation for the shrinkage factor and stated that the final cold width of a slab is influenced by 1) thermal shrinkage and 2) contraction counteracted by expansion. They expressed the shrinkage factor as the ratio of mould width to cold slab width. The effects that were considered are the thermal shrinkage, expansion due to ferrostatic pressure and the influence of slab width. An expression for the thermal shrinkage was determined by modelling the strand as a rigid bar taking into account the

effect of the taper. The expansion of the shell depends on the thickness and temperature of the shell combined with ferrostatic pressure. The relationship between the spray cooling, ferrostatic pressure, casting speed, shell thickness and shell temperature was studied. They concluded that the expansion depends on the ferrostatic pressure, spray cooling and casting speed. They also concluded that there is a slight dependency of the slab width change on the mould width.

If all the previous findings are combined then the cast width change is influenced by the following factors as concluded by Assar et al<sup>[39]</sup>:

1. Grade Chemistry
2. Strand width (mould width)
3. Cooling curve (secondary cooling intensity)
4. Casting speed

## **2.2 Fundamentals of the chemistry to cast width error relationship**

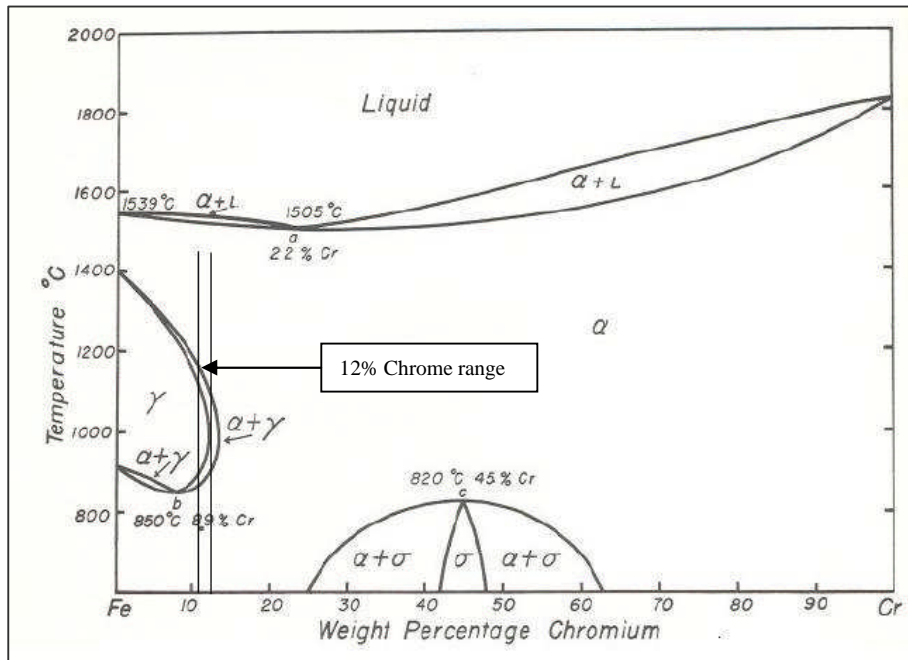
In this section the mechanism of how the phase ratio between austenite and ferrite influences the width error of the strand will be proposed. It is important to understand how the theoretical relationship of the phase ratio between austenite and ferrite influences the width error in practise, because this is the fundamental building block on which this study is based. An overview of the theory behind the chemistry to cast width error relationship will be described first before the mechanism is proposed.

## 2.2.1 Theoretical fundamentals

This study deals specifically with 12% chrome non-stabilised ferritic stainless steel and the relationship of the heat composition and associated cast width change of the heat. The main reason why 12% chrome ferritic steels (non-stabilised) would exhibit a relationship between heat composition and cast width change (width error) is that it goes through a dual phase region between austenite and ferrite (termed the “gamma loop”) during the secondary cooling stage. Research<sup>[26]</sup> on the transformation behaviour and hot strength of the 12% Chrome non-stabilised steel has indicated that as long as the ratio of austenite and ferrite keeps on fluctuating, the width error variation will persist. This dual phase characteristic plays a role in the width change due to three reasons:

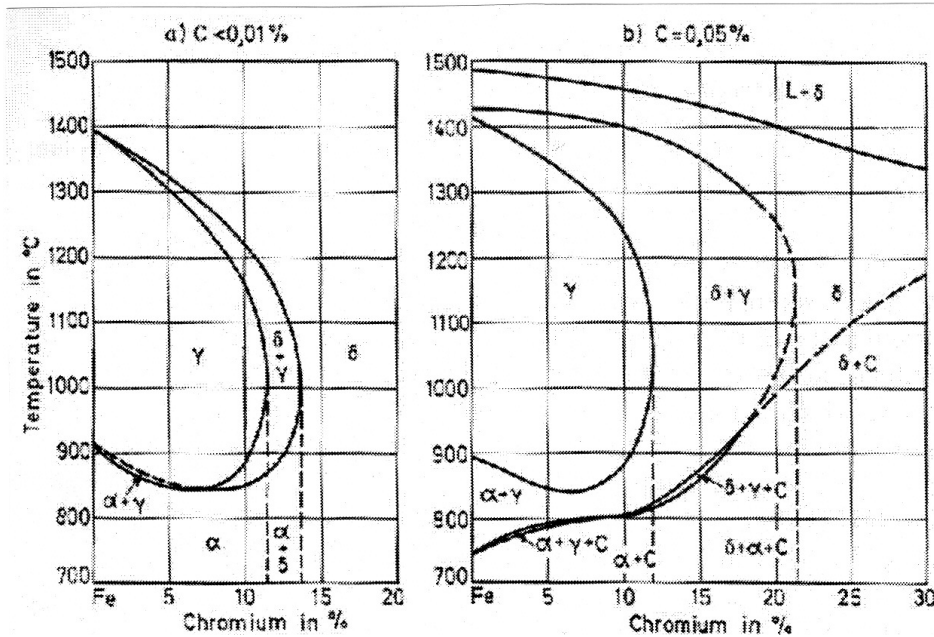
1. The temperature range at casting exit is in the order of 800°C to 950°C where the steel is still expected to be in the dual phase region. The ratio between austenite and ferrite depends on the chemical composition and the cooling intensity<sup>[27]</sup>. The lattice structure of austenite is FCC (Face Centred Cubic) and the lattice structure of ferrite is BCC (Body Centred Cubic). The BCC structure occupies a higher volume than the FCC structure and therefore changes in the ratio of austenite to ferrite will be accompanied by a change in the width<sup>[27]</sup>. The dual phase region is evident from the Fe-Cr binary phase diagram in Figure 4a<sup>[23]</sup>.

**Figure 4a: Fe-Cr Binary phase diagram<sup>[7]</sup>**



The 12 % Chrome material therefore will go through the gamma loop as it cools from above 1250°C. Rowlands<sup>[25]</sup> also indicates that the 12% chrome content places it at the critical boundary of the gamma loop and therefore austenite or ferrite stabilisers (formers) can change the structure of the steel at elevated temperatures. Elements that are important ferrite stabilizers in stainless steels are chrome, titanium and molybdenum. Important austenite stabilisers are carbon, nickel, nitrogen and manganese. Small changes in any of these elements can change the austenite to ferrite ratio in a 12% chrome, non-stabilised ferritic stainless steel. Figure 4b indicates the change to the austenite/ferrite region when a strong austenite stabiliser is added like carbon. The “gamma loop” (austenite/ferrite) region is shifted to higher chromium levels and the transformation temperature from delta ferrite to austenite is increased.

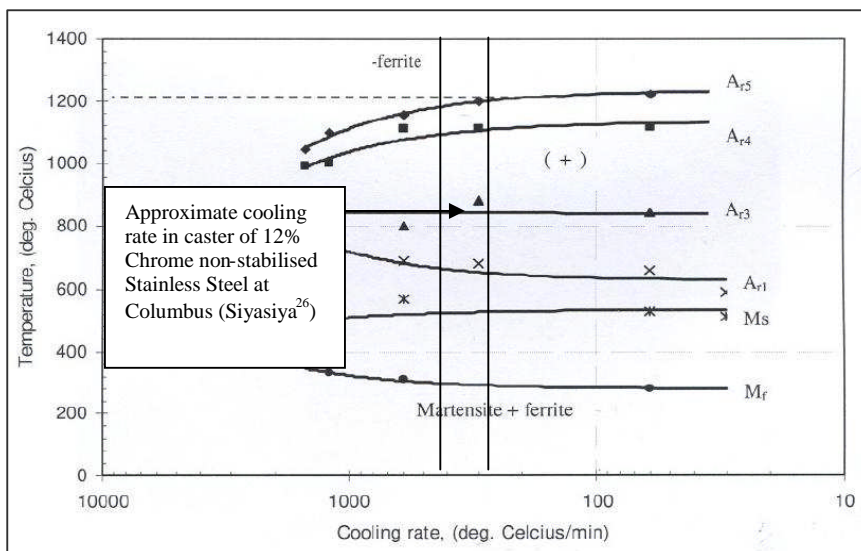
**Figure 4b: Effect of carbon content on the austenite/ferrite region<sup>[27]</sup>**



It can be seen that the delta ferrite and austenite phase region has been expanded and the transformation temperature from delta ferrite to austenite has increased. Variations in the austenite and ferrite stabilisers therefore influences the gamma loop which influences the phase ratio's and the transformation temperatures.

The CCT diagram depicted in Figure 5<sup>[26]</sup> also indicates the dual phase region of the 12% Chrome ferritic (non-stabilised) material. The typical cooling rate experienced during secondary cooling is also plotted.

**Figure 5: CCT diagram for typical 12% Chrome non-stabilised ferritic stainless steel<sup>[27]</sup>**



The dual phase temperature range corresponds to the temperature range between mould exit and continuous casting machine exit. Due to the dual phase nature between austenite and ferrite together with slight variations in heat composition, there will always be a certain ratio between the austenite and ferrite fraction that will be determined by the heat composition (ferrite and austenite stabilisers).

2. The hot strength of austenite is more than the hot strength of ferrite which will influence the ability of the formed shell to contend with the ferrostatic pressure if the phase ratio between austenite and ferrite changes. The more the ratio favours austenite, the “stronger” the shell will tend to be and vice versa if the ratio favours ferrite. A “stronger” shell will be able to contend more effectively with any forces that will attempt to deform it (mainly ferrostatic pressure). A shell consisting predominantly of austenite will also be less sensitive to shell thickness variations caused by fluctuating in-mould conditions. The opposite is true for a shell consisting predominantly out of ferrite, it will be sensitive to any forces acting upon it and shell thickness variations will also play a large role in the tendency of the shell to deform.
3. The inherent resistance to creep of austenite is more than that of ferrite<sup>[50]</sup>. Hence changes in the phase ratio between the austenite and ferrite influences the creep characteristics of the strand. When a metal or an alloy is under constant load or stress, it may undergo progressive plastic deformation over a period of time. This time-dependent strain is called creep.<sup>[46]</sup> Creep properties are usually determined by means of a test in which a constant load or stress is applied to the specimen and the strain is recorded as a function of time. This creep curve exhibits various stages. Directly on loading the specimen undergoes an instantaneous strain. This

is followed by the primary stage. The creep rate then declines gradually until it reaches a constant value which is called the secondary stage. During the tertiary stage the creep rate increases again until final fracture<sup>[45]</sup>. Creep processes are diffusion-controlled and they become of particular importance in materials experiencing extensive time at elevated temperature especially if a stress or load is also present. The elevated temperature that is important is typically above  $0.4T_m$  where  $T_m$  is the absolute melting point<sup>[47]</sup>. If a continuous casting process is considered, then typically the strand is at a temperature  $> 0.4 T_m$  for the whole of the secondary cooling stage and is therefore prone to undergo creep at high temperature. Temperature plays a major role in the creep rate with a higher creep rate experienced with a higher temperature<sup>[45,48]</sup>. The creep resistance of austenitic stainless steels are increased by the following elements: carbon, nitrogen, chromium, molybdenum, tungsten, vanadium, boron, titanium and niobium<sup>[50]</sup>. According to Austin et al<sup>[48]</sup> the following elements have an influence on the creep properties of ferrite (ferrite itself does not have a high resistance to creep at elevated temperatures), nickel, silicon and cobalt only increases the creep resistance at elevated temperatures marginally while carbon, chromium, manganese and molybdenum increase the resistance to creep markedly. Jamieson et al<sup>[49]</sup> also found that nitrogen and silicon increases the initial creep strength of ferrite.

### 2.2.2 MEDUSA model

In order to quantify and calculate the constitution and transformation properties of 12% chrome stainless steels (non-stabilised) the Columbus research and development department has developed a model to calculate the phase fractions and transformation tempos of 12% Chrome stainless steels (non-stabilised) given the chemistry composition. The model is called MEDUSA<sup>[30]</sup> (**M**athematical **E**valuation of **D**ilatometry **U**sing **S**tatistical **A**nalysis). The model was developed using dilatometry, mechanical testing and image analysis of 500 laboratory heats and 300 plant heats. Most of the experimental work used to derive the MEDUSA model was done on 12% chromium heats but chromium levels up to 25 % were also included in the model. The effect of all residuals was examined by deliberately making steels with a wide composition range to stabilise the model. Continuous cooling diagrams were constructed for all 800 heats used to develop the model. The phase balance at 1000°C was also measured to allow, along with the CCT diagrams, the calculations of the hardenability curve. The following parameters are calculated (amongst others), by the MEDUSA model and are of value for the present study. All the equations pertaining to the specific parameter is given in Appendix C.

#### ***Gamma max(%)***

This is the amount of austenite present at 1100°C in ferritic stainless steels

#### ***AC1 temp (°C)***

This is the temperature at which austenite begins to form when ferritic stainless steel is heated at 1°C/min.

#### ***CR95 (°C/min)***

Cooling rate at which 95% of the austenite present at 1000°C transforms to ferrite. This is a measure of the transformation response of austenite to ferrite.



### ***A<sub>max</sub> (Austenite potential)***

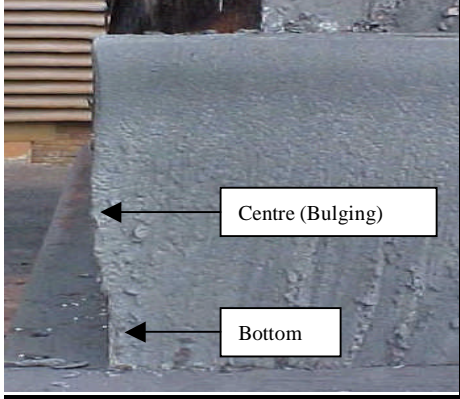
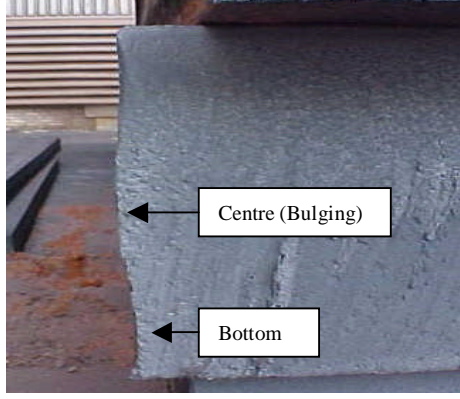
This is the percent austenite present in ferritic stainless steels at 1000°C and approximates to the maximum austenite potential of the steel.

## **2.3 Proposed mechanism for the chemistry to cast width relationship.**

The mould set-up in terms of the hot bottom width is in the order of 30 mm smaller than the aim cast width. A schematic representation of the mould set-up and aim strand width is given in Appendix A. This means that the material actually creeps/flows wider during the secondary cooling stages. The dual phase characteristic of the solidified shell was discussed earlier, and it follows from the difference in hot strength and creep properties between austenite and ferrite that the width change can be different for different heats. Intuitively it is clear that a low creep rate will lead to an under-width strand and a fast creep rate will lead to an over-width strand. Another mechanism that will play a role in the final strand width, is the amount of slab bulging on the narrow sides. If the creep rate is low and the shell is strong, the slab dimensions should remain intact and little or no bulging should be evident. The opposite is true for a shell with high creep rate and low strength, where bulging due to the ferrostatic pressure should be evident. These mechanisms are easy to evaluate in the plant by inspection and measurement of slabs and comparison to the width error. Figure 6 gives examples and pictures of actual slabs that were measured and inspected. The results are given by comparing the width as measured in the centre of the slabs, to the width measured at the bottom of the slab. Measurements were taken from slabs that had cast width errors of +19.9 mm and +18.8 mm respectively, because the mechanisms causing the width change were present. From the

measurements of the slabs it is clear that the width change is a combination of excessive creep of the shell and bulging of the narrow sides. According to the measurements, the contribution of each mechanism is on the order of 50%. The characteristics of the shell creeping and bulging in terms of creep/bulging rate and position in the continuous caster is unclear, but research<sup>[26,28]</sup> has shown that bulging must take place within the first couple of metres after mould exit, where the shell is still thin. It is also expected that the creep rate will be higher during the initial stages of secondary cooling because the temperature is high. One way of combating bulging, is to increase the water flow rates in the secondary cooling, in order to form a cooler shell that will be able to withstand the ferrostatic pressure. “Hard cooling” on the 12% Chrome ferritic non-stabilised steels has in the past resulted in micro cracks forming on the slab surface, due to low hot ductility problems in the unbending zone. It should also be kept in mind that not all slabs are bulging, and increasing the cooling intensity will be effective to a limited extent. The strategy decided upon in this study was to attempt to predict the heats that will exhibit bulging/creep problems and increase/decrease the water flow rates on these selected heats in order to limit the amount of bulging and change the creep properties. The increase in water flow rates is kept below the level that would induce micro crack problems mentioned earlier.

**Figure 6: Slabs measured in plant to indicate width error mechanism**

Slab number 3494843: Width error measured by laser +19.9 mm wide	Slab number 3494873: Width error measured by laser +18.8 mm wide
	
Slab Measurements: Centreline cast width error = +20 mm Bottom cast width error = +12 mm	Slab Measurements: Centreline cast width error = +19mm Bottom cast width error = +10 mm

## 2.4 Description of the model being used in the production environment

This model was initiated when it was found that there was a relationship between the heat composition and the resulting width change of the heat. The work done by Siyisiya<sup>[26,27,28]</sup> indicated that there is a relationship between the heat chemistry and resulting width change of a specific heat on 12% chrome non-stabilised ferritic material based on the dual phase characteristic of the solidified shell. Logic suggested that the important parameters that should correlate with the width change are those that influences the austenite to ferrite ratio at high temperatures. It was found that good results were achieved by not using only the heat composition in terms of austenite and ferrite stabilisers but by also using the calculated parameters from the MEDUSA model

as described earlier. Experience indicated that the combination of the following input parameters gave satisfactory results: AC1, Amax, Gamma max, Kaltenhauser ferrite factor, addition of the absolute carbon and nitrogen and the CR95. The model was developed using very simple analysis techniques. A set of rules was generated from historical data. The width change was divided into three groups (narrow from aim, acceptable and wide from aim) and the combination of the parameters that resulted in the width error falling into one of the three groups were determined. The combinations of parameters are expressed in terms of rules. The rules give the freedom of not having to use all the input parameters. The amount of parameters per rule can range from one to six. The model therefore consists out of a mixture of rules but all having the form of IF-then statements. A typical example of a rule incorporating four parameters is illustrated below:

If AC1 < x AND Gmax > y AND CR95 < z AND (C+N) > zz THEN .....

The parameters are calculated for each heat and then sent to the model for evaluation. If the model classifies it as being problematic (either narrow or wide), then the secondary cooling is changed automatically to compensate for the expected width error. The secondary cooling is changed by selecting a different pre-defined secondary cooling strategy. The philosophy is therefore to selectively change the secondary cooling on selected heats to better control the width error. This model has been running “live” in the plant since November 2004. More details and preliminary results of the current model can be found in De Beer<sup>[7]</sup>.

## 2.5 Data mining

### 2.5.1 Introduction

A functional general definition of data mining according to Weaver<sup>[34]</sup> is the use of numerical analysis, visualization, or statistical techniques to identify non-trivial numerical relationships within a dataset, to derive a better understanding of the data and to predict future results. Data mining is also described Anon<sup>[2]</sup> as an analytical process designed to explore data in search of consistent patterns, and/or systematic relationships between variables, and then to validate the findings by applying the detected patterns to new subsets of data. The ultimate goal of data mining is prediction, and predictive data mining is the most common type of data mining and has the most direct business application<sup>[2]</sup>. Data mining is conceptually based on statistical techniques like the traditional Exploratory Data Analysis (EDA). However, an important difference between Exploratory data Analysis (EDA) and data mining, is that data mining is more concerned with the application than the basic nature of the underlying phenomena. In essence, data mining is less concerned with what the exact relationships are between the variables, but is more concerned to obtain a solution that can be used for accurate prediction<sup>[2]</sup>. The application of data mining techniques is found in a diverse array of industries like steel manufacturing<sup>[1],[20]</sup>, the chemical industry<sup>[5]</sup> and even drug discovery<sup>[34]</sup>, to name a few. A survey done in November 2003 of 213 data mining users<sup>[3]</sup>, indicated the most regularly used data mining techniques are the following:

Table 1: Survey of data mining technique usage.

Technique	Percentage of 213 votes
Decision tree / Rules	16%
Clustering	12%
Statistics	12%
Neural Networks	9%
Logistic regression	9%
Visualization	7%
Association rules	5%
Nearest Neighbour	5%
Text Mining	4%
Web Mining	4%
Bayesian nets / Naïve Bayes	3%
Sequence analysis	3%
Support Vector Machine	3%
Hybrid Methods	3%
Genetic Algorithms	2%
Other	3%

Source: Anon2<sup>[3]</sup>

From Table 1, it is clear that the first four techniques account for approximately 50% of the surveyed users, and the conclusion can be made that these techniques are the most popular. For this reason, it was decided to use two of these techniques in an attempt to describe and predict the patterns of heat chemistry versus strand width error. Each of the techniques consists of many branches and the chosen branches relevant to this study will be discussed.

The process of data mining consists out of three main stages<sup>[2]</sup>.

Stage 1: Data exploration

Stage 2: Model building and validation

Stage 3: Deployment.

A more detailed description of the process is described in the course outline offered by Olivia<sup>[21]</sup> on data mining. The chapter headings can be used as a process map for data mining.

### ***Step 1: Defining the objective***

The first step in any modelling process is to define the objective<sup>[21]</sup>. The objective usually determines what techniques are used and how they are applied. The objective of this study is to test various techniques for accuracy in a specific application and to develop an on-line prediction model that is sufficiently accurate.

### ***Step 2: Gathering the data***

Accurate, actionable, accessible data are the lifeblood of any successful model<sup>[21]</sup>. Unfortunately a model will only be as good as the quality of the training data set. Extreme care must be taken to ensure that the data are correct and in the correct form. The data collected, must be relevant to the dependant variable that must be predicted. Data can be collected manually but extracting data from an existing database is the common way of data collection for industrial processes.

### ***Step 3: Preparing the data for modelling***

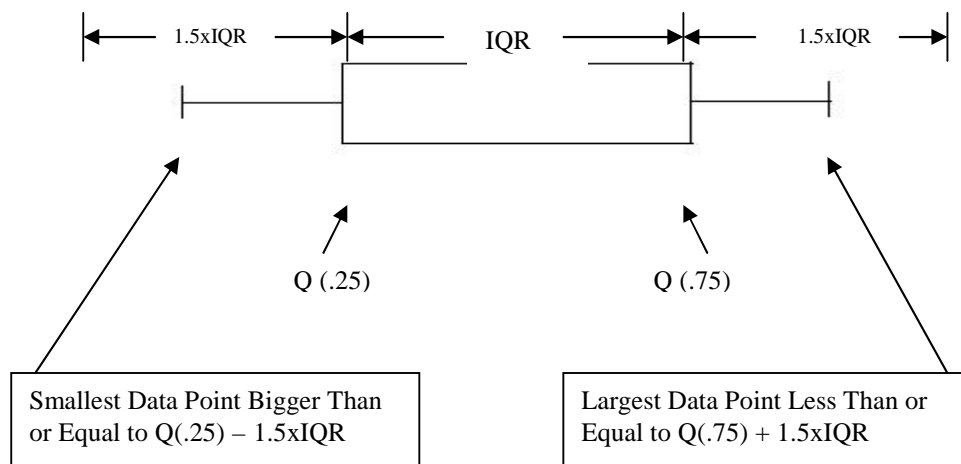
The average modeller spends 60% of his or her time preparing the data for modelling<sup>[21]</sup>. The old rule of “garbage in garbage out” will apply if the training data set is not correct. Some typical operations during this stage are to check for missing data or outliers. Missing data is a common problem in databases of industrial processes. The origin of missing data in an industrial process database can be anything from measurement errors, level 1 to level 2 computer communication problems, or level 2 problems, to mention just a few. Two typical ways of dealing with missing values is to populate the missing values with predetermined values, by using for example, the average of the specific field’s data or by using some kind of perturbation technique or even using random numbers within a specific range. The other option is to eliminate records, where some of the fields (either independent or dependant variables) are missing. Outliers are also common in industrial process databases and the reason for the outlier can be amongst others, measurement

errors, default values or database errors. One technique of identifying outliers is described by Vardeman<sup>[32]</sup> and is based on a box plot. The “box” is constructed by using the first or lower quartile and third or upper quartile presented as  $Q(0.25)$  as the lower and  $Q(0.75)$  as the upper quartile. An interquartile range (IQR) can now be defined as:

$$IQR = Q(0.75) - Q(0.25)$$

The IQR can be used to determine outliers by identifying as outliers all points larger than  $Q(0.75) + 1.5 \times IQR$  and all points less than  $Q(0.25) - 1.5 \times IQR$ . Figure 7 indicates this technique graphically.

**Figure 7: Box plot indicating outliers by using IQR (Interquartile range)**



***Step 4: Selecting and transforming variables***

Typical operations during this stage will include Anon<sup>[3]</sup> data transformations, selecting sub-sets of records and in data sets with a large number of variables, the variables will be screened. Redundant variables that do not contribute to the outcome of the model will be eliminated. This process will bring the number of variables into a manageable range. The accuracy of a model can also be influenced if redundant or inactive variables are included.



### ***Step 5: Processing and evaluating the model***

During this step different models are fitted to the training data set. Each model will give an output of predictions. The area of interest is usually the accuracy in terms of prediction of the different models. The best model is the model with the most accurate prediction. There are different techniques for evaluating and comparing the performance of predictive models. According to Anon<sup>[3]</sup> the following techniques are available for model evaluation: Bagging (Voting, Averaging), Boosting, Stacking (Stacked Generalizations) and Meta-Learning. The basic principle of these techniques is “competitive evaluation of models” which is to apply different models to the same data set and then choosing the best.

### ***Step 6: Validating the model***

By definition, models should perform well on the development data<sup>[21]</sup>. A true test of a model is however the performance on a new data set from a different time span or for example different market. According to Olivia<sup>[21]</sup>, there are three powerful methods to ensure good model fit. 1) Scoring on alternative data sets gives a good indication of the model performance; 2) Bootstrapping uses simple resampling techniques to find confidence levels around the predictions; 3) Key Variable analysis calculates important factors as they are effected by the model.

### ***Step 7: Implementing and maintaining the model***

The final stage is the implementation of the chosen model into the application area. In this study, the end product will be an on-line model capable of predicting the width error of a 12% Chrome ferritic non-stabilised heat before it is cast. Model maintenance is a very important part of any implementation process. The type of maintenance will depend on the type of model implemented, the systems or software where the model is run, to name a few. A very common area for model maintenance in an industrial process is to

update the model each time a change to the process is made that can influence the model predictions. It might even be necessary to recreate a model after a change. Wrong predictions can be the result if the model is not adequately maintained.

### **2.5.2 Statistical analysis**

Statistics are seen as a “Classical” technique. The “Classical” techniques are in contrast to the so-called “next generation techniques”, such as neural networks and decision trees. By strict definition, statistical techniques are not data mining techniques, because the statistical techniques have been around before data mining became a recognized technique in industry. However, statistical techniques are driven by the data and are used to discover patterns in the data that can be used for prediction<sup>[31]</sup>.

#### **What is the difference between statistics and data mining ?**

According to Thearling<sup>[31]</sup>, it is a very difficult question to answer. The problem is that if a statistical technique is successful, the reason for its success is exactly the same as for a data mining technique (clean data, clear target and good validation). The other issue that makes a clear distinction very difficult, is that the application area for statistical techniques and data mining techniques are exactly the same with even the same purpose (prediction, classification discovery). There are however some benefits of data mining above statistical techniques.

1. Data mining techniques tend to be more robust towards “messier” real world problems.
2. Data mining techniques are more forgiving for less expert users.

3. With the advances of computing technology it becomes viable to use very complex data mining techniques that might be better suited to a specific data set than classical statistical techniques.

### **2.5.2.1 Computationally simple descriptive statistics**

Statistics are all about data. In industrial processes there are usually a vast number of variables combined with a vast number of measurements collected in the process databases. The following questions are usually asked about the dataset under investigation<sup>[31]</sup>.

- What patterns are in the database?
- What is the chance an event will occur?
- Which patterns are significant?
- What is a high level summary of the database that will give an idea of what is contained in the database?

A great advantage of statistics is that it provides a high level view of the dataset without having to understand individual records. There are several statistical techniques that can be employed for giving an overview of a dataset and there are also some techniques that can be used for prediction. The term used for “overview” statistics is “descriptive” statistics. Vardeman<sup>[32]</sup> splits descriptive statistics in two categories; 1) Computationally simple descriptive statistics and 2) Computationally intensive descriptive statistics. Predictive methods fall mostly into the latter category.

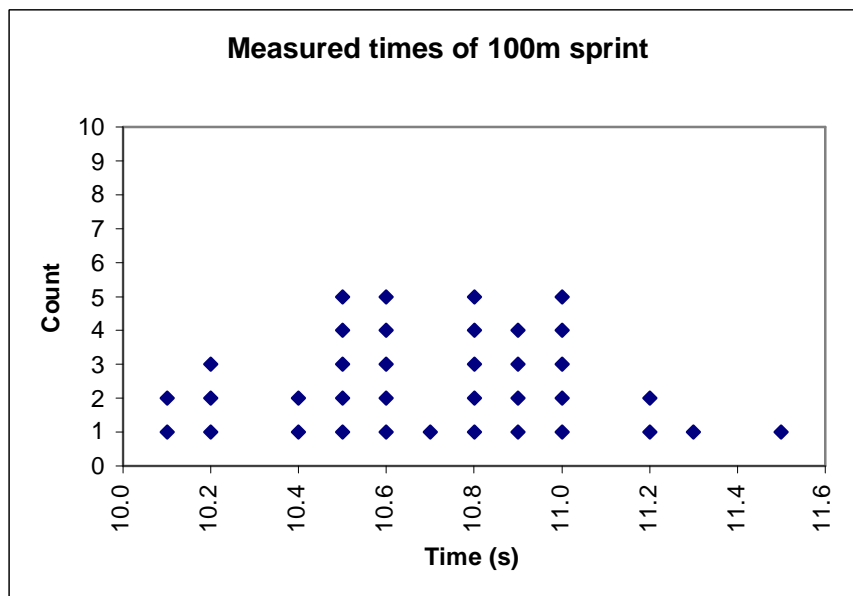
The typical place to begin in the analysis of data is to visually plot the data on simple graphs. Not much more analysis would be needed in simple engineering problems than a

visual summary. The following techniques exist for simple visual representation of univariate data: 1) Dot diagrams, 2) Stem-and-leaf plots, 3) Frequency tables, 4) Histograms, 5) Scatterplots and run charts.

### 2.5.2.1.1 Dot Diagrams<sup>[32]</sup>

Dot diagrams are useful tools for small amounts of univariate quantitative data. One makes a dot diagram by plotting each observation as a dot placed at a position corresponding to its numerical value along a number line. Figure 8 indicates a typical Dot diagram. The diagram gives the measured times for a 100 m sprint by an individual athlete.

**Figure 8: Sample Dot Diagram**



If the data has significant figures of typically two or more then it is usually necessary to round the value of the data points off to enable plotting. It becomes then impossible to read the original data values from the graph. This is a disadvantage of the dot diagram visual presentation. An advantage of the dot diagram is the ease of implementation and relatively large amount of information obtained.

### 2.5.2.1.2 Stem-and-leaf plots<sup>[32]</sup>

A stem-and-leaf diagram presents the same amount of visual information as the dot diagram but the original values are preserved. A stem-and-leaf plot is made by using the last few digits of each data point to indicate its position. Figure 9 below indicates a stem-and-leaf diagram for a series of 100m sprint times.

**Figure 9: Stem-and-leaf plot of 100m sprint times**

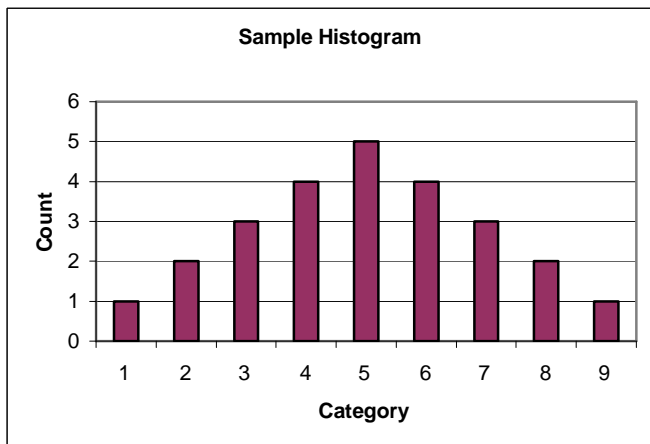
10	.	21	.	28	.	29	.	45	.	57	.	6	.	88	.	9	.	9	.	9	.	95	.	99
11	.	01	.	05	.	19	.	22	.	4	.	48	.	5	.	6	.	78						
12	.	3	.	46	.	5	.	59	.	8														
13	.	2	.	5	.	6	.	8																

A more useful way of presenting a stem-and-leaf diagram is to plot two diagrams back-to-back. Stem-and-leaf and dot diagrams are useful tools to get a “feel” for the data and to provide an overview of the data. They are however not commonly used in engineering reports or presentations. Frequency tables or histograms are more commonly used for formal presentation of data.

### 2.5.2.1.3 Frequency diagrams and histograms<sup>[32]</sup>

A frequency table is made by first breaking the interval containing the data into smaller appropriate intervals of equal lengths. The frequency of each interval can be calculated by counting the amount of data points in the interval. Relative and cumulative frequencies can be calculated. Plotting the frequencies as a type of bar chart creates a histogram. The histogram gives insight into the shape of the distribution.

**Figure 10: Sample Histogram**



#### **2.5.2.1.4 Scatterplots and Run Charts<sup>[32]</sup>**

Dot diagrams, stem-and-leaf plots, frequency tables and histograms are univariate tools. Mostly real engineering problems are multivariate. Relationships between the variables are usually also important and can give some insight into the process being studied. A scatterplot is a two-dimensional plot of data and can quickly indicate a relationship between two variables if it exists. If the X-axis is a time axis then the scatterplot is called a run chart. Seeing patterns on a run chart gives insight into the change of variables over time.

#### **2.5.2.2 Computationally intensive descriptive statistics<sup>[32]</sup>**

The techniques described in this section are computationally more complicated and are usually performed using some kind of statistical software package. The discussion will start with simple straight line fitting using the least squares approximations and then moving to more advanced methods of multivariate linear regression using polynomials and surface fitting.

### 2.5.2.2.1 Fitting a straight line using least squares approximations

A very commonly used method to describe bivariate data is to fit a straight line to it using the following generic equation:

$$Y \approx \beta_0 + \beta_1 X \dots\dots\dots(1)$$

The advantage of describing data with an equation is that 1) Summarization is made possible, 2) Interpolation can be performed, 3) Limited Extrapolation can be done and 4) Process optimisation is made possible based on findings using the equation. The principle of using least squares in the fitting of an equation for Y containing some parameters to an n-point data set, values of the parameters are chosen so as to minimise

$$\sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \dots\dots\dots(2)$$

Where  $Y_1, Y_2, \dots, Y_n$  are the observed responses and  $\hat{Y}_1, \hat{Y}_2, \dots, \hat{Y}_n$  are corresponding responses predicted or fitted by the equation.

$\beta_0$  and  $\beta_1$  are calculated with the following equations:

$$\beta_1 = \frac{\sum X_i Y_i - \frac{(\sum X_i)(\sum Y_i)}{n}}{\sum X_i^2 - \frac{(\sum X_i)^2}{n}} \dots\dots\dots(3)$$

$$\beta_0 = \bar{Y} - \beta_1 \bar{X} \dots\dots\dots(4)$$

The next question that must be answered is how good the fit with the obtained equation is. There are two commonly used methods for checking the quality of the fitted line 1) Sample correlation, 2) Coefficient of determination.

The **Sample Linear Correlation** between x and y exhibited in a sample of n data points  $(x_i, y_i)$  is:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \dots\dots\dots(5)$$

The value of r falls in the range  $-1 < r < 1$  range with values close to -1 and 1 indicating strong relationships between y and x. The sign indicates if the relationship is proportional or inversely proportional.

The **Coefficient of determination** for an equation fitted to an n-point data set via least squares and producing fitted Y values  $\hat{Y}_1, \hat{Y}_2, \dots, \hat{Y}_n$  is the quantity :

$$R^2 = \frac{\sum (Y_i - \bar{Y})^2 - \sum (Y_i - \hat{Y}_i)^2}{\sum (Y_i - \bar{Y})^2} \dots\dots\dots(6)$$

The value of  $R^2$  is in the range  $0 \leq R^2 \leq 1$  and has the interpretation of the fraction of the raw variation in Y accounted for using the fitted equation.

Another important concept in fitting a line with the least squares method is the calculation of residuals. When one fits an equation to a set of data, the hope is usually that the equation extracts the main message of the data, leaving behind only the variation in  $Y_i$  that is uninterpretable. In essence one hopes that the predicted/calculated  $\hat{Y}$ 's looks like  $Y_i$  except for some small random variation. One way of checking this phenomenon is by calculating and plotting the residuals.

If the fitting of an equation or model to a data set with responses  $Y_1, Y_2, \dots, Y_n$ , produces fitted  $\hat{Y}_1, \hat{Y}_2, \dots, \hat{Y}_n$ , then the corresponding residuals are  $e_i = Y_i - \hat{Y}_i$ .



The idea is that when the residuals are plotted against some sensible quantity the result should be some patternless graph. If there is some pattern to the graph, then it is evident that there is some extra driving force not taken into account by the equation.

#### 2.5.2.2.2 Fitting surfaces using least squares approximations

This section deals with linear multivariate regression. The typical problems in industry are rather multivariate than univariate, making multivariate regression particularly useful in industrial problems. The first curve to be considered is the generalization of the straight line equation which is a  $K^{\text{th}}$ -order polynomial in a  $K$  variables equation given by:

$$Y \approx \beta_0 + \beta_1 X + \beta_2 X^2 + \dots + \beta_k X^k \dots\dots\dots(7)$$

The solution of the equation is usually done by software on a personal computer and the software is freely available. The same techniques of correlation and coefficient of determination can be used to check the quality of the fitted curve. The coefficient of determination and visual examination should be used to determine the best polynomial for the data. Residual plots are also useful tools to determine the appropriateness of the polynomial.

If a surface needs to be fitted to a set of data points consisting out of  $x_1, x_2, \dots, x_k$  quantitative variables on some response  $Y$ , then the surface to be fitted will have an equation of the form (Linear polynomial in  $K$  variables):

$$Y \approx \beta_0 + \beta_1 X + \beta_2 X_2 + \dots + \beta_k X_k \dots\dots\dots(8)$$

The calculation of the  $\beta_{1..k}$  parameters is done using software on a personal computer. The coefficient of determination and residual plots can be used to check the quality of the fitted surface. With increasing variables it becomes more difficult to plot the surface to

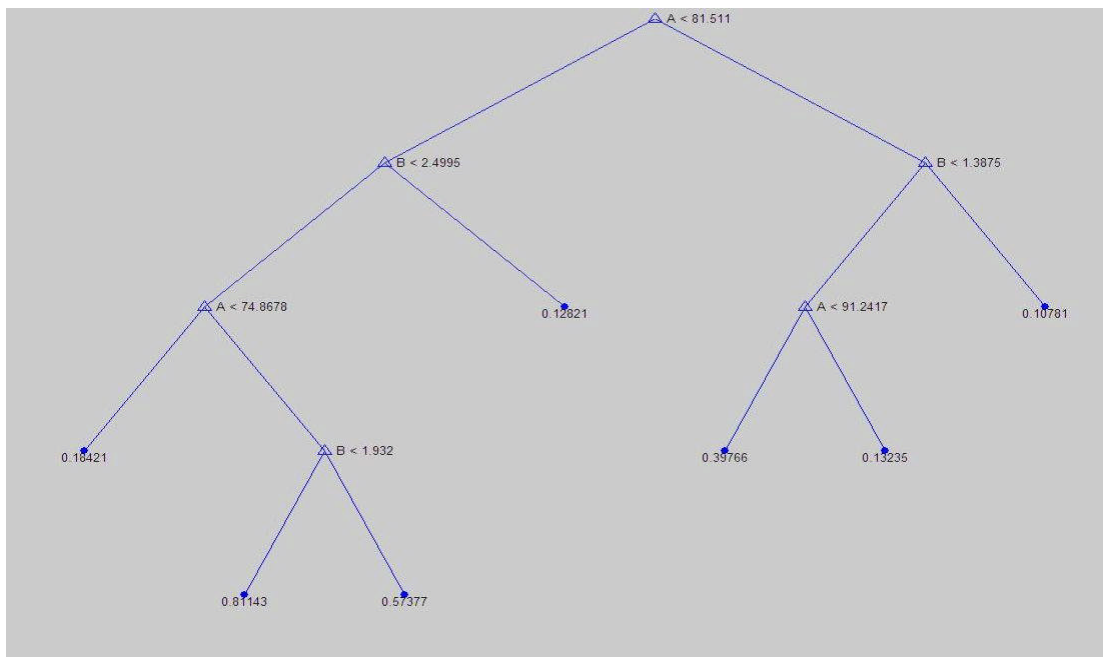
check it visually against the data, and consequently the coefficient of determination and residual plots becomes increasingly important.

## 2.5.3 Decision trees

### 2.5.3.1 Introduction

Decision tree classifiers are used successfully in many diverse areas like radar signal, character recognition, remote sensing, medical diagnosis, expert systems and speech recognition and stainless steel quality predictions<sup>[1]</sup>, to name a few. Arguably the most important feature of decision trees is their ability to break down a complex problem into smaller problems that are easier to understand and hopefully the combined outcome of the solution to the smaller problems is the solution being searched for. The decision tree approach falls into an array of techniques called multistage decision-making. Other approaches in the multistage decision making array are table look-up rules and sequential approaches. The basic idea of any multistage process is to break up a complex decision into a series of smaller but easier decisions<sup>[11]</sup>. Kweku and Bryson<sup>[10]</sup> describes a decision tree as a supervised knowledge discovery process in which prior knowledge regarding classes in the database is used to guide the discovery. Perhaps a very simple definition of decision trees is given in the Matlab<sup>®</sup> help file stating that a regression tree is a sequence of questions that can be answered as “yes” or “no” plus a set of fitted response values. Each question asks whether a predictor satisfies a given condition. Predictors can be continuous or discrete. Depending on the outcome of one question you either proceed to another question or arrive at a response value. An example of a decision tree is given in Figure 11.

***Figure 11: Typical structure of a decision tree***



Landgrebe<sup>[11]</sup> gives an excellent description of some relevant areas regarding decision trees.

1. A graph  $G = (V,E)$  consists of a finite non-empty set of nodes (or vertices)  $V$  and a set of edges  $E$ . If the edges are ordered pairs  $(v,w)$  then the graph is said to be directed.
2. A directed graph with no cycles is called a directed acyclic graph. A directed (or rooted) tree is a directed acyclic graph satisfying the following properties:
  - There is exactly one node called the root, which no edges enter. The root node contains all the class labels.
  - Every node except the root has exactly one entering edge
  - There is a unique path from the root to each node
3. If  $(v,w)$  is an edge in a tree, then  $v$  is called the father of  $w$  and  $w$  is called the son of  $v$ . If there is a path from  $v$  to  $w$  then  $v$  is a proper ancestor of  $w$  and  $w$  is a proper descendant of  $v$ .

4. A node without a proper descendant is called a leaf. All other nodes (except the root) are called internal nodes.
5. The depth of a node  $v$  in a tree is the length of the path from the root to  $v$ . The height of node  $v$  in a tree is the length of a largest path from  $v$  to leaf. The height of a tree is the height of its root. The level of node  $v$  in a tree is the height of the tree minus the depth of  $v$ .
6. An ordered tree is a tree in which the sons of each node are ordered.
7. A binary tree is an ordered tree such that 1) each son of a node is either distinguished as a left or a right son, 2) No node has more than one left son nor more than one right son.
8. The balance of a node  $v$  in a binary tree is  $(1+L)/(2+L+R)$ , where  $L$  and  $R$  are the number of nodes in the left and right subtrees of  $v$ . A binary tree is  $\alpha$ -balanced with  $0 < \alpha \leq 1$ , if every node has balance between  $\alpha$  and  $1-\alpha$ . A 0.5-balanced tree is said to be a complete tree.
9. The average number of layers from root to the internal nodes is referred to as the average depth of the tree. The average number of internal nodes in each level of the tree is referred to as the average breadth of the tree. In general, the average breadth of the tree will reflect the relative weight given to accuracy while the average depth gives the relative weight given to efficiency.

### 2.5.3.2 Advantages and disadvantages of decision trees

According to Zhang<sup>[37]</sup> and Landgrebe<sup>[11]</sup> decision trees have the following advantages.

1. Decision trees have no strict assumption for the distribution of the target variable.
2. There are no multicollinearity problems when input variables are highly correlated.

3. Decision trees deal with non-linear models easily without any variable transformation.
4. Decision trees can clearly indicate the relative importance of input variables with respect to their influences on the model target, and can also indicate the interactions between variables.
5. Decision trees can easily incorporate ordinal (such as low, medium, high), nominal (classes with names) and interval variables in the same model.
6. Global complex decision regions can be approximated by the union of simpler local decision regions at various levels of the tree.
7. Decision trees have the ability of choosing different subsets of features at different internal nodes of the tree so that the feature subset chosen optimally discriminates among the classes in that node. This feature may provide performance enhancement above that of single stage classifier.

According to Zhang<sup>[37]</sup> and Landgrebe<sup>[11]</sup> decision trees have the following disadvantages.

1. Overlap between classes in nodes is possible. This can cause the number of terminals to be much larger than the number of actual classes and thus reducing the efficiency.
2. Errors may accumulate from level to level in a large tree. One cannot optimise the accuracy and efficiency simultaneously. For any given accuracy there is a bound of efficiency that must be satisfied.
3. Designing an optimal tree might prove to be difficult. The performance of a decision tree strongly depends on how well it is designed.
4. Decision trees require a relatively large amount of training data.

5. Decision trees cannot express linear relationships in a simple and concise manner.
6. They cannot produce a continuous output due to their binary nature
7. Inherent replication of isomorphic subtrees<sup>[19]</sup>

### 2.5.3.3 Design of decision trees

Software developments during recent years have made the generation of decision trees very easy. The objective, however, in any data mining exercise or modelling process, is to create the most appropriate and accurate model. This is where the data analyst will have to create a wide variety of models and then using some performance criteria, choose the best model. According to Kweku and Bryson<sup>[10]</sup> the typical areas that will need focus during the design are the choice of minimum number of cases per leaf, splitting criteria, minimum number of cases per split, maximum number of branches from a node and the maximum depth of a tree. They also suggest that one should not only look at the last obtained decision tree as the best, but that using other performance criteria some of the rejected trees might also prove to be useful. Landgrebe<sup>[11]</sup> claims that the main objectives of a decision tree are to:

1. Classify correctly as much of the training sample as possible.
2. Generalise beyond the training sample so that unseen samples can be classified with as high accuracy as possible.
3. Be easy to update as more training data becomes available.
4. Have as simple a structure as possible.

The design of any decision tree can then be broken down into the following tasks, 1) The appropriate choice of the tree structure, 2) The choice of feature subsets to be used at each internal node and, 3) the choice of the decision rule or strategy to be used at each internal node.

Some of the common optimality criteria for tree design are: minimum error rate, min-max path length, minimum number of nodes in the tree, minimum expected path length and maximum average mutual information gain Landgrebe<sup>[11]</sup>. The optimisation is done by creating a series of decision trees. In practice the optimisation methods have the limitation of being computationally very expensive.

The various heuristic methods for construction of decision trees can be classified in four categories<sup>[11]</sup>:

1. Bottom-up approach where one starts with the information classes until one is left with a node containing all the classes.
2. Top-down approach where starting from the root node, using a splitting rule, classes are divided until a stopping criteria is met. The main issues in this approach are 1) Choice of splitting criteria, 2) Stopping rules and 3) Labelling the terminal nodes.
3. Hybrid method where one uses a bottom-up procedure to direct and assist a top-down procedure.
4. Tree growing-pruning approach where in order to avoid some difficulties in choosing a stopping rule, one grows the tree to its maximum size where the terminal nodes are pure or almost pure and then selectively prune the tree.

#### **2.5.3.4 Performance measures for decision trees**

It was mentioned earlier that choosing the most appropriate and accurate decision tree is very difficult. One would typically use one performance criteria to determine the most suitable decision tree. Kweku and Bryson<sup>[10]</sup> give an excellent review of performance

measures for decision trees. Their opinion is that the predictive accuracy rate is the most commonly used method of performance measure. In binary trees, accuracy rate can however, be a function of sensitivity (True positives/actual positives) and specificity (True negatives/actual negatives). Discriminating power of leafs is also a measure of the performance of a decision tree. Ideally one would like leafs that are totally pure (the probability of all classes except one is zero for each leaf). Stability is a very important aspect of any decision tree because it gives an indication of the accuracy stability when the decision tree is applied to other data sets than the training set. Most decision trees are created with prediction in mind and therefore the stability on new data sets is very important.

## **2.5.4 Fuzzy logic**

### **2.5.4.1 Introduction**

A very good overview of the history and origin of fuzzy systems is given by Brulé<sup>[6]</sup>. The following paragraph is a summary of his discussion. With the development of a concise theory of logic, the early philosophers like Aristotle developed the so-called “Laws of thought”. One of these laws, the “Law of excluded Middle” states that there are only true and false answers to any proposition or idea. Heraclitus opposed the idea saying that some things can be both true and false. Plato was the first to propose that there was a third region (beyond true and false) but Lukasiewicz proposed the first systematic alternative to the bi-valued logic of Aristotle. In the early 1900’s, Lukasiewicz described a three-valued logic along with the necessary mathematics. The third region can be described as “possible” and he assigned numerical values to the region between true and false. Knuth proposed a three-valued logic similar to Lukasiewicz but he used an integral range from  $[-1,0,1]$  rather than  $[0,1,2]$ . His approach did however not gain acceptance



and passed into relative obscurity. In 1965, Lotfi A. Zadeh published his seminal work called “Fuzzy sets”<sup>[36]</sup> which described the mathematics of fuzzy set theory and by extension fuzzy logic.

The idea of fuzzy logic control was born out of allowing partial set membership rather than crisp set membership<sup>[9]</sup>. Fuzzy logic is not a data mining technique but rather a control idea. The first fuzzy logic controller was built in England in 1973 and the aim was to stabilize a small steam engine Sowell<sup>[29]</sup> The products of data mining like the discovered patterns in data, or discovered rules can be applied in a fuzzy control system in order to control a certain process. Fuzzy logic represents an opposite approach to conventional process control by not using a mathematical model of the process to predict response, but is rather based on “fuzzy” rules or sets often based more on human experience than mathematical fundamentals. According to Pentz<sup>[22]</sup>, fuzzy logic can be invested with human experience instead of a mathematical model. Yet the control can be just as good or even better and problems that were impossible to control in the past are suddenly possible with fuzzy logic. One of the main application areas of fuzzy logic is where historically empirically collected data is used to determine the rules for the fuzzy control system. Some advantages of fuzzy logic are given by the newsletter of the Seattle Robotics Society<sup>[9]</sup>:

- It is inherently robust because it does not require precise, noise-free input and can even be programmed to fail “safely” should a problem occur. The output control is a smooth control function despite the wide range of input variables.
- The fuzzy logic system is based on user-defined functions and can therefore be easily adapted to new scenarios.

- Fuzzy logic is not limited to a few feedback input parameters and one or two control outputs, nor is it necessary to measure or calculate any rate-of-change parameters in order for it to be implemented. Any input data that provides some indication of the system performance is sufficient. This allows the sensors to be imprecise and inexpensive keeping the overall costs down and complexity low.
- Because of the rule-based operation, any reasonable number of inputs can be processed and numerous outputs generated. However, care should just be taken that the amount of input parameters does not create a lot of complex rules. Another option is to break the control system up into smaller separate fuzzy control systems.
- Fuzzy logic can control non-linear systems that would be difficult or impossible to model mathematically.

A simple example to illustrate the idea of fuzzy logic is to consider the cast strand width error of a continuous casting machine. Let's assume the width error set is defined as  $[-20,20]$ . Intuitively we can divide the interval into regions to which we can connect a linguistic description like "Very narrow", "Narrow", "Wide" and "Very Wide". The regions are defined as:

Very Narrow =  $[-20,-10)$

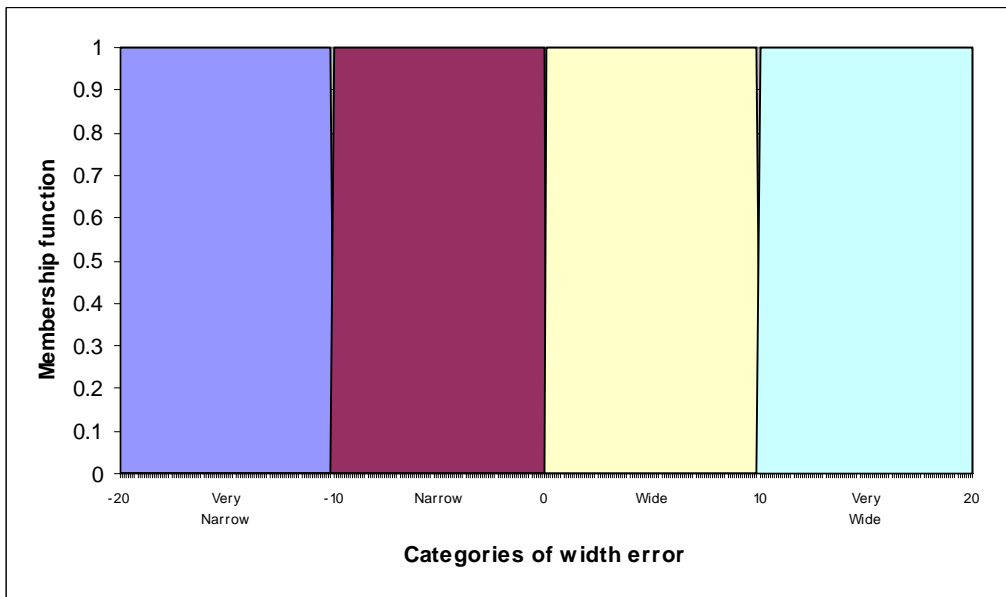
Narrow =  $[-10,0)$

Wide =  $[0,10)$

Very Wide =  $[10,20]$

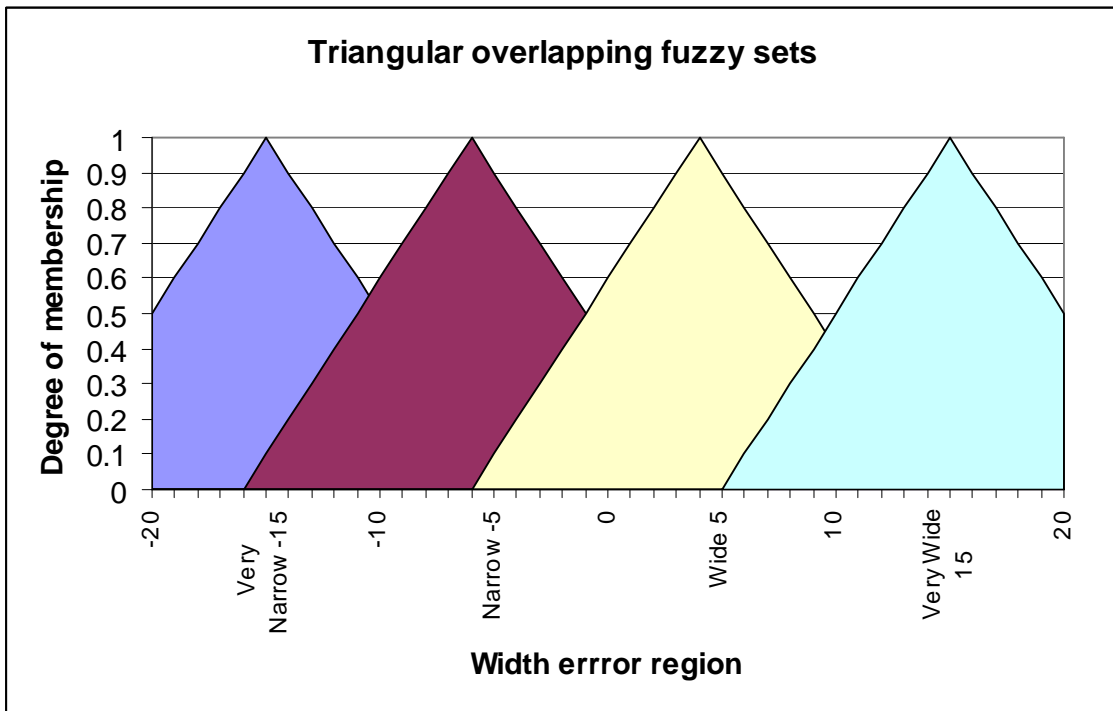
If classical set theory is applied and True = 1 and False = 0, then a graphical representation of the above set would be given in Figure 12.

***Figure 12: Bivalent Sets to characterize the strand width error***



The problem with classical set theory is the sudden sharp change from one “region” to another. For example, if a slab is 9mm wide from aim it will be called “wide”, if it is 10mm wide from aim it will be called “Very Wide”. Intuitively one can see the problem with applying classical set theory to real world problems. If one would apply fuzzy set theory that allows a more continuous transition from one region to another, together with allowing shared membership in different regions, one can apply for example triangular membership functions and the slab width error domain can be covered by the membership as illustrated by Figure 13.

**Figure 13: Triangular membership functions illustrating fuzzy set theory**



It can be seen that there is smooth transition between areas. Fuzzy set theory therefore makes it possible to characterise certain “fuzzy” areas where the decision between true and false is not so clear.

In general, a fuzzy system consists out of four components: Fuzzy rule base, fuzzy inference engine, fuzzifier and defuzzifier<sup>[33]</sup> that will be discussed in detail.

**2.5.4.1.1 Fuzzy rule base**

A fuzzy rule base consists out of a set of fuzzy IF-THEN rules. It is the heart of the fuzzy system because usually all the other components play a role to implement these rules. In general these rules have the following form:

$$Ru^{(k)} : \text{IF } x_1 \text{ is } A_1^1 \text{ and } \dots \text{and } x_n \text{ is } A_n^1, \text{ THEN } Y \text{ is } B^{(k)} \dots\dots\dots (9)$$

where  $A_i^1$  and  $B^{(k)}$  are fuzzy sets in  $U_i \subset R$  and  $V \subset R$ , respectively and  $\mathbf{x} = (x_1, \dots, x_n)^T \in U$  and  $y \in V$  are the linguistic input and output. The rules from equation (9) are called

canonical fuzzy IF-THEN rules because they include many other parts of fuzzy rules and special cases such as:

- Partial Rules,
- “Or Rules”,
- Single Fuzzy Statements,
- “Gradual Rules”, and
- Non-Fuzzy rules (conventional production rules).

#### **2.5.4.1.2 Fuzzy Inference Engine**

In a fuzzy inference engine, fuzzy logic principles are applied to combine the fuzzy IF-THEN rules in the fuzzy rule base into a mapping from a fuzzy set  $A'$  in  $U$  to a fuzzy set  $B'$  in  $V$ . There are basically two ways to infer 1) Composition-based inference and 2) Individual-rule based inference. In compositional-based inference all the rules in the fuzzy rule base are combined into one fuzzy relation in  $U \times V$ . This single relation is then viewed as one IF-THEN rule. There are two ways to view the rules that will influence the method of inference. The first view is to see the rules as individual conditional statements. A reasonable operator to use in such a case would be union (see 2.2.3.3). The second view is to see the rules as strongly-coupled conditional statements meaning that all the conditions of all the rules must be satisfied in order for the whole set of rules to have an impact. A reasonable operator for this view would be intersection (see 2.2.3.3). There are other inference methods which will not be discussed in this text but all details can be found in Wang<sup>[33]</sup>. Other inference methods are 1) Product Inference Engine, 2) Minimum Inference Engine, 3) Lukasiewics Inference Engine, 4) Zadeh Inference Engine, and 5) Dienes-Rescher Inference Engine.

### 2.5.4.1.3 Fuzzifiers

The fuzzifier is defined as the mapping from a real-valued point  $x^* \in U$  to a fuzzy set  $A'$  in  $U$ . One important criteria for a fuzzifier is that it should be able to suppress any noise in the input values. Three typical fuzzifiers are the following:

#### 2.5.4.1.3.1 Singleton Fuzzifier

The singleton fuzzifier maps a real-valued point  $x^* \in U$  into a fuzzy singleton  $A'$  in  $U$  with membership function value 1 at  $x^*$  and 0 at all other points.

$$\mu_{A'}(x) = 1 \text{ if } x = x^* \dots\dots\dots(10)$$

$$\mu_{A'}(x) = 0 \text{ otherwise}$$

#### 2.5.4.1.3.2 Gaussian Fuzzifier

The Gaussian fuzzifier maps  $x^* \in U$  into fuzzy set  $A'$  in  $U$  with the following membership function:

$$\mu_{A'}(x) = e^{-\frac{(x_1-x_1^*)^2}{a_1}} \oplus \dots \oplus e^{-\frac{(x_n-x_n^*)^2}{a_n}} \dots\dots\dots(11)$$

where  $a_i$  are positive parameters and the t-norm  $\oplus$  is usually chosen as the algebraic product.

#### 2.5.4.1.3.3 Triangular Fuzzifier

The triangular fuzzifier maps  $x^* \in U$  into fuzzy set  $A'$  in  $U$  using the membership function

$$\mu_{A'}(x) = \left(1 - \frac{|x_1-x_1^*|}{b_1}\right) \oplus \dots \oplus \left(1 - \frac{|x_n-x_n^*|}{b_n}\right) \quad \text{if } |x_i-x_i^*| \leq b_i, i = 1,2,3\dots n \dots\dots(12)$$

$$\text{else } \mu_{A'} = 0$$

with  $b_i$  positive parameters and  $\oplus$  the t-norm chosen as the algebraic product.

#### 2.5.4.1.4 DeFuzzifiers

The defuzzifier is defined as a mapping from fuzzy set  $B'$  in  $V \subset R$  (which is the output of the inference engine) to a crisp point  $y^* \in V$ . Conceptually the task of the defuzzifier is to calculate a point in  $V$  that represents the fuzzy set  $B'$ . The concept is almost similar to calculating the mean value of a series of random values. Because  $B'$  was constructed using some technique, a choice must be made how to best represent point  $y^*$  in  $V$ . The following criteria are recommended to choose the best scheme:

- Plausibility – The point  $y^*$  should represent  $B'$  from an intuitive point of view.
- Computational simplicity
- Continuity – A small change in  $B'$  should not result in a large change in  $y^*$ .

The common defuzzification techniques are discussed next:

##### 2.5.4.1.4.1 Centre of Gravity Defuzzifier

The centre of gravity defuzzifier specifies the  $y^*$  as the centre of the area covered by the membership function of  $B'$ , and is given as:

$$y^* = \frac{\int_V y \mu_{B'}(y) dy}{\int_V \mu_{B'}(y) dy} \dots\dots\dots(13)$$

where  $\int_V$  is the conventional integral. The advantage of the centre of gravity method lies in the plausibility but the disadvantage is that it is computationally intensive.

##### 2.5.4.1.4.2 Centre Average Defuzzifier

The fuzzy set  $B'$  is the union or intersection of  $M$  fuzzy sets (due to the inference technique) hence a good average will be the average weights of the different fuzzy sets. The heights of the  $M$  sets can be taken as the weights of the  $M$  sets. Let  $\bar{y}^l$  be the centre of the  $l$ 'th fuzzy set and  $w_l$  be its height, then the centre average defuzzifier will be given by:

$$y^* = \frac{\sum_{l=1}^M y_l w_l}{\sum_{l=1}^M w_l} \dots\dots\dots(14)$$

The centre average defuzzifier method is the most commonly used defuzzifier because it is intuitively plausible and computationally simple.

**2.5.4.1.4.3 Maximum Defuzzifier**

Conceptually the maximum defuzzifier chooses the  $y^*$  as the point in  $V$  at which  $\mu_{B'}(y)$  achieves its maximum value. This method is not highly regarded in literature because the feeling is that the results may be contradictory to the intuition of maximum membership. Another disadvantage is that small changes in  $B'$  may result in large changes in  $y^*$ .

**2.5.4.2 Membership functions**

The generation of the applicable rules for the fuzzy controller usually depends on the system to be modelled. The rules can be based purely on experience or the rules can be generated with data mining processes from historical empirical data. It should also be kept in mind that a rule can be a fundamental mathematical equation and all rules do not have to be derived from empirical data. This freedom of rule generation is a big advantage of fuzzy systems. The implementation of the rules is usually based upon membership functions. A description of a membership function according to Kaehler<sup>[9]</sup> is that it is a graphical representation of the magnitude of participation of each input. It associates a weighting with each of the inputs, defines functional overlaps between inputs and ultimately determines an output response. The rules are the input membership values as weighting factors to determine their influence on the fuzzy output sets of the final output conclusion. Once the functions are inferred, scaled, and combined, they are defuzzified into a crisp output that drives the system. There are different membership



functions associated with each input and output response. Some features to note about membership functions are:

- Shape – can be triangular, bell, trapezoidal, hasesine and even exponential. More complex functions require more computing resources.
- Height – usually normalised to 1
- Width – Base of the function
- Shouldering – Locks height at maximum if an outer function. Shouldered functions evaluated as 1.0 past their centre.
- Centre – Points in the middle of the membership function
- Overlap – Overlapping of membership functions.

### 2.5.4.3 Fuzzy set Operations

As is the case with classical set theory, fuzzy set theory also have set operations. The following summary was adapted from Anon3<sup>[3]</sup>.

#### 2.5.4.3.1 Union

The membership function of the union of two fuzzy sets A and B with membership functions  $\mu_a$  and  $\mu_b$  respectively is defined as the maximum of the two individual membership functions. This is called the maximum criterion denoted by

$$\mu_a \cup b = \text{Max}(\mu_a, \mu_b) \dots \dots \dots (15)$$

The union operation in fuzzy set theory is equivalent to the **OR** operator in Boolean algebra.

### 2.5.4.3.2 Intersection

The membership function of the intersection of two fuzzy sets A and B with membership functions  $\mu_a$  and  $\mu_b$  respectively is defined as the minimum of the two individual membership functions. This is called the maximum criterion denoted by

$$\mu_a \cap b = \text{Minimum}(\mu_a, \mu_b) \dots \dots \dots (16)$$

The union operation in fuzzy set theory is equivalent to the **AND** operator in Boolean algebra.

### 2.5.4.3.3 Complement

The membership function of the Complement of a Fuzzy set A with membership function  $\mu_a$  is defined as the negation of the specified membership function. This is called the negation criterion denoted by

$$\mu_{\bar{A}} = 1 - \mu_A \dots \dots \dots (17)$$

The Complement operation in fuzzy set theory is equivalent to the **NOT** Boolean operator.

## 2.6 Summary of literature survey

The literature survey covered topics from data mining techniques and the theory behind the cast width error and chemistry relationship on 12% chrome ferritic non-stabilised stainless steel. The data mining techniques that were covered in detail included statistical regression and general statistical graphical techniques, decision trees and fuzzy logic. The cast width error to chemistry relationship was discussed by first looking at the fundamental metallurgical principles to why such a relationship might be possible. The reason for this relationship is mainly due to the fact that 12% chrome non-stabilised

ferritic stainless steel have a dual phase region between austenite and ferrite in the temperature range 1250°C and 850°C. The hot strength of austenite and ferrite is different with austenite having more hot strength than ferrite. The ratio between austenite and ferrite is fluctuating as a function of the heat chemistry. The variation between heats can lead to width error variation. The second issue with regards to the width error to chemistry relationship that was discussed deals with the proposed mechanism of the cast width error. Plant measurements and visual observation seemed to indicate that the width error was a result of creep from the complete shell combined with bulging on the narrow sides.

## Chapter 3

### Data preparation for modelling

This section deals with one of the most important factors of any data mining exercise, gathering appropriate, relevant and accurate data to be used in the modelling process. Unfortunately it can easily be the case of “garbage in, garbage out”. This chapter consists out of four sections covering the themes of gathering the data, preparing the data for modelling, clean-up of the data set and selection of significant variables.

#### 3.1 Gathering the data

The data that was used for this project is “live” plant data. The plant databases are updated continuously as new products are manufactured. Through experience over the years, it has been attempted to capture as many of the critical process parameters as possible. The data is usually stored on the databases for a period of three years. The data relevant to this project are the specific chemistry analysis of a specific heat of steel and the strand width of the heat. The chemistry of each heat is analysed by taking a sample from the molten metal and then analysing it in the laboratory for chemical composition. The chemical analysis of each heat is then entered onto the corporate database called “P4”. The width of the strand is measured continuously by means of a lazer at the exit of the casting machine. The measured value is stored once every second. The cast length is 32 meters from mould exit to where it is measured by the lazer. The strand temperature is still at approximately 900°C when the lazer measures it and consequently a conversion factor (shrinkage factor) is used to transform the hot width to a cold width measurement for a specific steel grade (see Appendix D). The shrinkage factor is unique for the different steel grades and was determined empirically by comparing the hot lazer

measurements of a slab to the cold measurements as taken in front of the reheat furnace before hot rolling. The slabs are charged cold into the reheat furnace at Columbus Stainless and the slab width is measured with a single probe in the middle of the slab. The shrinkage factors were not determined as part of this project. Experience over the years have indicated that the shrinkage factors are very accurate and converts the hot slab width to cold width measurements very accurately. The cold width measurement that is continuously measured by the lazer (converted from hot width) is then stored on the steelplant database as an average value per slab (The strand that is continuously cast is cut into sections of typically 12 meter lengths known as slabs). This average value of the slab width as measured by the lazer is then taken as the width measurement of the slab. Appendix G gives a typical relationship between the raw hot lazer data (actual measured), cold lazer data (including shrinkage factor) and the value representing the slab width for a single slab. Also included in Appendix G is the standard deviation for the raw lazer data for a specific slab, indicating that the standard deviation for the raw lazer measurements is very small in comparison to the standard deviation between the different slabs. After the slabs have been cast, they are hot rolled. In order to be hot rolled the slabs must first be reheated to typically above 1000°C. The slabs are measured with a single probe when entering the reheat furnace. The slabs are typically below 200°C when measured by the probe and hence can be assumed to be the “cold” width of the slab. This “cold” measurement of the slabs is also used to calibrate the conversion factor used for converting the “hot” lazer width measurement to a “cold” width measurement. The fact that two independent “cold” width measurements are available per slab, makes it possible to check the validity and accuracy of each measurement station. If both measurements are used, then measurement “drifts” and erroneous readings can be identified. The data for this project was extracted from the databases

using Microsoft Access. The variables that were extracted from the databases included the following: Slab width as measured by the lazer, slab width as measured in front of the reheat furnace, absolute levels of Carbon, Nitrogen, Manganese, Nickel, Chrome, Silicon and Titanium and then also the calculated parameters from the chemistry including the AC1 temperature, Amax, CR95, Gmax and ferrite factor (see Appendix C for descriptions). The chemistry parameters were chosen as those that is believed to have an influence on the ferrite/austenite ratio as discussed in chapter 2.

### **3.2 Preparing the data for modelling**

The data was split into two sets. A training set and a validation set. The same training and validation sets were used to train/derive and compare the accuracy of the different models. The training set that was used consisted out of the data for the 12 % chrome ferritic non-stabilised material for the time period of 2002/12/28 to 2005/01/12. Every record represents the chemistry composition of a heat. The average weight of a heat is 100 ton and therefore the training set represents approximately 24900 tons of steel with 249 combinations of unique chemistry (heats), with the resulting width error of each chemistry. Every heat that is cast results in approximately five slabs of steel. Every slab has a unique number that is used for tracking purposes in the plant. This means that all the slabs from one heat have the same chemistry but have different processing parameters at the continuous casting machine. Although the width measurement is done continuously while the strand is cast, the width measurement is recorded per slab. The only way to get an indication of the resulting width error of a certain chemistry combination is to work with the individual slabs of the same heat. In this project the width error associated with a certain heat composition was calculated by taking the average of the width error of the slabs associated with the heat. The validation data set

that was used was taken from all the 12% chrome (non-stabilised) ferritic material cast between 20/01/2005 and 21/04/2006 with the standard cooling practice (without secondary cooling modification). This resulted in 126 unique chemistry combinations and associated width error.

### **3.3 Data Clean-up**

Accurate data that is a true reflection of plant conditions are absolutely critical and necessary for a successful model to be developed. Due to the fact that “live” plant data was used for this project, it was critical to filter inaccurate data out of the data set. Inaccurate plant data can usually be attributed to inaccurate measurements in the plant or data base anomalies like unpopulated fields. The accuracy of the lazer was checked by comparing the lazer measurements with the probe measurements taken before the reheat furnace. A common problem in the plant was dirty lazer lenses that resulted in inaccurate readings from the lazer captured on the database. The average width error for each heat as measured by the lazer were compared to the average width error as measured by the probe, and the records where the difference was more than 10mm were discarded as being untrustworthy measurements. Records where unpopulated fields were encountered were deleted. The inter quartile range (IQR) method was used to identify outliers with the intention of deleting them, but it was realized that these outliers were necessary to train an accurate model because they define the extreme points. Due to operational issues it happened sometimes that a heat with chemistry similar to the “outliers” must be cast, and in these cases it was necessary that the model must be able to recognise and handle these chemistries that will result in severe strand width deviations. After the two data sets were “cleaned”, the training set consisted out of 249 records representing 249 unique chemistries and their associated width errors. The validation set had 126 records.

### 3.4 Selection of significant variables

Due to the fact that the chemistry parameters were chosen as those that would have an influence on the phase balance between austenite and ferrite (austenite and ferrite stabilisers), it was necessary to find the parameters that indicate the strongest correlation with the width error. The correlation coefficient was determined between each parameter and the width error. The parameters with the strongest correlations were chosen as the parameters that were used in the modelling process. A parameter that was significant through plant experience was the sum of the Carbon and Nitrogen content. Table 2 gives the results of the correlation coefficient between the parameters and the width error.

*Table 2: Correlation Coefficients between parameters and width error.*

<b>Parameter</b>	<b>Correlation Coefficient</b>
Carbon	-0.208
Nitrogen	-0.334
Chrome	-0.011
Manganese	-0.102
Nickel	-0.279
Silicon	0.199
Titanium	0.135
AC1	0.464
Amax	-0.394
CR95	0.433
Gmax	-0.479
Ferrite factor	0.427
Carbon + Nitrogen	-0.362

The positive and negative signs for the correlation coefficient indicate if the strand width tends negative or positive with an increase in the parameter. A reasonable cut-off was chosen as 0.35 correlation coefficient. From Table 2 the correlation coefficients that complied with this criteria are as follows in descending order:



Gmax  
AC1  
CR95  
Ferrite factor  
Amax  
Carbon + Nitrogen

These parameters were therefore chosen as input parameters for the modelling processes.

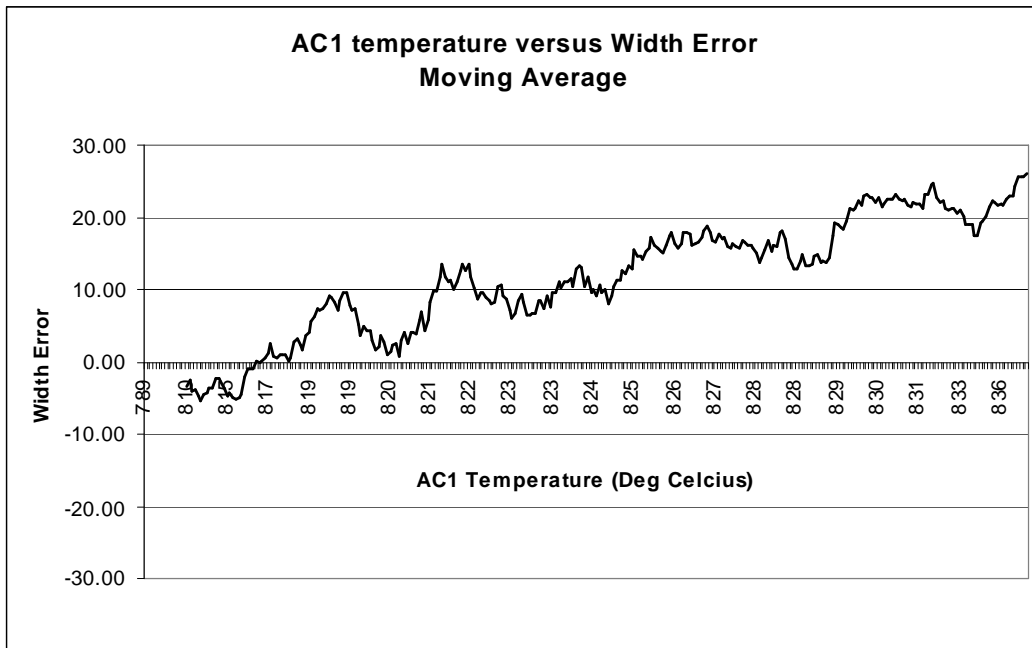
They are also the same parameters that are used in the model that is currently running live in the plant.

### **3.5 Visual representation of data**

The graphs in this section were adapted from De Beer<sup>[7]</sup>. The graphs gave a visual representation of the relationship of the different parameters versus the strand width error. It can visually be seen from Figures 14 to 19 that there is a relationship between the parameters and the strand width error of 12% chrome non-stabilised ferritic stainless steel.

Figure 14 indicates that higher AC1 values lead to a greater positive width error. The AC1 temperature influences the width error due to its relationship with the ferrite start temperature as calculated by the MEDUSA model<sup>[30]</sup>. The higher the ferrite start temperature, the higher the AC1 temperature. A high AC1 temperature therefore means that ferrite will form early in the continuous casting process increasing the soft ferrite fraction of the shell and thus increasing the probability of shell deformation.

**Figure 14: Relationship of AC1 temperature to strand width error.**



From Figure 15 it can be seen that the higher the Amax percentage, the narrower the width error tends to be. This can be explained by the fact that the higher the Amax value is, the more austenite will be present in the shell that will strengthen it, hindering creep and bulging.

**Figure 15: Relationship of Amax to strand width error.**

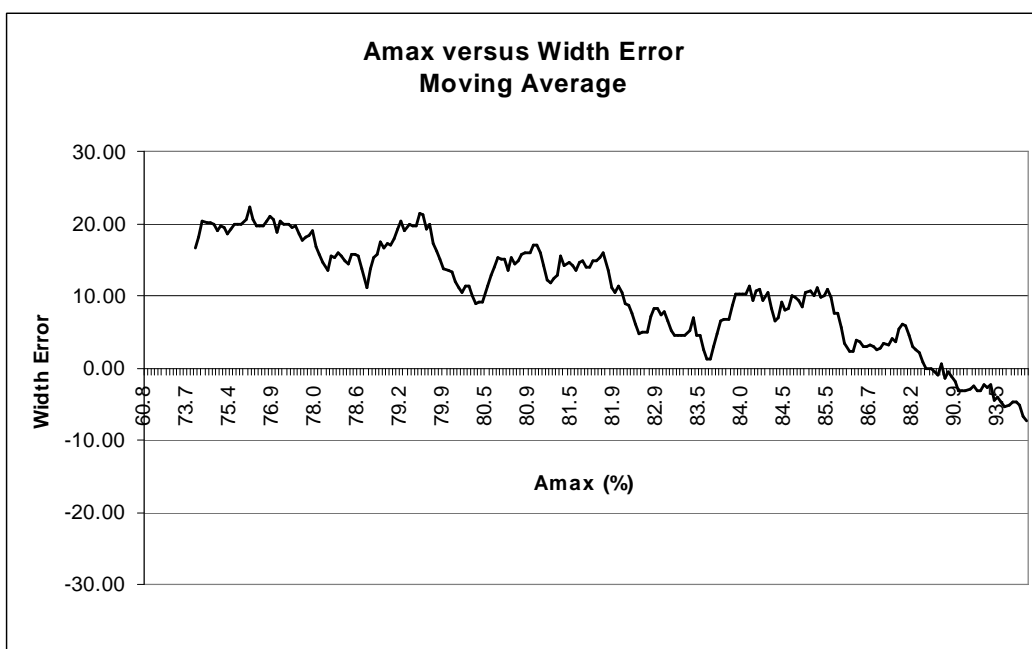
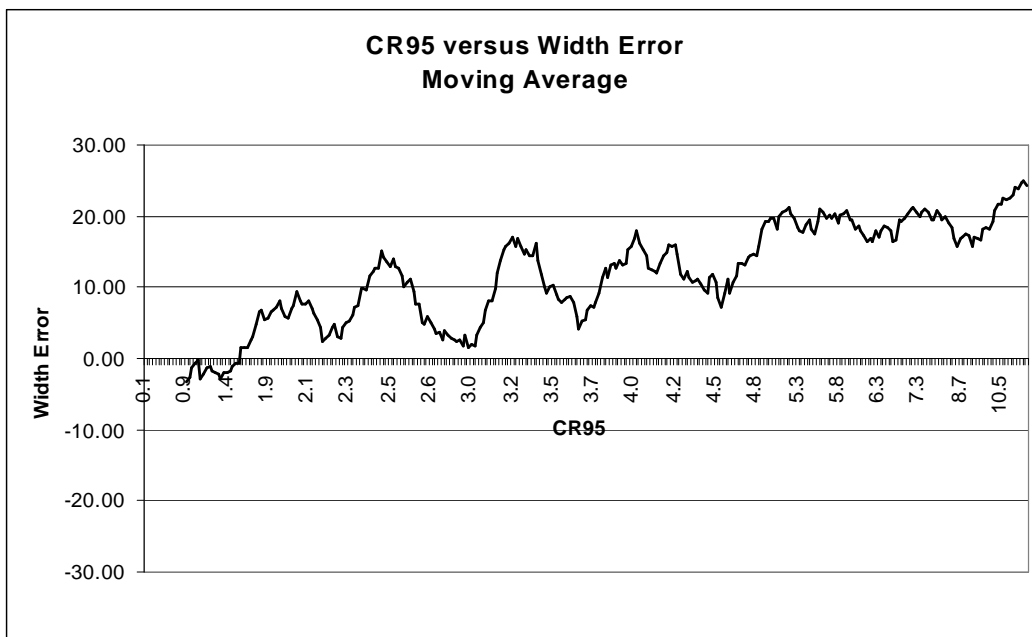


Figure 16 gives the relationship between the width error and the CR95 value. It is evident that a higher CR95 value favours a larger positive width error. This relationship is explained in terms of transformation rate of austenite to ferrite during solidification. A heat with a high CR95 value will transform the austenite present at 1000°C faster to ferrite, than a heat with a lower CR95 value. If the transformation rate from austenite to ferrite is high (under constant cooling) then the austenite fraction of the structure will decrease (transform to ferrite) quickly and relatively high up in the casting machine bow. This will leave room for a weaker shell that might be prone to deformation. The opposite is true for a slow transformation of austenite to ferrite that will tend to result in a stronger shell for a longer period.

**Figure 16: Relationship of CR95 to strand width error.**



The Gamma max trend versus the strand width error is depicted in Figure 17. As expected it followed the same trend as the Amax. The higher the Gamma max is, the more difficult it would be for the shell to deform due to the higher austenite fraction.

**Figure 17: Relationship of Gamma max to strand width error.**

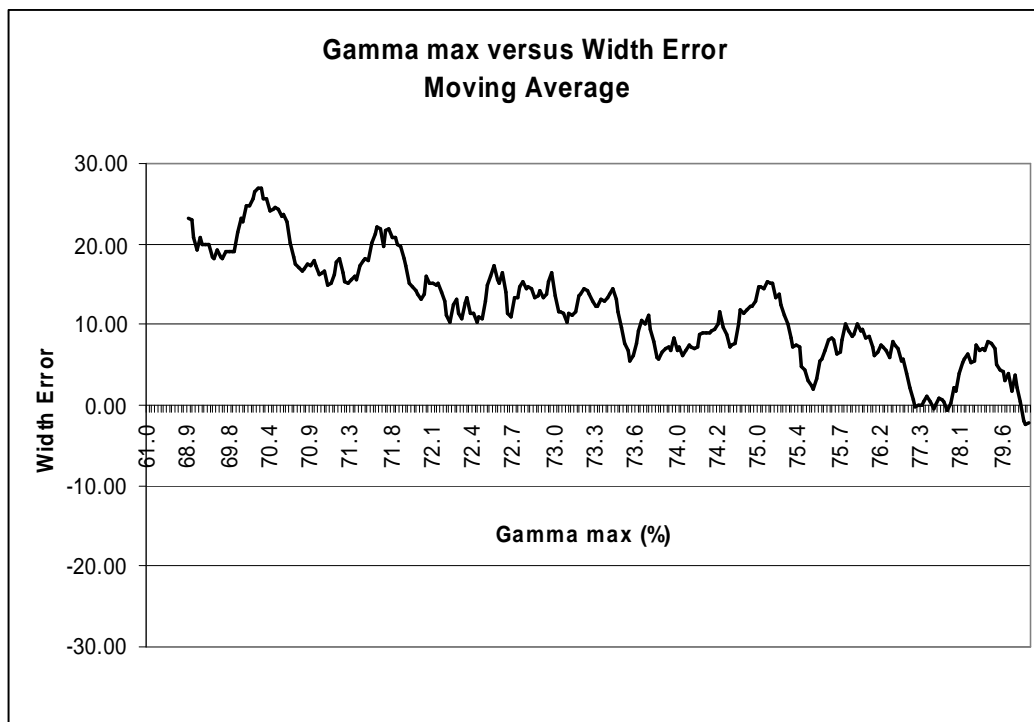


Figure 18 indicates that the higher the ferrite factor is, the wider from aim the strand width tends to be. If the ferrite volume fraction is high, the shell will tend to be softer and therefore tend to deform more easily. Figure B1 in Appendix B indicates the relationship between the ferrite factor and the measured austenite start temperature  $A_{r5}$  for a 12% chrome non-stabilised ferritic stainless steel. It can be seen that the transformation to austenite starts at earlier times and the transformation temperature is increased for a chemistry with a low calculated ferrite factor. Earlier transformation and transformation at higher temperatures for heats with a low ferrite factor, versus heats with a high ferrite factor mean that the shell will be stronger for heats with a low ferrite factor versus heats with a high ferrite factor.

**Figure 18: Relationship of ferrite factor to strand width error.**

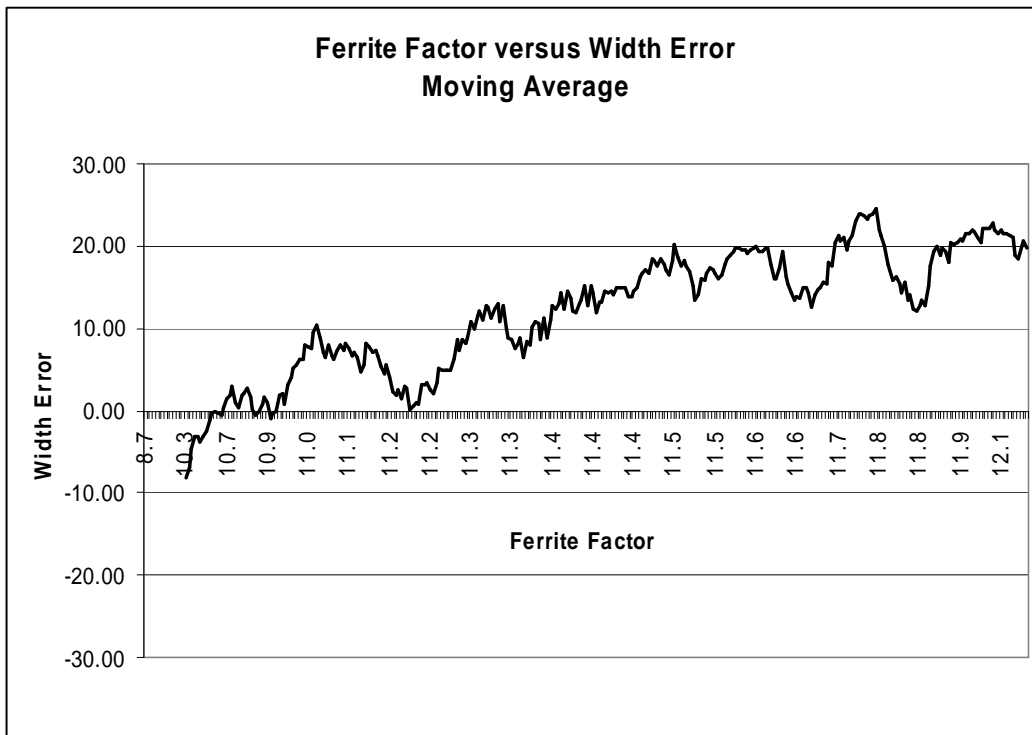
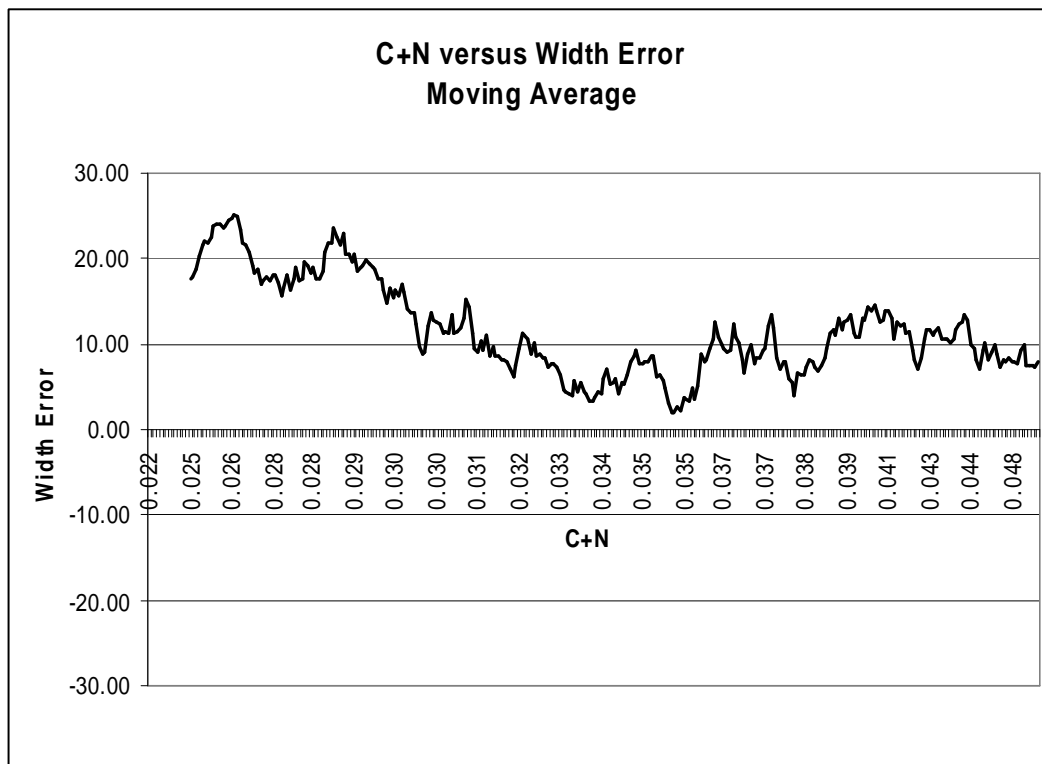


Figure 19 indicates that the width deviation from the aimed width tends to decrease (from being wide) with an increase in the parameter C + N. This is due to the fact that both carbon and nitrogen are austenite stabilisers and therefore an increase in these parameters will change the ferrite to austenite ratio in the shell, favouring austenite, and this will result in a stronger shell.

**Figure 19: Relationship of Carbon and Nitrogen to strand width error.**



The different modelling techniques will first be applied to the training set to derive applicable models. The quality of the derived models in terms of accuracy on the training data set will be discussed. Comparison of the different models will be done in chapter 4 when applied to the validation set of data.

### 3.6 Model evaluation criteria

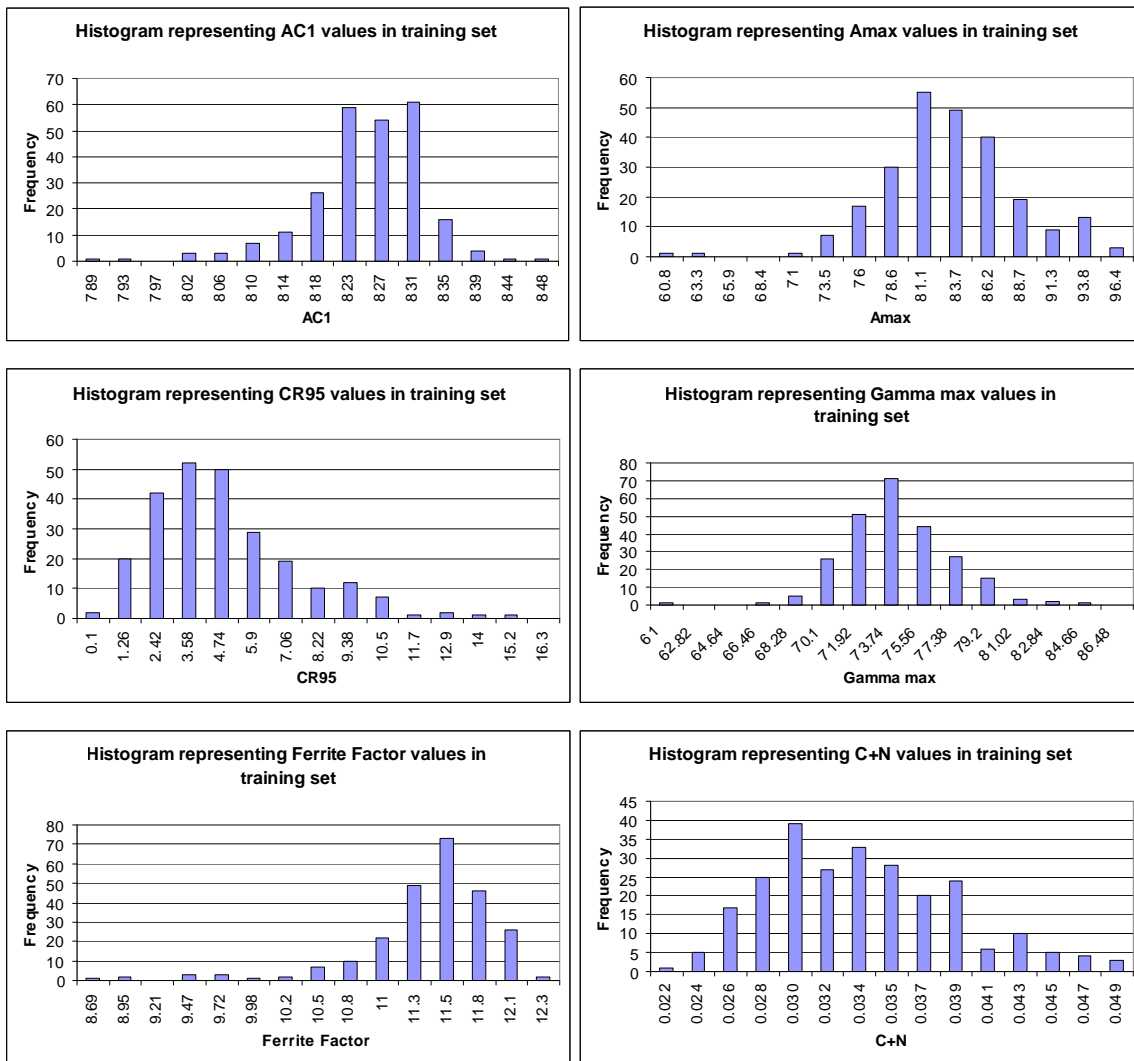
The acceptable width criteria is defined as a width error (actual cold slab width – aim cold cast width) of between 0mm and 15mm. Width errors of <0mm means negative values and the slab is narrower than aim. Width errors >15mm represent slabs that are considered to be excessively wide. Narrow and excessively wide slabs cause production problems during subsequent processes. The results are grouped into three groups with group one representing the narrow slabs, group two representing the acceptable slabs and group three representing the excessively wide slabs. The performance of the models in

terms of accuracy will be determined on its ability to predict the width error within one of the defined groups. The measure of accuracy will be the percentage of heats correctly predicted for each group. An example illustration is if it is assumed that a population consists out of 100 records with 20 records having width errors less than zero (group 1), 50 records having width errors between 0 and 15mm (group 2) and 30 records having width errors more than 15mm (Group 3). When a model is derived and predictions are made, then for each of the 100 records there will be an actual value and a predicted value. If both the actual and predicted value falls within the same group, then the model predicted accurately. If, for example the records in group 1 is considered, then there are 20 records each with its own actual value and predicted value. Let's assume 15 records of the 20 records have both actual and predicted values inside group 1, then the model is evaluated as being 75% correct in group 1.

### **3.7 Modelling using statistical regression**

Figure 20 gives the distribution of the parameter values using histograms. This summary gives a high level overview of the training set that will be used to derive the different models. This statistical technique falls into the category of non-computationally expensive statistical methods. No predictive model can be derived using the histograms and is only useful to give a better higher level understanding of the data under review.

**Figure 20: Histograms of parameter values of the training data set.**



It can be seen from Figure 20 that the Amax, Gamma max and AC1 indicate normal distributions. Both CR95 and ferrite factor indicate a portion that is normal distributed but both graphs indicate a long tail to one side. It seems as if the outlier chemistries influence the ferrite factor and CR95 values more than it did the other parameters.

Response Surface Methodology (RSM) is a tool for understanding the quantitative relationship between multiple input variables and one output variable. A quadratic response surface was used to model the width error using the mentioned input parameters. The statistical toolbox in Matlab<sup>®</sup> was used to derive the RSM models. The



general equation describing the quadratic response surface of an input vector  $\mathbf{x}$  and response vector  $\mathbf{y}$  has the form:

$$\begin{aligned}
 y = & \beta_0 + \beta_1 x_1 + \beta_2 x_2 \dots + \beta_n x_n \dots \dots \dots (\text{linear}) \\
 & + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \dots \dots + \beta_{1n} x_1 x_n + \dots + \beta_{(n-1)n} x_{(n-1)} x_n \dots \dots (\text{Interaction}) \dots \dots (22) \\
 & + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{33} x_3^2 + \dots \dots + \beta_{nn} x_n^2 \dots \dots \dots (\text{Quadratic})
 \end{aligned}$$

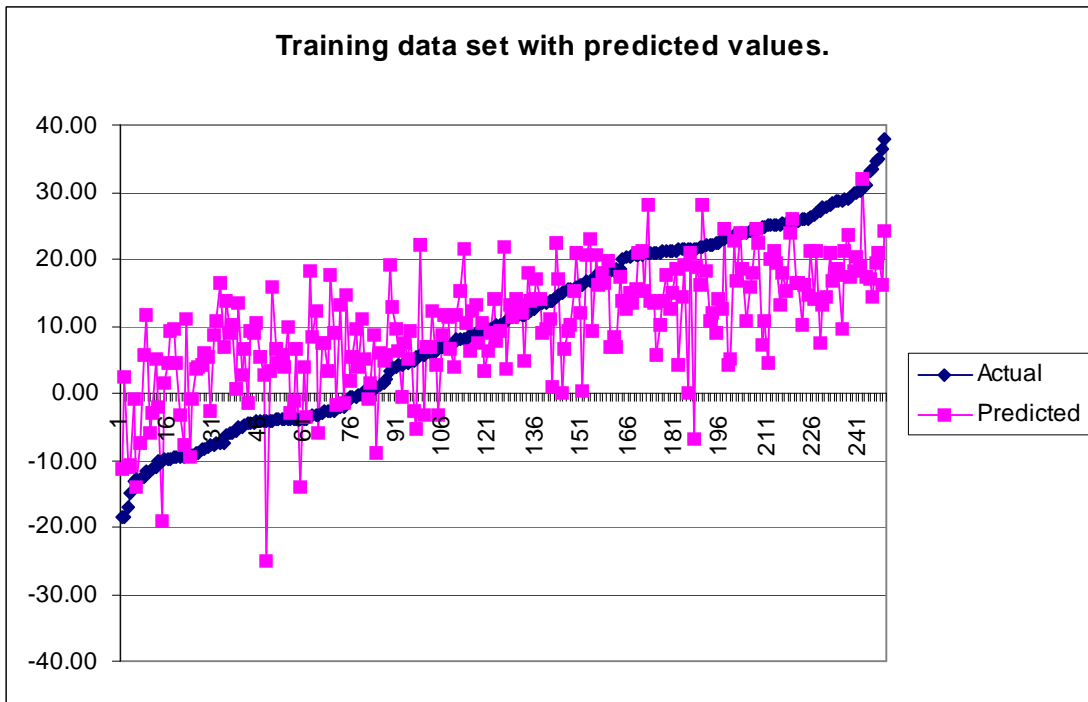
The “rstool” in Matlab<sup>®</sup> was used to derive the quadratic response surfaces used in this section. The “rstool” has the functionality of deriving separate response surface models including the linear section alone, linear and quadratic alone and linear and interaction terms separately and a response surface including the linear, interaction and quadratic terms. In total, four response surface models were developed, one model representing each of the before mentioned models.

Equation 23 represents the derived linear equation.

$$\begin{aligned}
 \text{WidthError} = & 0.374 \times AC1 - 0.182 \times A \text{ max} + 0.664 \times CR95 - 0.561 \times G \text{ max} + 1.683 \times \text{FerriteFactor} \dots (23) \\
 & - 431.486 \times (C + N) - 249.854
 \end{aligned}$$

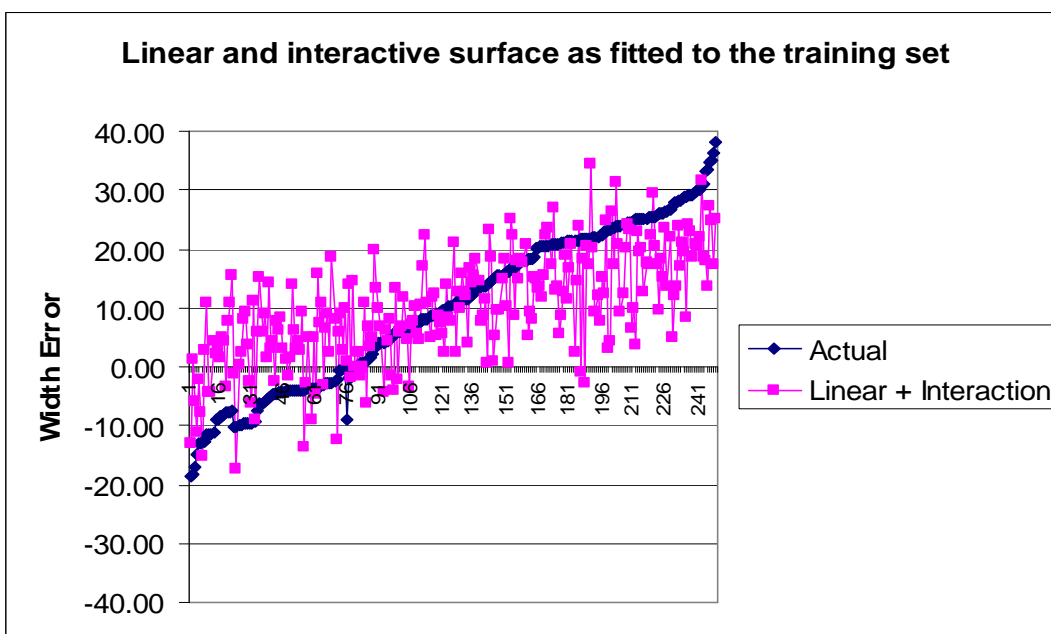
Figure 21 indicates the training set together with the predicted values using equation 23.

**Figure 21: Training data set with predicted values using linear statistical regression model**



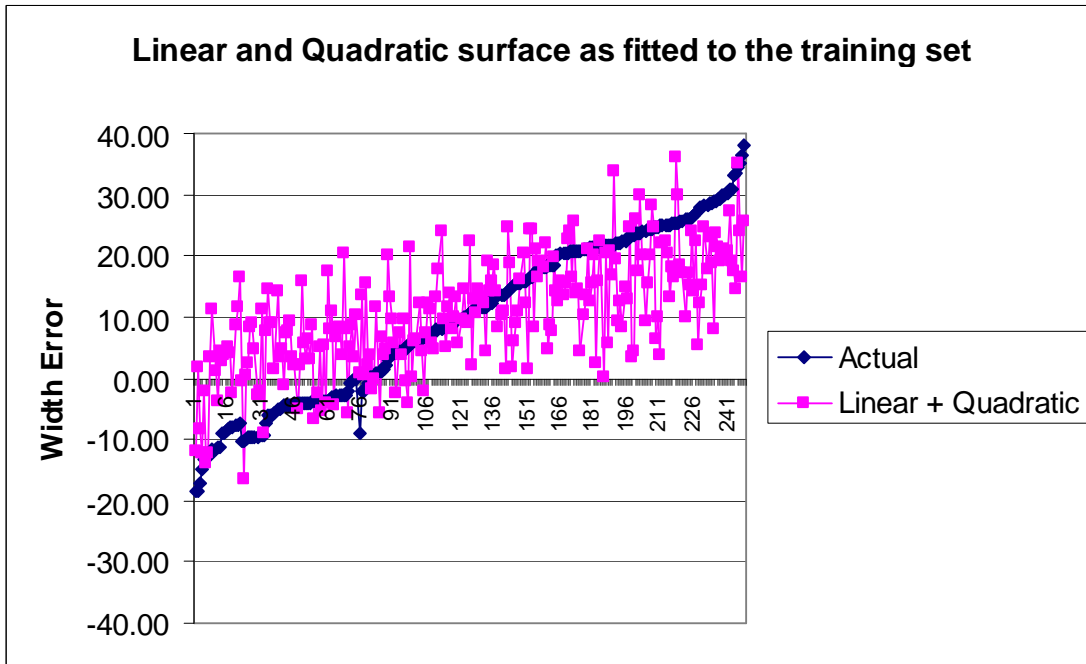
The derived equation representing the linear and interaction terms combination is given in Appendix E. Figure 22 indicates the relationship of the predicted values versus the actual values.

**Figure 22: Linear and interactive term surface versus actual values**



The derived equation representing the linear and Quadratic terms combination is given in Appendix E. Figure 23 indicates the relationship of the predicted values versus the actual values.

***Figure 23: Linear and quadratic term surface versus actual values***



The derived equation representing the complete quadratic response surface is given in Appendix E. Figure 24 indicates the relationship of the predicted values versus the actual values.

***Figure 24: Full quadratic response surface versus actual values***

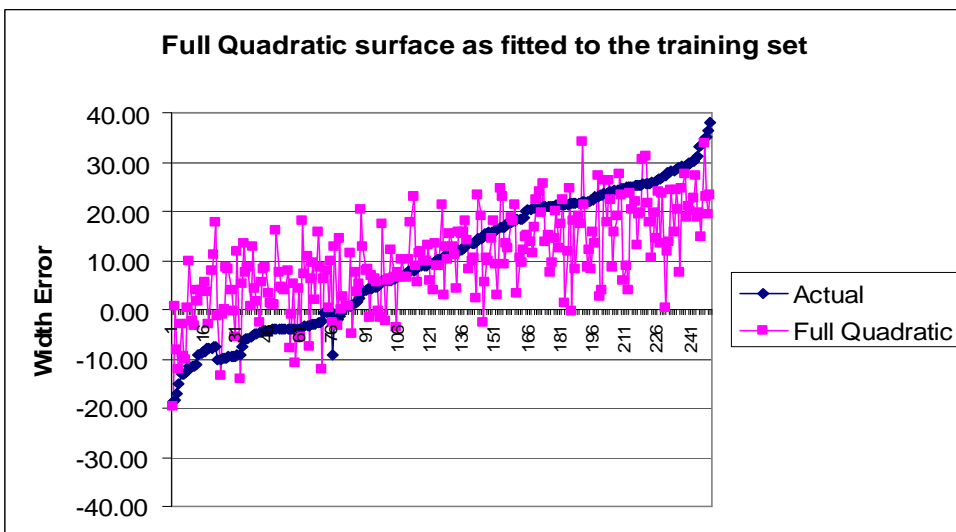


Table 3: Results across the width error range from the different response surfaces.

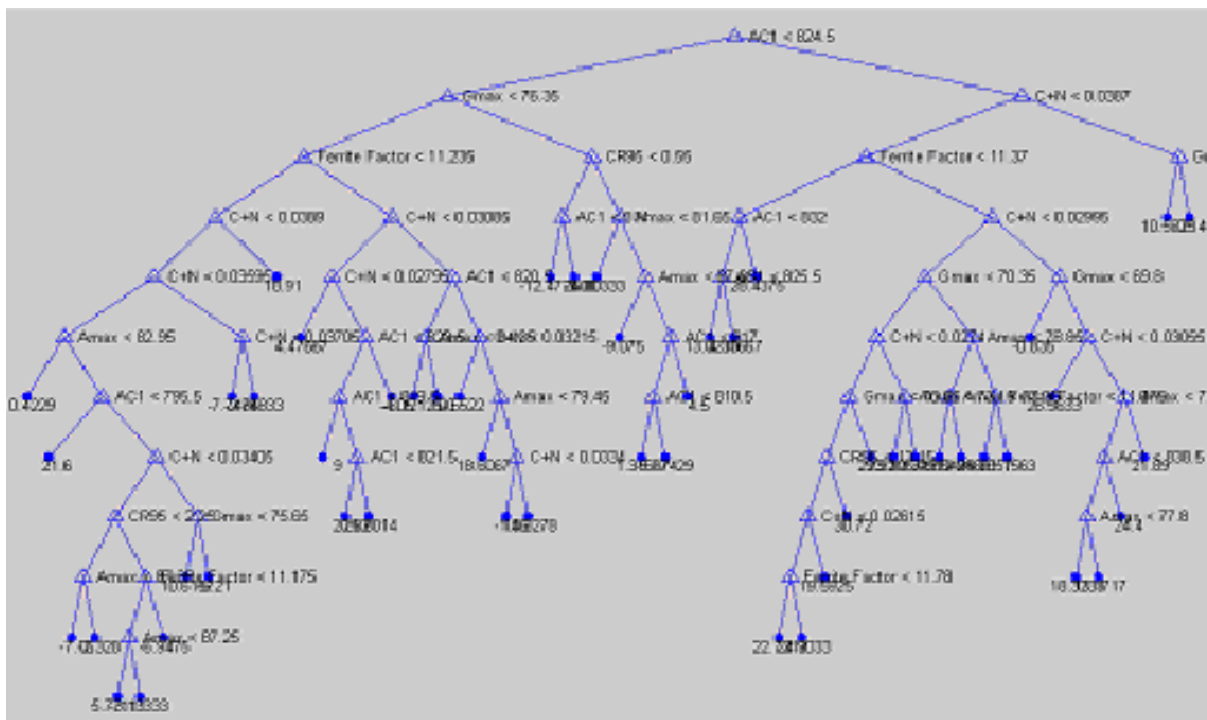
Group	Width error range	# of Records	Linear % Correct	Linear + Interaction % Correct	Linear + Quadratic % Correct	Full Quadratic response surface
1	< 0 mm	79	30%	30%	26%	30%
2	0 - 15mm	65	75%	72%	74%	71%
3	>15mm	105	59%	59%	63%	57%
		<b>249</b>	<b>54%</b>	<b>53%</b>	<b>54%</b>	<b>52%</b>

Table 3 indicates that the response surface models has given results that are very close to each other. The overall accuracy ranges from 54% to 52% correct, which indicates no significant difference between the different models. All the models indicated the best results in group 2. The only model that gave different results from the others in group 1 and group 2 is the linear and quadratic combination model.

### 3.8 Modelling using a Regression Decision Tree

In the preceding section, regression was applied to fit a response surface to the training data set. In this section a regression decision tree will be derived from the training data set. The statistics toolbox found in Matlab<sup>®</sup> was used to derive the regression decision tree. The un-pruned version of the decision tree is displayed in Figure 25 (Quality as obtained from Matlab<sup>®</sup>).

**Figure 25: Un-pruned version of regression tree**



The un-pruned tree consists of 41 levels. Pruning can be done to the basic tree to optimise the structure. With a tree like this one, having many branches, there is a danger that it fits the current data set well, but would not do a good job at predicting new values. Some of its lower branches might be strongly affected by outliers and other artefacts of the current data set. If possible, one would prefer to find a simpler tree that avoids this problem of overfitting. The following procedure can be followed to optimise a tree size by using cross validation.

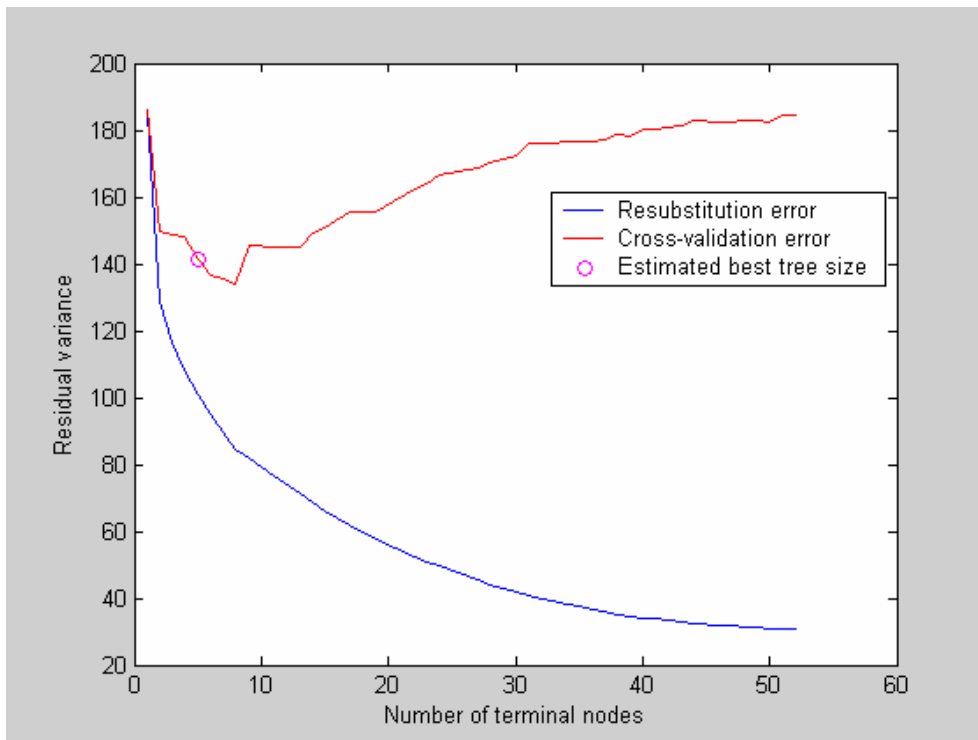
First, compute a resubstitution estimate of the error variance for this tree and a sequence of simpler trees and plot it as the lower (blue) line in Figure 26. This estimate probably underestimates the true error variance. Then compute a cross-validation estimate of the same quantity and plot it as the upper (red) line (Figure 26). The cross-validation procedure also provides an estimate of the pruning level, “best”, needed to achieve the best tree size (Matlab®).

The matlab code that was used is as follows:

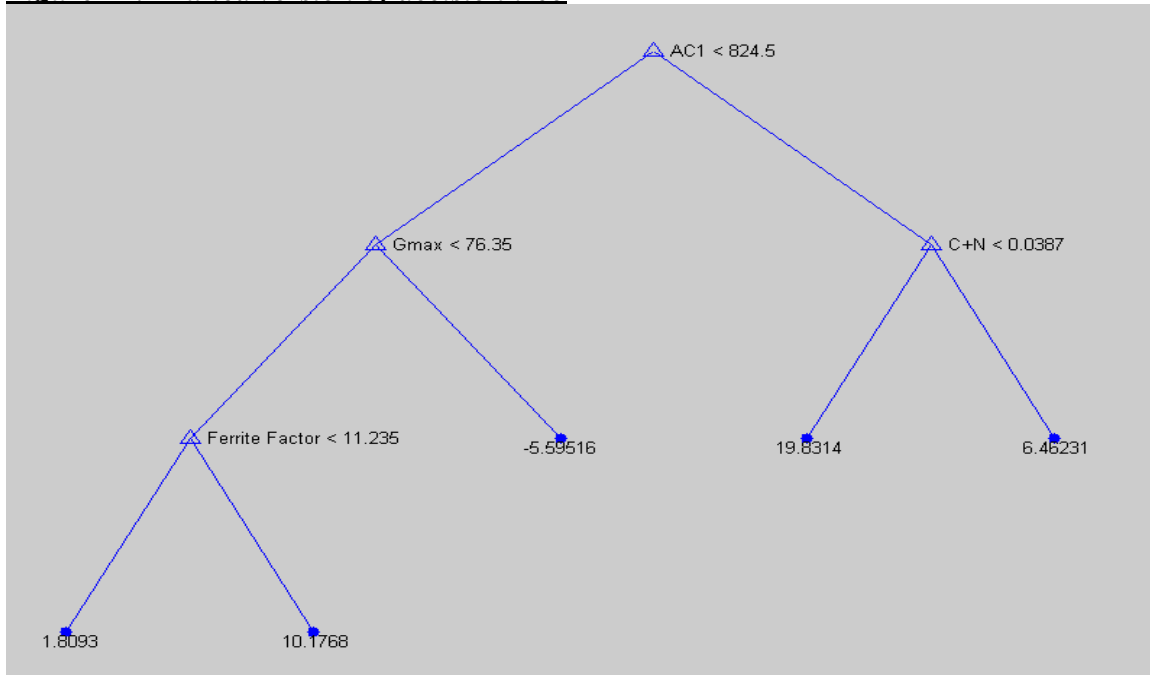
```
[c,s,ntn] = treetest(t,'resub');
[c2,s2,n2,best] = treetest(t,'cross',x,y);
plot(ntn,c,'b-', n2,c2,'r-', n2(best+1),c2(best+1),'mo');
xlabel('Number of terminal nodes');
ylabel('Residual variance');
legend('Resubstitution error','Cross-validation error','Estimated best tree size');
best
t0 = treeprune(t,'level',best);
```

Figure 27 indicates the pruned decision tree as per above Matlab<sup>®</sup> procedure. The pruned model consists out of four levels.

***Figure 26: Resubstitution and Cross-validation error to estimate best tree size***

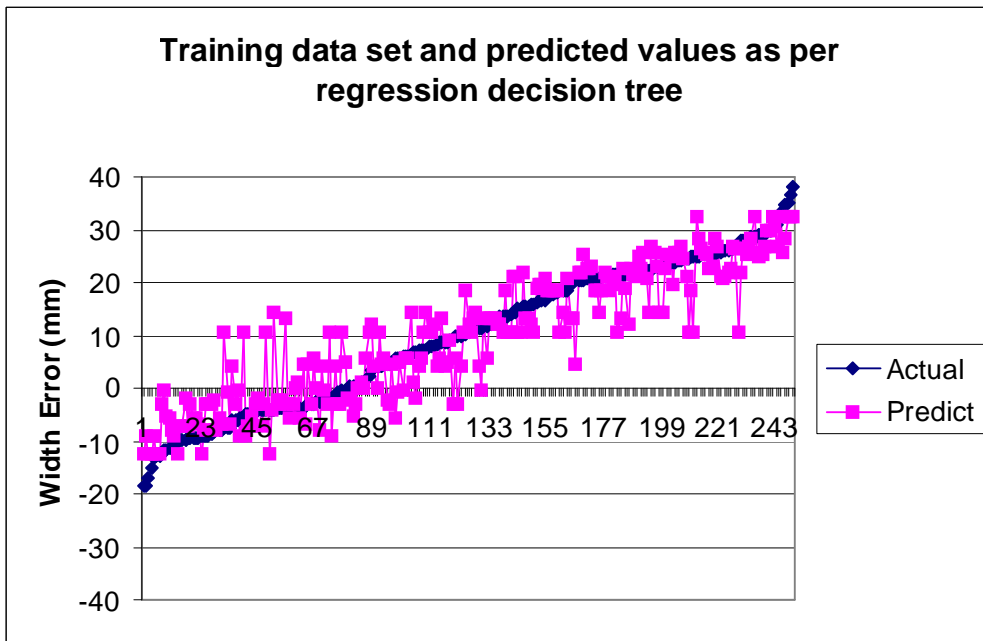


**Figure 27: Pruned version of decision tree**



The best pruning level was determined as level 37 of the original 41 of the un-pruned tree. A comparison of the actual width error and the predicted width error using the pruned decision tree is given in Figure 28:

**Figure 28: Actual width error and predicted width error using pruned decision tree**



The correlation coefficient between the actual width error and the predicted width error is 0.91, which is very good.

*Table 4: Results across the width error range from the decision tree as applied to the training data set.*

<b>Group</b>	<b>Width error range</b>	<b># of Records</b>	<b>% Correct</b>
1	< 0 mm	79	81%
2	0 - 15mm	65	71%
3	>15mm	105	80%
		<b>249</b>	<b>78%</b>

The overall accuracy of 78% is acceptable. The decision tree tends to be more accurate on the extremes of the width error ranges on the narrow and wide side.

### **3.9 Modelling using fuzzy logic**

The problem of modelling the cast strand width error suits the principles of fuzzy logic very well. Two approaches to the fuzzy logic methodology will be taken in this study. The first approach is to use triangular membership functions as the fuzzy logic model basis, while the second methodology is based on polynomial membership functions.

#### **3.9.1 Fuzzy logic model based on triangular membership functions**

The aim is to categorise the input parameters into one of three pre-defined groups. The groups are defined in terms of the width error. Group one is defined as heats with negative width error (i.e. narrow from aim), group two is defined as those having acceptable width error and group three is defined as those heats with large positive expected width errors (i.e. wide from aim). In terms of the actual plant implementation, the groups will be used to select an appropriate secondary cooling pattern. The fuzzy

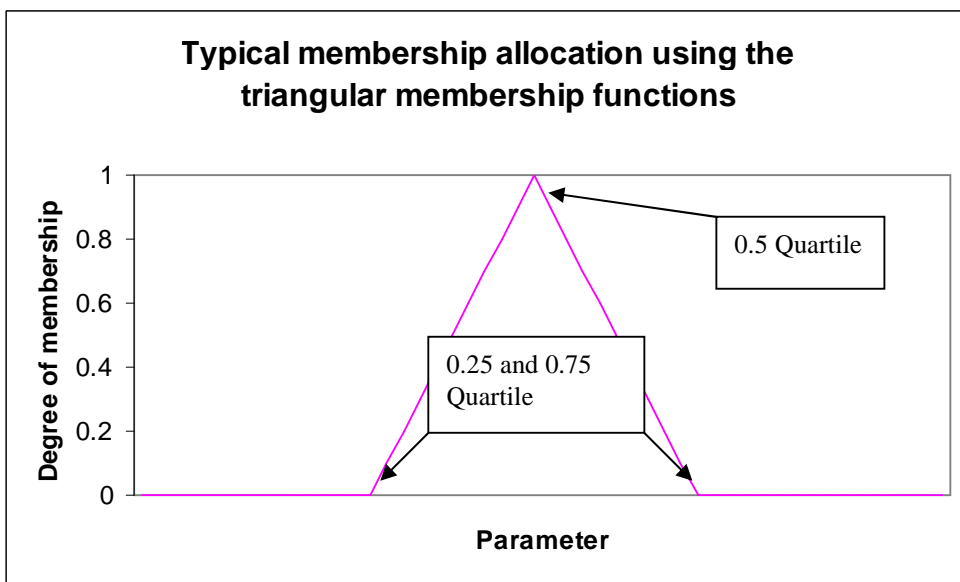


logic membership functions were defined as straight lines using equations of type:

$$y = mx + c \dots\dots\dots(23)$$

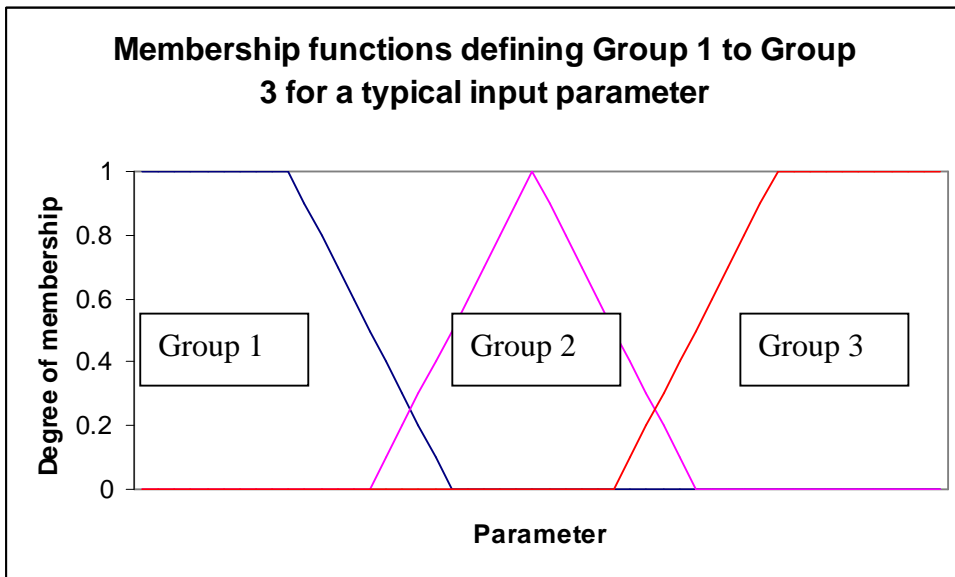
Each parameter was viewed separately and divided into the three groups according to the width error. The 0.25, 0.5 and 0.75 quartiles were determined for each subgroup in each parameter. The 0.5 quartile for each subgroup was chosen to have a membership of one and then the 0.25 and 0.75 quartiles were chosen as having zero membership to the specific subgroup.

***Figure 29: Typical membership allocation using triangular membership functions***



If the membership functions are combined then the result is a distribution with triangular sets. Figure 30 indicates the combined membership functions as applied across a typical parameter range.

***Figure 30: Typical membership functions as defined across a parameter range***



For certain parameters, group one and group three were swapped due to the fact that some parameters ( $A_{max}$ ,  $G_{max}$  and  $C+N$ ) are inversely proportional to the width error and others are proportional to the width error.

The membership functions were determined using the training data set. Each parameter was used separately to determine the membership equations across the parameter range. Each value of each parameter was then fuzzified using the specific membership functions. The de-fuzzification was done using a centre-average approach. Each parameter gave an individual prediction of the width group the heat belongs to between the three predefined width groups. The average of the six predictions were then used to determine the final prediction of the width group. Figure 31 indicates the individual membership functions as determined by the process described above.

***Figure 31: Triangular membership functions of the six input parameters to the fuzzy logic model***

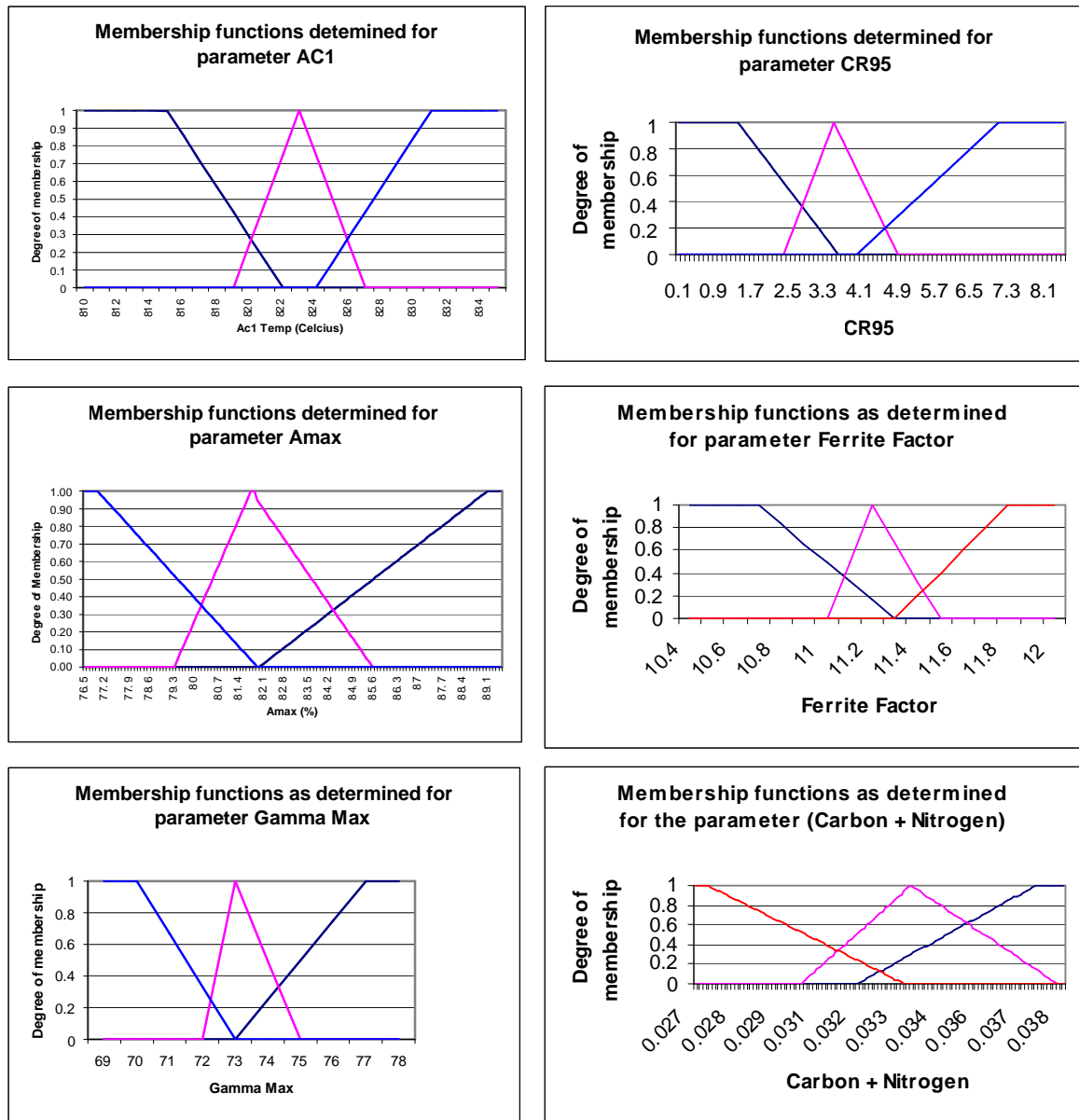


Table 5 indicates the results obtained using the training data set as input to the fuzzy logic model.

Table 5: Results across the width error range from the fuzzy logic model as applied to the training data set.

Group	Width error range	# of Records	% Correct
1	< 0 mm	79	56%
2	0 - 15mm	65	40%
3	>15mm	105	74%
		<b>249</b>	<b>59%</b>

Table 5 indicates that the fuzzy logic model is the most accurate in the width group representing the material that is considered to be very wide from aim. The fuzzy logic model is the least accurate in the width group considered to have acceptable width error values. The implication of such inaccuracy is that material that has acceptable width error will be changed to fall into another width error group causing more deviation in the width error deviations. The overall accuracy of 59% on the training set is not considered to be very good.

### 3.9.2 Fuzzy logic model based on Polynomial membership functions

The basis of this model is to use a polynomial for each parameter to predict a width error separately. The final width error is then taken as the average of the predictions resulting from the six polynomials derived from the six input parameters. The optimum order of polynomial was found for each parameter by deriving six polynomials for each parameter ranging from pure linear to a polynomial of the 6<sup>th</sup> order. The optimum degree of polynomial was chosen as the one with the best correlation to the actual width error. The six input parameters as well as the width error values were standardised to have values between 0 and 1. All polynomials were derived and correlations checked using Matlab<sup>®</sup>. The code for this purpose can be found in Appendix F. Table 6 indicates the correlation

results. The values in bold represent the optimum degree of polynomial for the specific parameter.

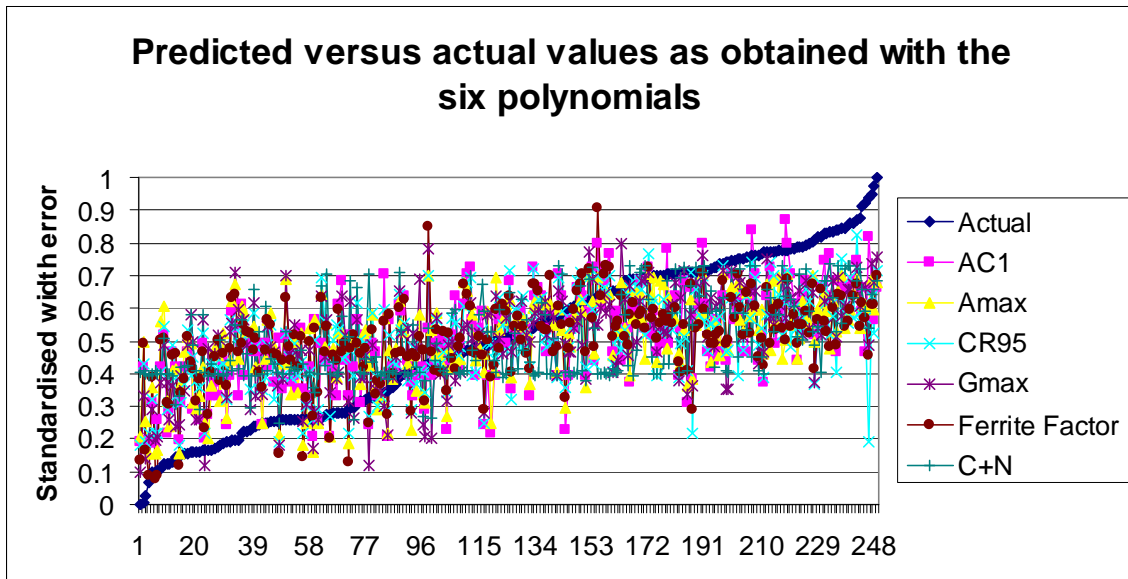
**Table 6: Correlation coefficients for the different degree of polynomials and input parameters**

Parameter	Degree of Polynomial					
	1	2	3	4	5	6
AC1	0.5667	0.5861	<b>0.6179</b>	0.6180	0.6250	0.6250
Amax	0.5080	0.5353	<b>0.5573</b>	0.5575	0.5581	0.5581
CR95	0.5191	0.5556	<b>0.5600</b>	0.5640	0.5696	0.5696
Gmax	0.5732	0.5755	<b>0.6012</b>	0.6016	0.6036	0.6068
Ferrite factor	0.5323	<b>0.5429</b>	0.5493	0.5651	0.5669	0.5721
C + N	0.4541	0.4728	0.4730	0.5016	<b>0.5189</b>	0.5200

The individual optimum correlations are not necessarily taken as the maximum, but are chosen where the change in correlation seems to have stabilised. For the input parameters AC1, Amax and CR95 a third order polynomial is chosen as optimum, for the ferrite factor a second order polynomial is chosen and for the sum of carbon and nitrogen a fifth order polynomial is chosen.

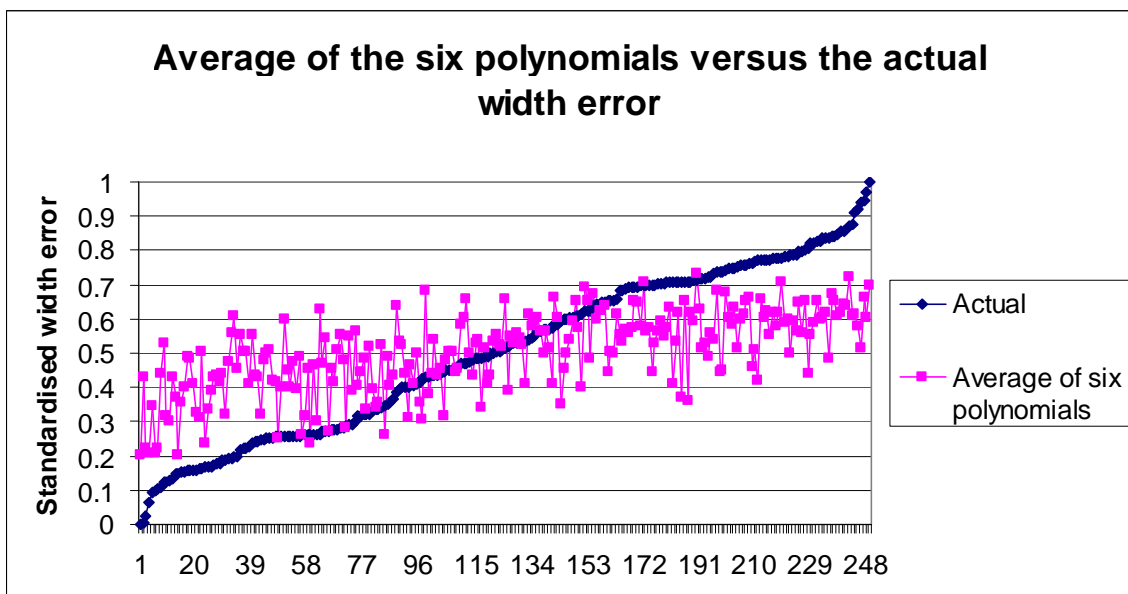
The polynomial for each parameter was derived and is given in Appendix F. Figure 32 gives a comparison between the predicted values from the six polynomials and the actual width error.

**Figure 32: Predicted versus actual values as obtained from the polynomials**



It can be seen from Figure 32 that all the polynomials tend to be less accurate on the extremities. The result of the average of the six predictions is given in Figure 33.

**Figure 33: Average of the Predicted values from the six polynomials versus the actual width error**



It is clear from Figure 33 that the polynomials are not accurate on the extremities of the width error range. The trend of the predictions seems to be correct but the slope does not match the actual value.

*Table 7: Accuracy across the width range as obtained with the average of the polynomials*

<b>Group</b>	<b>Width error range</b>	<b># of Records</b>	<b>% Correct</b>
1	< 0 mm	79	25%
2	0 - 15mm	65	80%
3	>15mm	105	49%
		<b>249</b>	<b>49%</b>

The accuracy obtained (49 %) is not very good. The other negative factor about this method is that it is not accurate on the extremities where a high accuracy is needed in order to identify problematic heats before they are cast.

### **3.10 Summary of the results of the derived models**

Prediction models were developed in the preceding sections to describe the relationship between the six input parameters and the resulting width error. Three approaches were used which included a statistical regression approach using response surface methodology, decision tree models (pruned and un-pruned) and models based on fuzzy logic (two approaches). Table 8 gives a summary of the results obtained with the different models as applied to the training set. Only the best results per modelling approach is given.

*Table 8: Summary table of the results obtained with the different models as applied to the training set..*

Group	Width error range	# of Records	Statistical	Decision tree	Fuzzy logic
			% Correct	% Correct	% Correct
1	< 0 mm	79	30	81	56
2	0 - 15mm	65	75	71	40
3	>15mm	105	59	80	74
		<b>249</b>	<b>54</b>	<b>78</b>	<b>59</b>

It can be seen from Table 8 that the decision tree model resulted in the most accurate model as trained on the training data set.



## Chapter 4

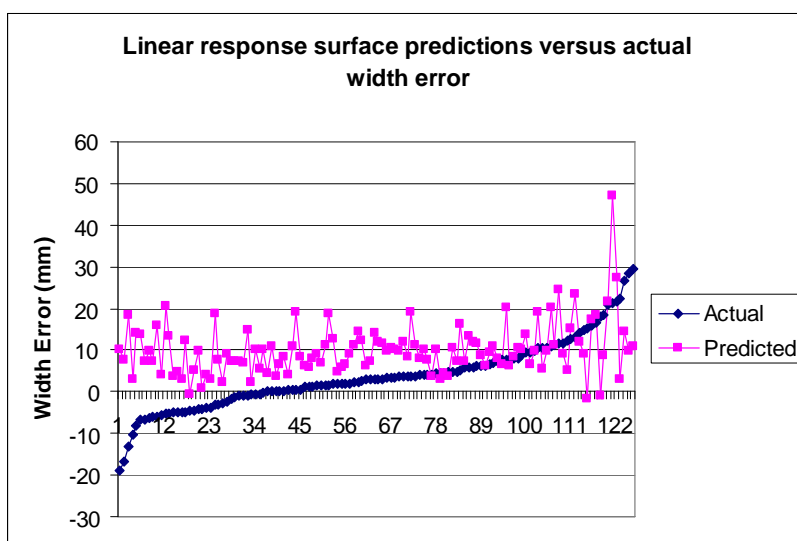
### Evaluation of the derived models on a Validation data set.

In the previous sections three different predictive models were derived using the training data set. The models consist of statistical regression, decision tree model and models based on fuzzy logic. On the training data set the decision tree model was the most accurate, or in other words the best fit on the training data set was obtained with the decision tree model. In this section the derived models are applied to the validation set. The validation set consists out of data that are independent from the training set and therefore represents typical “live” data that the model would encounter during operating conditions in the plant. The validation set consists out of 126 unique heat chemistries and their associated strand width error.

#### 4.1 Application of the derived statistical model on the validation data set

The statistical regression model (equation 22) was applied on the validation data set using Microsoft Excel. The results obtained are given in Figure 34.

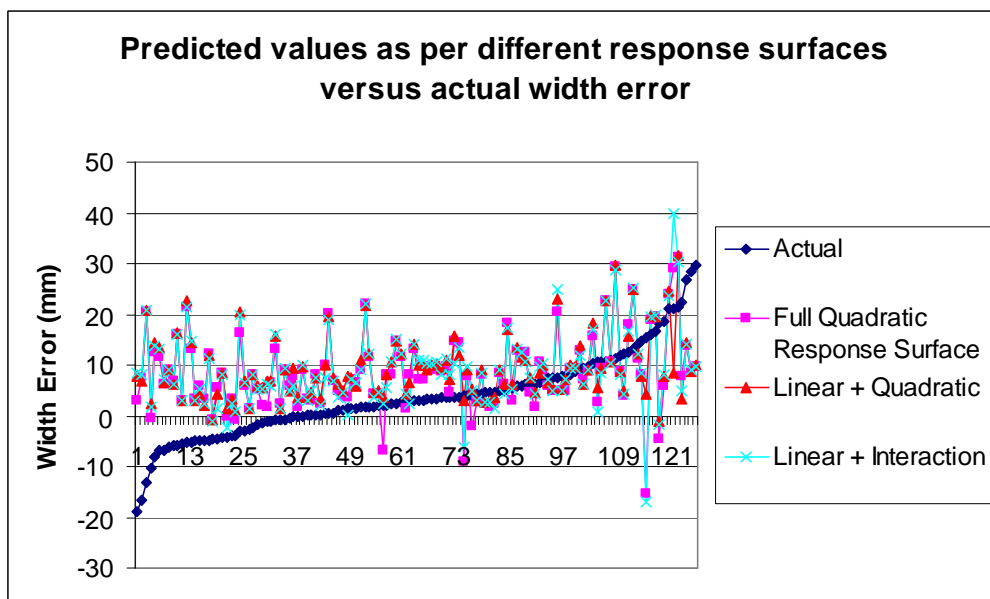
***Figure 34: Linear Response surface model applied to the validation data set***



Visually it can be seen that the fit is not very good. The calculated correlation coefficient is only 0.21. If the three width error groups are considered then Table 9 indicates that the model is most accurate in the width error group 0 – 15mm. The two groups on the extremes of the populations have very low accurate predictions. It is also clear from Figure 34 that the predicted values are predominantly found between 0 and 20mm. This also explains the very good accuracy obtained in the 0 – 15mm band. Basically all the predictions are in this interval and therefore the chance of the prediction being accurate is very good. The high accuracy rate is however not believed to be due to the accuracy of the model but rather due to the fact that all predictions are in this interval anyway.

The more advanced response surface models were also applied to the validation data set. Figure 35 represents the predictions obtained with the different response surfaces versus the actual width error.

**Figure 35: Response surface models applied to the validation data set**



It can be seen from Figure 35 that the response surfaces do not pick up the trend of the actual data. The response surfaces are not accurate in the extremities of the width error range. This is unacceptable for a potential on-line prediction model where it is critical that the model be accurate on the extremities of the width error range. Table 9 gives a summary of the accuracy across the width range, as achieved with the more advanced response surface models.

*Table 9: Summary of the accuracy obtained with the surface response models as applied to the validation data set.*

Group	Width error range	# of Records	Linear	Full Quadratic	Linear + Quadratic	Linear + Interaction
			% Correct	% Correct	% Correct	% Correct
1	< 0 mm	38	3%	8%	3%	8%
2	0 - 15mm	76	87%	88%	92%	91%
3	>15mm	12	42%	0%	0%	0%
		<b>126</b>	<b>57%</b>	<b>55%</b>	<b>56%</b>	<b>57%</b>

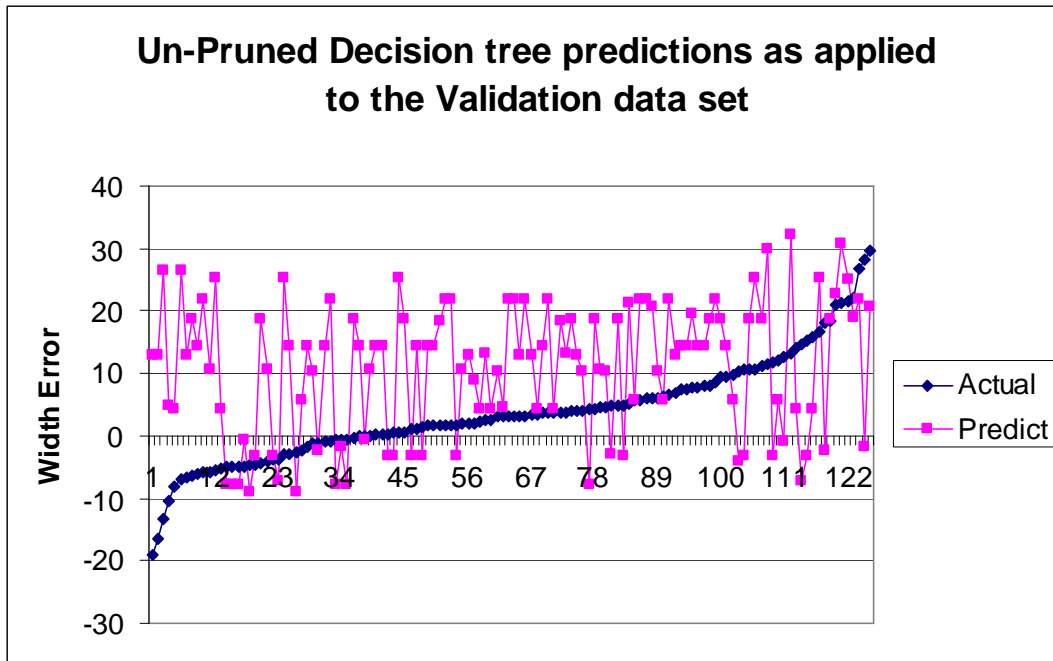
From Table 9 it can be seen that all the models are accurate in the width error range represented by group 2, but all the models perform very poorly in the width error ranges represented by group 1 and group 3. From Figure 38 it is clear that the predictions tend to be in the range of group 2, and hence, a high accuracy is obtained in group 2. The accuracy is therefore not thought to be due to the models being accurate in that range, but rather due to the fact that most of the predictions fall in that range.

#### **4.2 Application of the derived decision tree model on the validation data set**

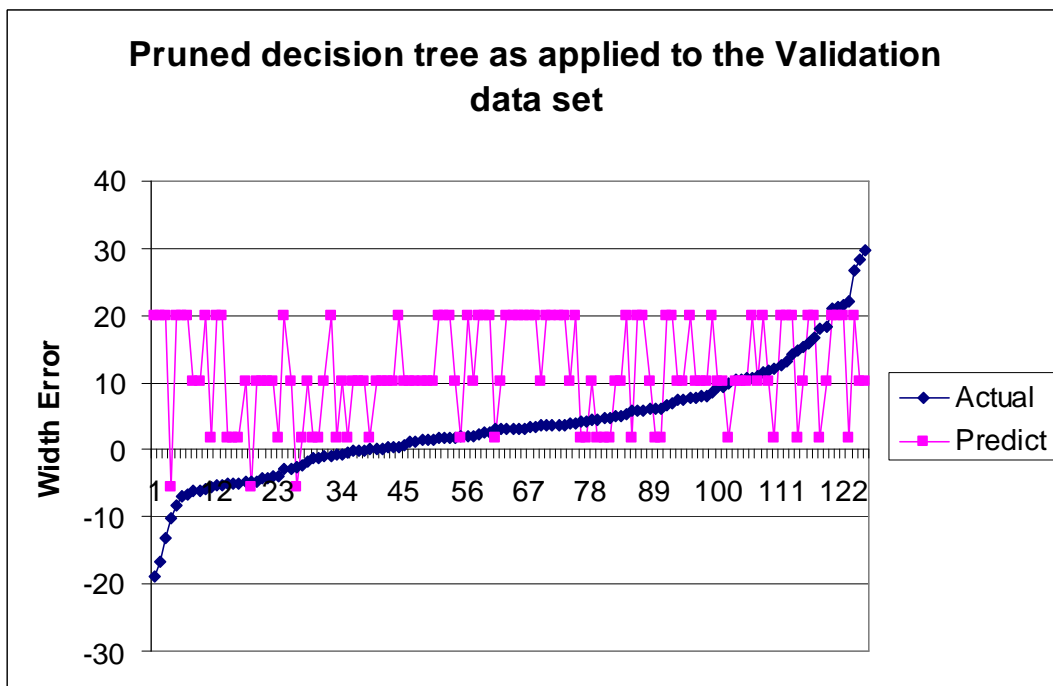
The decision tree model derived from the training data set, was applied to the validation set in two formats: Pruned and un-pruned to determine the best model between the two

versions. Figure 36 and 37 depict the results obtained with the two decision tree models as applied to the validation set.

***Figure 36: Decision tree model results as applied to the validation data set.***



***Figure 37: Pruned decision tree model results as applied to the validation data set.***



The accuracy in terms of the three groups is given below in Tables 10 and 11.

*Table 10: Accuracy obtained per width error group with the decision tree model.*

Group	Width error range	# of Records	% Correct
1	< 0 mm	38	36%
2	0 - 15mm	76	47%
3	>15mm	12	66%
		<b>126</b>	<b>46%</b>

*Table 11: Accuracy obtained per width error group with the pruned decision tree model.*

Group	Width error range	# of Records	% Correct
1	< 0 mm	38	7%
2	0 - 15mm	76	59%
3	>15mm	12	50%
		<b>126</b>	<b>42%</b>

From Tables 10 and 11 it can be seen that both decision trees result in an overall accuracy rate of less than 50 % which is not very good. The un-pruned decision tree has more balanced results between the three width groups. The most accurate results were obtained in group 3 with the large decision tree and the pruned decision tree resulted in the most accurate results in group 2. The good results obtained in group 2, are a result of the fact that most of the predictions of the pruned decision tree fall within the range of group 2 and the accuracy of the results are not a reflection of the true accuracy in group 2.

### **4.3 Application of the derived fuzzy logic models on the validation data set**

The fuzzy logic model that is based on triangular membership functions was applied to the validation data set. The accuracy obtained per width error group is given in Table 12.

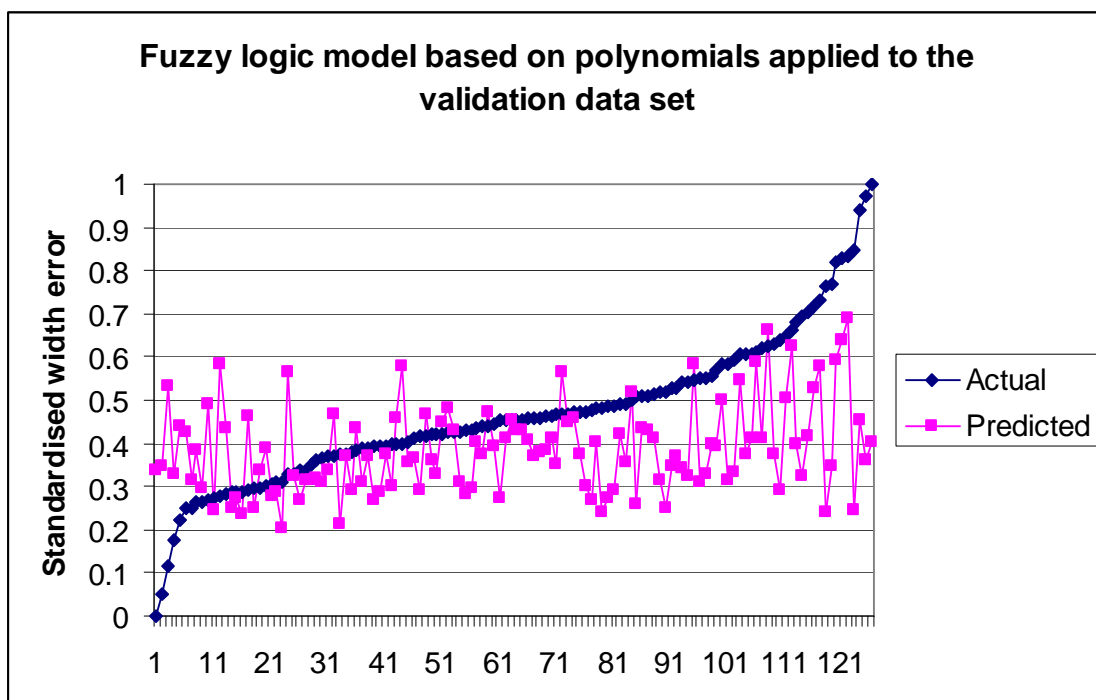
Table 12: Accuracy obtained per width error group with the fuzzy logic model based on triangular membership functions .

Group	Width error range	# of Records	% Correct
1	< 0 mm	38	32%
2	0 - 15mm	76	57%
3	>15mm	12	50%
		<b>126</b>	<b>48%</b>

The overall accuracy of 48% is not very good. The best results were obtained in group 2 with an accuracy rate of 57%.

The second model based on fuzzy logic was derived using polynomials as prediction functions for the individual parameters. Figure 38 indicates the relationship between the predicted values and the actual width error values.

Figure 38: Predicted values from the fuzzy logic based model using polynomials versus standardised actual width error values



The predictions from the fuzzy logic model fail to pick up the trend from the actual data. It is thus clear, that the model seems to predict between 0.2 and 0.6 across the width error range. This “flat” prediction might be due to the average taken from the six predictions. The accuracy across the width range is given Table 13.

Table 13: Accuracy across the width error range as obtained with the fuzzy logic model based on polynomials and applied to the validation data set.

<b>Group</b>	<b>Width error range</b>	<b># of Records</b>	<b>% Correct</b>
1	< 0 mm	38	52%
2	0 - 15mm	76	67%
3	>15mm	12	16%
		<b>126</b>	<b>58%</b>

The overall accuracy of 58% on the validation data set is reasonable, considering the results obtained with the other models. One positive about this model, is that it seems to be fairly accurate in group 1 and group 2 but unfortunately not in group 3.

#### **4.4 Application of the model currently in use by the plant to the validation data set**

The model currently used in the plant uses a slightly different philosophy to the models developed previously. The model is purely based on a set of rules derived empirically from a similar training set, used for the derivation of the models in the previous sections<sup>[7]</sup>. The model philosophy is to only target the heats that fall in the width error groups 1 and 3. If a heat is predicted to be within group 1, then the secondary cooling is adjusted to provide more creep during the secondary cooling stage in the continuous casting machine. This results in the strand being slightly wider and should fall typically into width group 2. The same principle applies to heats that are predicted to fall into

group 3. The secondary cooling intensity is increased and that results in the strand ending up slightly narrower, which should be sufficient seeing that the heat then falls within group 2. This philosophy results in a narrower width distribution of the material<sup>[7]</sup>. The model structure is such that the chemistry is tested against the set of rules, should one of the rules become active, then the secondary cooling would be adjusted. Should none of the rules become active no action will be taken by the model and it would be termed a “no decision”. The model therefore, does not have the capability to predict material to fall within group 2. However, to measure the accuracy obtained in group 2, the correct decision by the model would be not to change the cooling pattern. The accuracy of the model therefore in group 2 will be measured in terms of the percentage “no decisions”. Table 14 gives the results of the current model as applied to the validation set.

*Table 14: Results of the current model as applied to the validation set.*

<b>Group</b>	<b>Width error range</b>	<b># of Records</b>	<b>% Correct</b>	<b>% No Decision</b>
1	< 0 mm	38	88%	58%
2	0-15mm	76	67%	N/A
3	>15mm	12	71%	42%
		<b>126</b>	<b>73%</b>	<b>54%</b>

The percentage of “no decision” is relatively high at 54 %, and this material should be viewed as lost opportunities where a change could have been implemented, to improve the width error. The 54% also indicates that the model can be improved to be able to recognise more problematic material. Table 14 also indicates that where decisions were made in the model, it was 73% correct. This result is considered to be excellent if it is compared to the other models derived and tested in the previous sections.



#### **4.5 Summary of results obtained by applying the models to the validation data.**

In the preceding sections the three derived models and the current model that is used in the plant were applied to the same validation data set. This gives a good indication of the relative accuracy obtained with the different data mining techniques on this specific problem. The aim is to select the most appropriate model for this specific application of predicting the width error of a heat before it is cast by purely taking the chemistry into account. The results obtained with the statistical regression, decision tree and fuzzy logic models were very disappointing. The only model with a success rate of more than 50% accurate, is the statistical regression model with a 57% success rate. This relatively high success rate is mainly due to the fact that most of the predictions were in the range of 0mm to 15mm and therefore many predictions that were correct, were purely by chance, because the largest width error population is group two, representing the 0 mm to 15mm range. The current model is a rule-based model drawn up empirically from historical plant data. The expectation was that the more advanced data mining techniques would outperform the current rule-based model. This assumption was tested with the validation data set and the current rule-based model surprisingly outperformed the more advanced techniques comfortably. The current model achieved an accuracy of 73% which is excellent compared to the results of the other data mining techniques. The model that is currently used in the plant is therefore the best suited for this specific application. Table 15 gives a comparison of the four models under consideration in terms of their accuracy on the validation data set. Only the best results obtained per model type is given.

*Table 15: Combined results of the different models as achieved on the validation data set.*

<b>Model</b>	<b>Total Accuracy</b>
<b>Statistical regression</b>	57%
<b>Decision Tree</b>	46%
<b>Fuzzy logic</b>	58%
<b>Current model</b>	73%

From Table 15 it is clear that the current model is outperforming the other models by far. The combined result of 73% accurate is very good, considering the complexity of trying to predict the strand width error for a specific heat from the chemistry of the specific heat. Due to the fact that the model only made decisions on 48% of the heats that fell into group one and group three, it does not mean that >70% of all the problem heats can be picked up by the current model. Currently, the results only indicate that at least a large portion of the problem material can be picked up accurately and adjustments can be made to rectify the situation. What the results mean for the plant is a narrower width distribution which means better strand width control for 12% Chrome, non stabilised ferritic stainless steel.

## **Chapter 5: Plant results of the model best suited for this application**

### **5.1 Introduction**

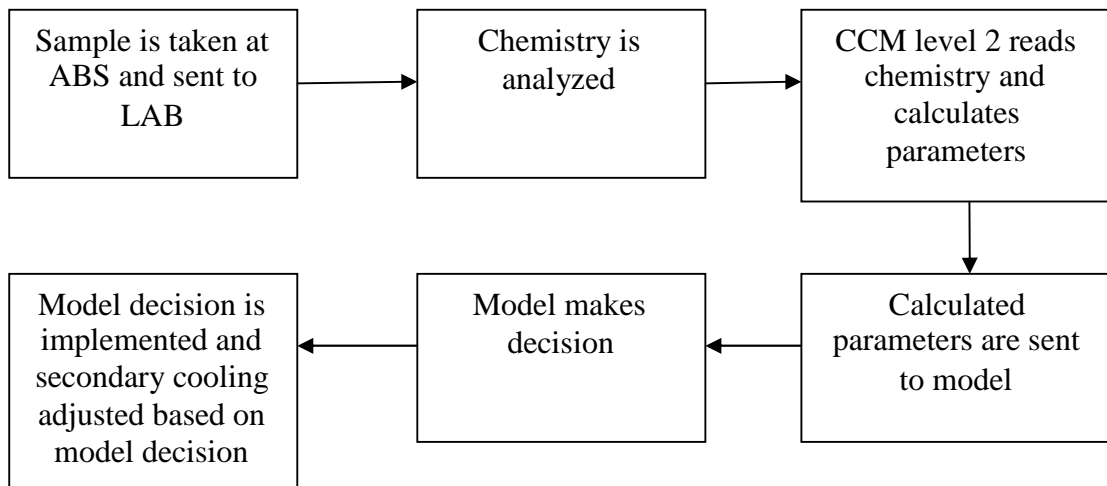
In the preceding chapters three different models were derived, having used the same training data set. The different models represented different data mining techniques and included statistical regression, decision trees and fuzzy logic. The derived models were all applied to the same validation data set and compared to the performance of the model currently in use also tested on the validation set. The results were then compared and it was found that the model that is currently used in the plant was far more superior in terms of prediction accuracy than the models from the other data mining techniques. This section details the implementation of the model in the plant, together with the results achieved with the model while using it as prediction tool in the production environment.

### **5.2 Implementation of the model in the production environment**

The model was implemented in the production environment in November 2004. Development and testing of the model started in January 2004. Initially the model was run “off-line” and then gradually it was used manually “off-line” for predictions and the secondary cooling would be adjusted based on the model predictions. In November 2004 the model was implemented as a fully automatic model running on the Level 2 computer system of the continuous casting process. The sequence of events as outlined in Figure 39 can be summarised in the following way: A chemistry sample is taken at the end of the process at the rinsing station (ABS). There is approximately a 10 to 15 minute gap

between the end of the rinsing station process and the start of the casting operation. After the chemistry sample has been taken, it is sent to the laboratory for analysis. After the analysis is done, the chemistry results are read by the level two computer system of the continuous caster. The width model is situated in the caster level two computer system. Once the chemical analysis has been received by level two, the AC1, Amax, CR95, Gmax and ferrite factor is calculated and sent to the model as input parameters. The input parameters are checked by the model by fitting each rule in sequential order to the input parameters. The individual rules are checked until a rule becomes active. When all the rules have been checked and none fits, the default/standard secondary cooling practice is used. If one of the rules becomes active, the secondary cooling practice would be changed to either more aggressive or less aggressive based on the result of the specific active rule.

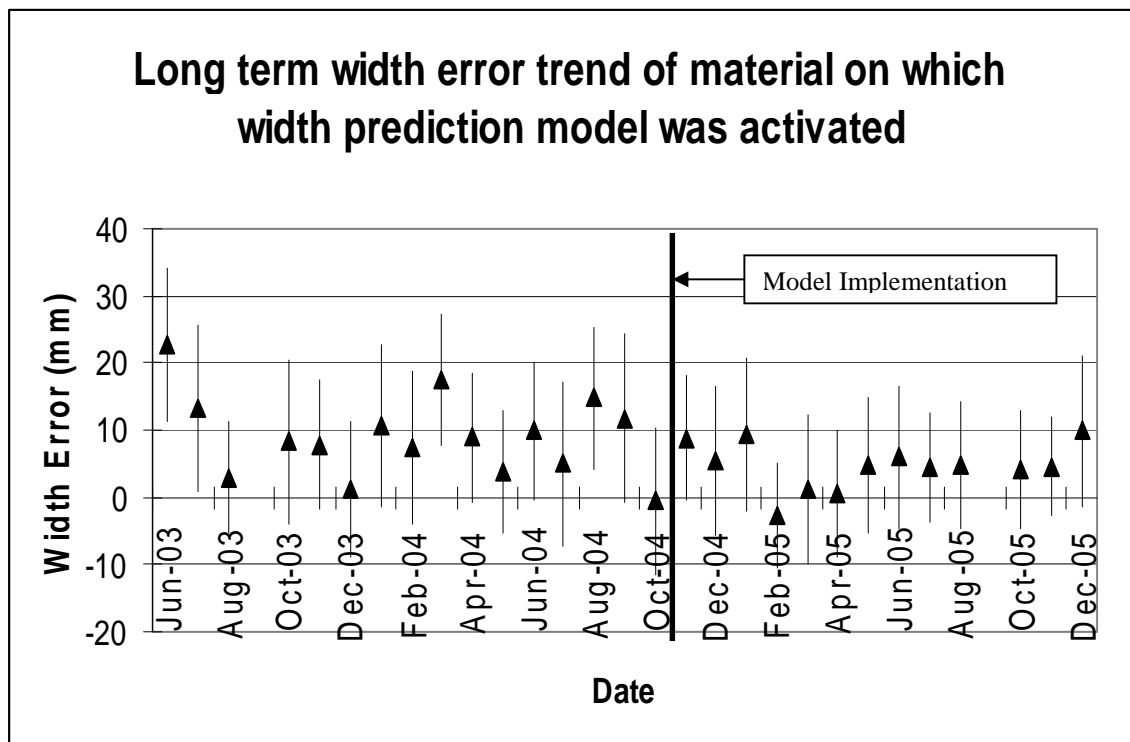
**Figure 39: Model decision making process as implemented in the plant**



### 5.3 Plant results from the implemented model

This discussion will begin by indicating the long term width error trend of the material on which the model is implemented and is depicted in Figure 40. The results are based on the cast width as measured by the lasers at the exit of the CCM. All start-up slabs and end-of-cast slabs were excluded from the results, due to the fact that they are very narrow from aim, due to processing at very slow casting speed. Outliers were filtered using the inter-quartile range approach.

**Figure 40: Long-term width error trend of 12% Chrome non-stabilised ferritic Stainless steel.**



The graph represents the average and a two sigma (standard deviation) range. The two sigma range represents approximately 75% of the population. It can be seen that the width error stabilised after the implementation of the model. The width error also seems to be lower on average. A more stable width error on average per month means that the

control on the strand width improved because the width error for each month is in a similar region. This was not the case before the model was implemented. Before the model was implemented it is clear that each month had a different average width error and indicates poor width control.

The model results were further evaluated by comparing them to a population where the model was not in operation. The reference population chosen were all material with internal steel type 41220 and 41311 that were cast between July 2003 and November 2004. This population consisted out of 1637 individual slabs representing 420 different heats, that translate into 420 unique chemistries with associated width error. The population representing the population where the model was active, consisted out of 1407 individual slabs representing 383 heats cast between December 2004 and December 2005. This population was split into two populations. This was necessary because the secondary cooling practice used for the heats with large positive width errors was changed in May 2005 to a less aggressive pattern, after it was suspected that the high cooling intensity resulted in some slabs cracking. The relationship between the cracked slabs and the secondary cooling practice, could however, never be confirmed. The population representing the period from December 2004 to May 2005 consisted out of 242 heats and the population representing June 2005 to December 2005 consisted out of 141 heats. Each heat represents approximately 110 tons and therefore the reference population (without model) represents 46 200 tons of cast steel and the population where the model was active represents 42 130 tons of cast steel in total. Table 16 gives a summary of the two populations.

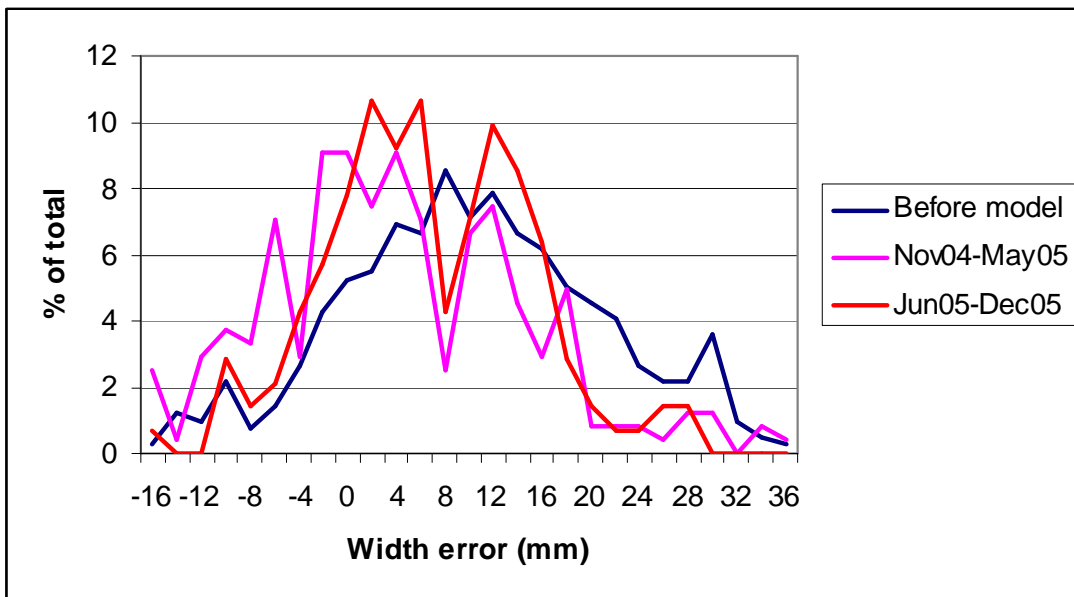
Table 16: Summary of reference population and model active population

	<b>Reference Population (Before Model)</b>	<b>Model Active Population (Nov 04 – May-05)</b>	<b>Model Active Population (June 05 – Dec 05)</b>
<b># of Heats</b>	420	242	141
<b># of Slabs</b>	1637	866	541
<b>Tons</b>	46 200	26 620	15 510
<b>Mean*</b>	9.27 mm	3.29 mm	5.68 mm
<b>Standard Deviation*</b>	10.46 mm	10.42 mm	8.33 mm
<b>Range* (max-min)</b>	54	55	44

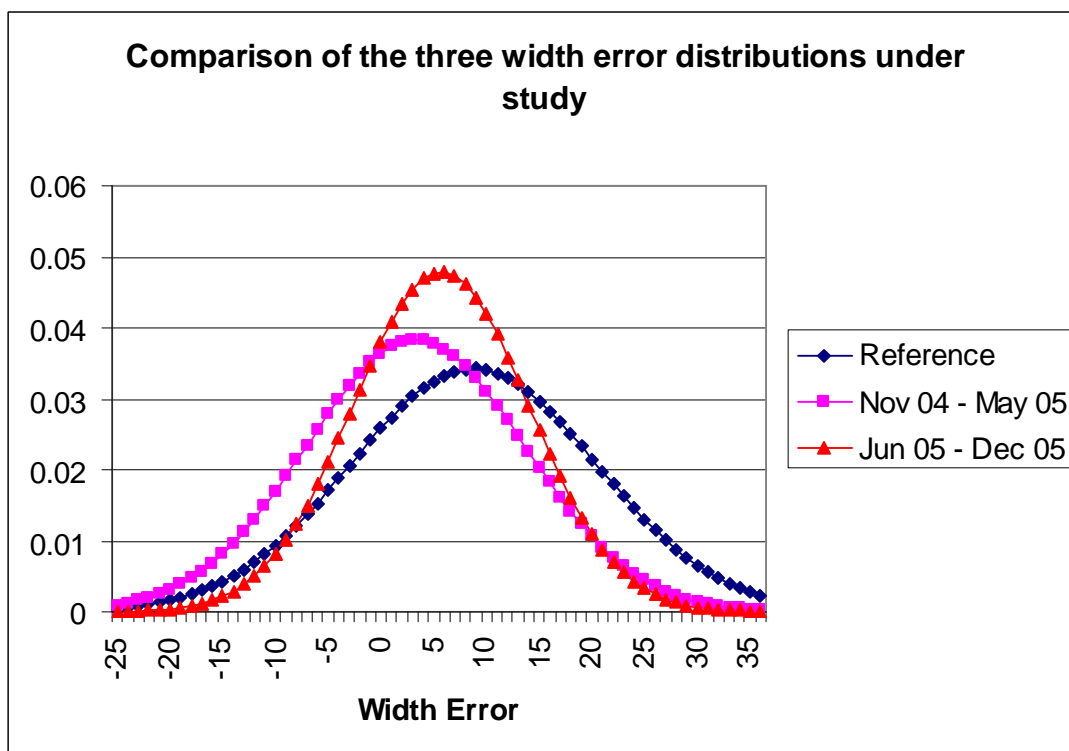
\* Based on heats

The three means were tested against each other using the student t-test with a 95% confidence interval and all three means were found to be significantly different from each other. It is evident that the mean of the reference population is higher than the two means of the populations where the model was active. The reason for this is that in September 2004 by management decision, the preferred range for the caster to supply ferritic slabs to the Hot mill was changed to 7.5 mm narrower. This resulted in a general width error population shift with a larger percentage of the material falling in group one of the width error definition (<0mm width error). Figure 41 is a frequency plot from the raw data to give a clearer picture of the raw data. The frequency is expressed as a percentage of the total of each group. Figure 42 indicates the distributions of the three populations if they are assumed to be normally distributed and have the means and standard deviations listed in Table 16. The standard deviation difference between the reference population and the June 2005 to December 2005 population is 20.4 % and the range differs by 18.5 %. The results indicate that the change made to the secondary cooling in May 2005 was very beneficial for the width error.

***Figure 41: Frequency diagrams comparison of the three populations under study***



***Figure 42: Normal distribution comparison of the three populations under study***



As is expected from the values in Table 16 the reference population and the population between November 2004 and May 2005 have similar normal distributions. The population representing June 2005 to December 2005 has a narrower distribution than the other two populations due to its smaller standard deviation. Figure 42 clearly indicates



the narrower width distribution of the population representing June 2005 to December 2005.

The two populations where the model was active were analysed in more detail. It is important to know how much of the material was cast with modified secondary cooling patterns and how the populations compare between those that were cast with modified cooling patterns and those that were cast with the standard secondary cooling pattern (Table 17).

*Table 17: Summary of the populations with and without secondary cooling modifications*

	<b>Nov 04 – May 05 Standard water pattern</b>	<b>Nov 04 – May 05 Modified water pattern</b>	<b>Jun 05 – Dec 05 Standard water pattern</b>	<b>Jun 05 – Dec 05 Modified water pattern</b>
<b># of Slabs</b>	352 of 818 total slabs	466 of 818 total slabs	157 of 529 total slabs	372 of 529 total slabs
<b>“Wide” cooling Practice</b>	52% of 818 slabs		68% of 529 slabs	
<b>“Narrow” cooling Practice</b>	4.5 % of 818 slabs		1.89 % of 529 slabs	
<b>Mean*</b>	7.21 mm	0.24 mm	5.41 mm	6.46 mm
<b>Standard Deviation*</b>	11.42 mm	9.31 mm	9.1 mm	9.3 mm
<b>Range*</b>	56 mm	57 mm	49 mm	54 mm

The means of the two populations of the material between November 2004 and May 2005 are significantly different (Student t-test 95% confidence interval). This can be attributed to the fact that the majority of the water patterns that were modified were changed to the “wide” pattern. The compensation achieved with the modified secondary cooling, was too much and the mean of 0.24 with a standard deviation of 9.3 means that

some material was cast that was too narrow and could have caused processing problems downstream. This problem was exaggerated by the fact that in September 2004 the aim cast widths on all ferritic material was changed to 7.5 mm narrower. This caused problems with the model compensation because the data used for training the model, and to derive the secondary cooling pattern were based on a population with an average width error of 7.5mm wider. The population of June 2005 to December 2005 is relatively smaller than the population of the first part of 2005 because of lower production volumes due to market conditions. The means of the two populations of the material from June 2005 to December 2005 are not significantly different (Student t-test, 95% confidence interval). This result indicates that the model changed material that would have been problematic to be within the range achieved with the standard cooling practice resulting in an improved (reduced) width error distribution during the time period June 2005 to December 2005 (Table 17, Figure 42). On the population from November 2004 to May 2005 the model changed 56% of the material cast to more suitable secondary cooling practices. On the population from June 2005 to December 2005 the model changed 70% of the material cast to more suitable secondary cooling patterns. The results are also contradictory to the 48% achieved on the validation data set in the previous chapter.

#### **5.4 Summary of plant results**

In summary it is noted that the width prediction model was successfully implemented in the plant to run “live” as a production tool. Only one major change was done to the model after implementation and that was to change the secondary cooling practice used on the material with predicted large positive width errors (too wide). The compensation done with the particular secondary cooling was too much and the effect was also

compounded by the fact that all the ferritic material had an aim cast width change of 7.5 mm narrower. The secondary cooling practice used on material with predicted large positive width errors were changed to a less aggressive practice. The results of the implemented model are detailed in the preceding sections and it can be seen that the model did in fact improve the width error of 12% Chrome ferritic stainless steel (non – stabilised) cast at Columbus Stainless. The narrower width error distribution implies that the width error improved.

## Chapter 6

### Conclusions and Recommendations

#### 6.1 Conclusions

1. A previous study<sup>[27]</sup> indicated that there is a relationship between the heat composition and cast width change of 12% chrome non-stabilised ferritic stainless steels. This study used parameters that describe the dual phase characteristic of the solidified shell to predict the expected width change of a heat based on the composition before the heat is cast.
2. The relationship of the chemistry and the resulting width error is due to the fact that during the secondary cooling temperature range (1200°C to 800°C) the 12% chrome, non stabilised ferritic stainless steels go through a mixed phase area between ferrite and austenite. Each heat that has a unique chemistry (still within internal specifications on the different elements) will result in a different austenite to ferrite ratio in the secondary cooling region. The hot strength of austenite is more than the hot strength of ferrite, the resistance to creep of austenite is more than ferrite and the volume of BCC ferrite is more than FCC austenite. These differences result in slabs being cast narrow from aim if the phase ratio favours austenite and vice versa if the ratio favours ferrite.
3. The rule based model that existed in the plant could not be improved by using more advanced data mining techniques. The project suggests that a good engineering understanding of the width change governing phenomena coupled with a simple data mining technique performs better than the more advanced data mining techniques.

## 6.2 Recommendations for future work

1. The model that is currently running “live” in the plant should be checked regularly and be updated with the latest information. The original idea of the model was that the rules would be updated regularly to enable the “intelligence” of the model to improve as a function of time.
2. From the results (table 17) it is clear that the model changes the secondary cooling on more than 60 % of the material that is cast. A vast majority of the changes made are the change to a more aggressive secondary cooling. It should be considered to make the more aggressive secondary cooling pattern the standard cooling pattern and only selectively use the current standard secondary cooling pattern. This will reduce the amount of secondary cooling changes through casts that should also promote more stable casting conditions.
3. The implemented model was derived from historical data including heat composition and width error. Should a change to the aim composition be made to material where the width model is active, then the model results should be very closely monitored to ensure the rules are still valid.
4. Recent developments at the continuous caster, include the continuous monitoring of the width error by production personnel. Should certain deviations be detected, the mould will be adjusted during the cast to correct the width error. The production personnel should receive training on the width model to understand what decision the width model had made and what effect it had on the width error. If the production personnel are not aware of the model, the situation can quickly arise where both the width model and

production personnel are changing the cast width. The results can be disastrous.

## References

1. **Ackerman, M.J.** *Steel Slab Surface Quality Prediction Using Neural Networks*. MSc, Potchefstroom University for CHE. Potchefstroom. South Africa.2003.
2. **Anon.** *Data Mining Techniques*. [Web:]  
<http://www.statsoft.com/textbook/stdatdmin.html>. [Date of access: 29/07/2005].
3. **Anon2.** *KDnuggets:Polls:Data mining techniques(Nov 2003)*. [Web:]  
[http://www.kdnuggets.com/polls/2003/data\\_mining\\_techniques.htm](http://www.kdnuggets.com/polls/2003/data_mining_techniques.htm). [Date of access:29/07/2005].
4. **Anon3.** *Fuzzy set operations – What do ya mean Fuzzy??!!*. [Web:].  
[http://www.doc.ic.ac.uk/~nd/surprise\\_96/journal/vol4/sbaa/report.fuzzysets.html](http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/sbaa/report.fuzzysets.html). [Date of access:03/03/2006].
5. **Bhupinder S. Dayal, Taylor P.A, Macgregor J.F.** *The design of experiments, Training and Implementation of Nonlinear Controllers Based on Neural Networks*. The Canadian journal of chemical engineering, volume 72. December, 1994.
6. **Brule', J.F. 1985.** *Fuzzy systems – A Tutorial*. [Web:].  
<http://www.austinlinks.com/Fuzzy/tutorial.html>. [Date of access:03/03/2006].
7. **De Beer, G.** *Continuous cast strand width control: Predictive width control on a 12% chrome corrosion resistant (Type EN 1.4003) steel at the slab caster at Columbus Stainless.*(In Proceedings of the 5<sup>th</sup> ECCC: Paper presented at Conference held in Nice, France, 20-22 June 2005).
8. **Jenkins, M.S.** *Heat transfer in the continuous casting mould*. PhD Thesis, Monash Univ. Clayton, Vic.,Australia.1999.
9. **Kaehler, S.D. 1998** *Fuzzy Logic-An introduction*. Encoder. Newsletter of the Seattle Robotics Society. [Web:]. [http://www.seattlerobotics.org/encoder/mar98/fuz/fl\\_part1](http://www.seattlerobotics.org/encoder/mar98/fuz/fl_part1) to [http://www.seattlerobotics.org/encoder/mar98/fuz/fl\\_part6](http://www.seattlerobotics.org/encoder/mar98/fuz/fl_part6). [Date of access:01/12/2006].
10. **Kweku-Muata, Osei-Bryson.** *Evaluation of decision trees: a multi-criteria approach*. Computers & Operations Research 31. 1933-1945. 2004. [Web:]  
<http://www.sciencedirect.com>.
11. **Landgrebe D, Safavian S.R.** *A Survey of Decision Tree Classifier Methodology*. IEEE Trans. Systems, Man, & Cybernetics. Vol. 21, No. 3, pp660-674, May 1991.
12. **Matlab®.** *Regression and Classification Trees*. Statistics toolbox. Version 6.5.1.
13. **Tarboton, J.N., Van Warmelo M.N. 1995.** “*The MEDUSA model*”. Internal Columbus Stainless R&D documents.
14. **MET-PRO-001 rev 1. 2004.** Internal Columbus Stainless quality control document.

15. **Meyn, U.** *General Melting and Refining Principles*. Internal Columbus Stainless Document. 2001.
16. **Mills, K.C, Fox, A.B, Lee, P.D, Sridhar, S.** 2001. *Modelling Mould Powder Behaviour in the Continuous Casting Mould*. (*In ICS*. 2001. Swansea, Wales).
17. **Mizukami, H., Yamanaka, A., Watanabe, T.** “High Temperature Deformation Behavior of Peritectic Carbon Steel during Solidification”, *ISIJ International*, Vol. 42 (2002), No. 9, pp. 964-973.
18. **Mörwald, K., Dittenberger, K., Ives, K.D.** 1998. “*Dynacs® cooling system – features and operational results*”. *Ironmaking and Steelmaking*, 25(4).
19. **Mues C, Baesens B, Files C.M, Vanthienen J.** *Decision diagrams in machine learning: an empirical study on real-life credit-risk data*. *Expert systems with Applications*, 27. 257 – 264. 2004.
20. **Normanton A.S, Barber B, Bell A, Spaccarotella A, Holappa L, Laine J, Peters H, Link N, Ors F, Lopez A, Laraudogoitia J.J.** *Developments in online surface and internal quality forecasting of continuously cast semis*. *Ironmaking and Steelmaking*, vol 31(5). 2004.
21. **Olivia Parr-Rud, C.** *Lifetime Value Predictive Modeling – Home Study Course*. [Web:] <http://www.oliviagroup.com/modeltraining.html>. [Date of access: 29/07/2005].
22. **Pentz, P.** *Fuzzy logic (A departure from the conventional approach)*. *Pulze*. May 1994. P18-20.
23. **Rhines, F.N, 1956.** *Material Science and Engineering series, “Phase Diagrams in Metallurgy Their Development and application”*. McGraw-Hill Book Company New York Toronto London. P340.
24. **Read-Hill, R.E. & Abbaschian, R.** 1994. *Physical Metallurgy Principles*. 3<sup>rd</sup> Ed. Boston, MA:PWS. 926 p.
25. **Rowlands, D.P.** “*The “secrets” of Stainless Steels*”. *Stainless steel information series No 7*. SASSDA. Publication date unknown.
26. **Siyasiya, C., Stumpf, W.E.** 2004. “*The transformation behaviour and hot strength of 3CR12 during the continuous casting process*”. Pretoria : UP. (M.Sc.).
27. **Siyasiya, C.** 22 April 2004. “*Brief Report to Columbus Stainless: The transformation behaviour and hot strength of 3CR12 during the continuous casting process*”. Internal Columbus Stainless R&D document.
28. **Siyasiya, C, Van Rooyen, GT and Stumpf, WE:** *Metallurgical factors that affect the strand width during continuous casting of DIN 1.4003 stainless steel*. *Journal of the South African Institute of Mining and Metallurgy*, vol. 105, pp. 473-481. (2005)



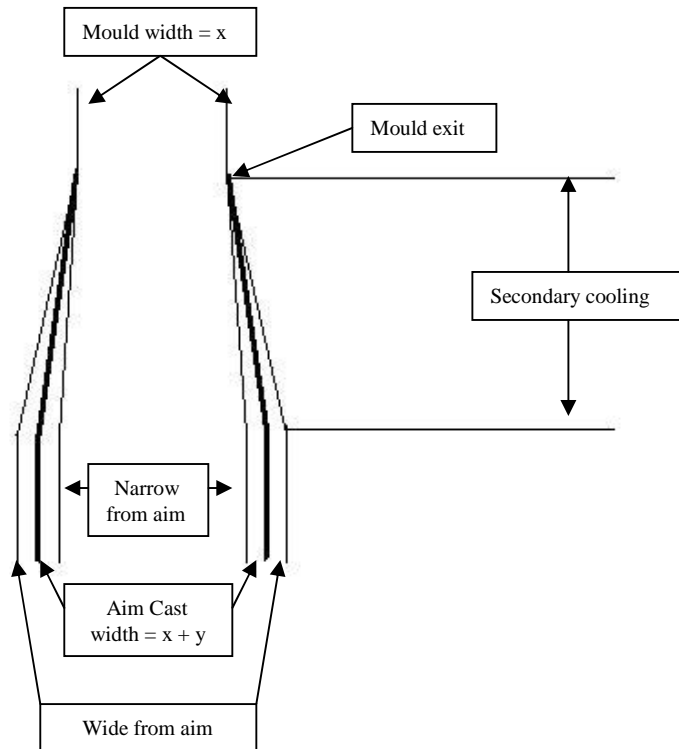
- 29. Sowell, T.** *Fuzzy Logic Tutorial - Fuzzy Logic for "Just plain folks"*. [Web:]. <http://www.fuzzy-logic.com/Ch2.htm>. [Date of acces: 01/12/2005].
- 30. Tarboton, J.N., Van Warmelo M.N. 1995.** "The MEDUSA model". Internal Columbus Stainless R&D documents.
- 31. Thearling K, Smith S, Berson A.** *An overview of Data Mining Techniques*. [Web:] <http://www.thearling.com/text/dmtechniques.htm>. [Date of access: 05/09/2005].
- 32. Vardeman, S.B. 1994.** *Statistics for engineering problem solving*. Boston, MA: PWS. 712 p.
- 33. Wang, L. 1997.** *A Course in fuzzy systems and control*. Upper Saddle River, NJ: Prentice Hall. 424 p.
- 34. Weaver, D.C.** *Applying data mining techniques to library design, lead generation and lead optimization*. Current opinion in Chemical Biology, 8:264-270. [Web:] <http://www.sciencedirect.com>. 2005.
- 35. Wikipedia. Free encyclopedia.** *Fuzzy Logic*. [Web:] [http://en.wikipedia.org/wiki/Fuzzy\\_logic](http://en.wikipedia.org/wiki/Fuzzy_logic). [ Date of access:03/03/2006].
- 36. Zadeh, L.A.** *Fuzzy sets*. Information & Control. Vol 8. pp 338-353. 1965.
- 37. Zhang B, Valentine I, Kemp Peter.** *Modelling the productivity of naturalised pasture in the North Island, New Zealand: a decision tree approach*. Ecological modelling, 186. 299 – 311. 2005. [Web:] <http://www.sciencedirect.com>
- 38. Mostert R, Brockhoff JP.** "Slab Width Control at Hoogovens Ijmuiden". Steelmaking Conference Proceedings, vol. 79. AIST, Warrendale, PA, 1996. p.19.
- 39. Assar MB, Knechtges RC, Segulin B, Crowley RC, Williams JG, Boudah R.** "Precise Slab Width Measurement and Control using Laser Technology". Steelmaking Conference Proceedings, vol. 84. AIST, Warrendale, PA, 2001. p.23.
- 40. Kocatulum B, Ozgu MR, Feagan GS, Giazzon JL.** "Slab Width and Length Control at the Burns Harbor Castors. Steelmaking Conference Proceedings, vol. 80. AIST, Warrendale, PA, 1997. p.209.
- 41. Nakamura. S. et al.,** "Slab Width Control for Hot Direct Rolling", Trans. ISIJ, Vol 28, 1988, pp. 110-116.
- 42. Evans, T.H. and Heisler, R.H.** "Slab width verification", 77<sup>th</sup> Steelmaking Conference Proceedings. Chicago March 20-23, 1994, Iron and Steel Society, Warrendale, PA, 1994, pp. 305-308.
- 43. Jones, D.A.** "Principles and prevention of Corrosion". 2<sup>nd</sup> ed. Upper Saddle River, NJ. 1996. P572.

- 44. Borneman, P.R.** “*Alloying Effects of Reactive Elements in Ferritic Stainless Steels*”. Speciality Steels and Hard Materials Conference Proceedings. 8-12 November. 1982. pp 307-314.
- 45. Piatti, G. Bernasconi, G.** “*Creep of Engineering Materials and Structures*”. Applied Science Publishers LTD. London. 1978. p420.
- 46. Smith, W.F.** “*Principles of Materials Science and Engineering*”. 2<sup>nd</sup> ed. McGraw-Hill. Singapore. 1990. P864.
- 47. Langdon, T.G.** “*Creep at low stresses: An Evaluation of Diffusion Creep and Harper-Dorn Creep as Viable Creep Mechanisms*”. Metallurgical and Materials Transactions A. Vol 33A. Feb 2002. pp 249-259.
- 48. Austin, C.R. St. John, C.R. Lindsay, R.W.** “*Creep Properties of Some Binary Solid Solutions of Ferrite*”. Metals Technology. August 1945. P84-105.
- 49. Jamieson, A. Gulvin, T.F. Baird, J.D. Barr, R.R. Preston, R.R.** “*Effects of Composition and Structure on the Creep Strength of Molybdenum Bearing Ferritic Steels*”. Publisher unknown.
- 50. Muller, B.W.** “*Evaluation of the Sag Properties in a Dual Stabilized Ferritic Stainless Steel*”. University of Cape Town. M.Sc. Thesis. 2002.

## Appendix A

*Schematic overview of the mould set-up relationship to aim cast width on 12% Chrome non stabilised ferritic stainless steel.*

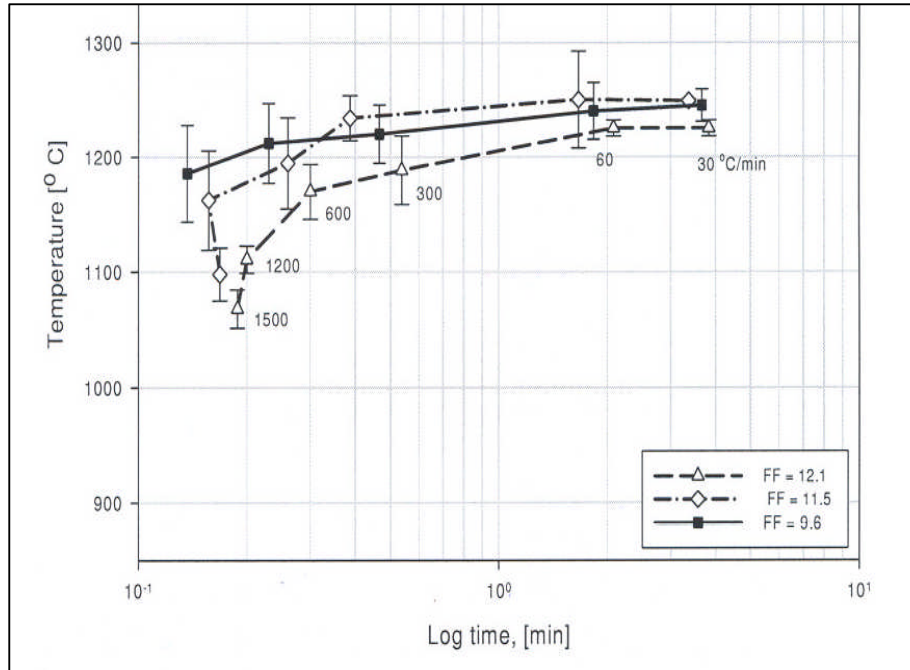
**Figure A1:** *Schematic overview of the mould set-up relationship to aim cast width on 12% Chrome non stabilised ferritic stainless steel.*



## Appendix B

*The relationship between the ferrite factor and the measured austenite start temperature  $A_{r5}$  for 3CR12.*

**Figure B1:** *The relationship between the ferrite factor and the measured austenite start temperature  $A_{r5}$  for 3CR12.*



Source: Siyasiya<sup>[26]</sup>

## Appendix C

### Equations used for metallurgical parameters

(All equations adopted from Columbus internal documentation MET-PRO-001 rev 1 and are the equations used to calculate the parameters on the Columbus systems)

Note that due to confidentiality of the equations a parameter is introduced that will be termed CLS. Certain coefficients in the equations will be substituted by this variable.

#### **Ferrite factor:**

This is the Kaltenhauser ferrite factor (1971). The equation used to calculate the ferrite factor is the following:

$$FF = (Cr + 6*Si + 8*Ti + 4*Mo + 4*Nb + 2*Al) - (4*Ni + 2*Mn + 40*(C+N)).....(C1)$$

#### **Gamma max:**

$$\begin{aligned} \text{Gamma max} = \\ (420*C + 470*N + 23*Ni + 9*Cu + 10*Mn + 180) - (11.5*Cr + 11.5*Si + 12*Mo + \\ 22*Al).....(C2) \end{aligned}$$

#### **Amax:** Also termed the austenite potential

$$Fe = 100 - (C + S + P + Mn + Si + Cu + Co + Ti + Mo + Cr + Ni + Al + Nb + V + N + B)$$

$$\begin{aligned} A_{max} = 1305.28 \\ +64.2529*Mn & \quad +41.7255*Co & \quad CLS*Ti & \quad +70.3538*Ni \\ -300.12*V & \quad +1678.48*B & \quad -14.924*Mn^2 & \quad -33.3896*Si^2 \\ +137.56*Ti^2 & \quad -27098*B^2 & \quad -0.678753*Fe^2 & \quad +6.67283*Si^3 \\ +5.16581*Cu^3 & \quad -17.4990*Ni^3 & \quad +0.005882*Fe^3 & \quad -26314.2*C*S \\ +174.992*C*Si & \quad -1003*C*Cu & \quad +2234.78*C*Ti & \quad -192.079*C*Ni \\ +1008.9*S*Si & \quad CLS*S*Ti & \quad -29554.6*S*Al & \quad +14098.5*S*N \\ -122679*S*B & \quad +18.8491*Mn*Si & \quad +15.2653*Mn*Cu & \quad CLS*Mn*Ti \\ -1.72602*Mn*Cr & \quad -30.0842*Mn*Ni & \quad -56.1561*Mn*Nb & \quad +71.2464*Mn*V....(C3) \\ CLS*Si*Ti & \quad -90.088*Si*Nb & \quad +229.274*Si*N & \quad -2183.42*Si*B \\ -5.2287*Cu*Co & \quad +124.587*Cu*Ti & \quad +83.949*Cu*Mo & \quad +4923.64*Cu*B \\ -197.289*Co*Mo & \quad +51.6542*Co*Ni & \quad -1167.69*Co*N & \quad -29.5763*Ti*Mo \\ +180.861*Ti*Ni & \quad CLS*Ti*Nb & \quad CLS*Ti*V & \quad CLS*Mo*V \\ +3729.98*Mo*B & \quad +13.5881*Cr*V & \quad +190.519*Ni*Al & \quad +1264.86*Al*Nb \\ CLS*Nb*V & \quad +1013.35*Nb*N & \quad -0.207238*C/Ti & \quad +11146.7*C/Cr \\ -0.065936*C/Nb & \quad -0.585814*Nb/C & \quad -6.46356*C/V & \quad CLS*V/C \\ +6.94717*N/V & & & \end{aligned}$$

#### **Ac1 TEMPERATURE**

$$Fe = 100 - (C + S + P + Mn + Si + Cu + Co + Ti + Mo + Cr + Ni + Al + Nb + V + N + B)$$

Ac <sub>1</sub> = 1692.52			
-60.3563*Mn	+320.661*Cu	CLS*Ti	-174.916*Cr
CLS*Ni	+942.384*Ti <sup>2</sup>	+10.3839*Cr <sup>2</sup>	-301.065*C <sup>3</sup>
+2.68843*Si <sup>3</sup>	CLS*Ti <sup>3</sup>	-0.163722*Cr <sup>3</sup>	-7338.67*C*Ti
+2430.18*C*Nb	+2559.53*S*Mn	-11400*S*Cu	CLS*S*Ni
+30125*S*Al	+6367.23*S*V	-371.607*Mn*Ti	-602.906*Mn*Al
-161.987*Mn*V	+750.507*Mn*N	CLS*Mn*B	+94.695*Si*Mo
+2.20712*Si*Cr	-12.6116*Cu*Co	-17.8166*Cu*Cr	-136.644*Cu*Ni
+296.83*Cu*Nb	CLS*Ti*Cr	-366.401*Ti*Ni	+2018.97*Ti*Al.....(C4)
CLS*Ti*Nb	-1149.71*Ti*V	CLS*Ti*N	CLS*Ti*B
+47.9937*Mo*Ni	-14.7511*Cr*Ni	+110.993*Cr*Al	+10.3851*Cr*V
-67.7747*Cr*N	-277.44*Cr*B	-1341.11*Ni*Al	+475.895*Ni*N
+8327.59*Ni*B	CLS*Al*Nb	-1.83959*Ti/C	CLS*Cr/C
-1.36387*Ti/N	-8.67322*Al/N	+2.65377*V/N	

### TRANSFORMATION RESPONSE (CR95)

$$Fe = 100 - (C + S + P + Mn + Si + Cu + Co + Ti + Mo + Cr + Ni + Al + Nb + V + N + B)$$

$$N = -6.69988$$

+53.8905*C	+4.26774*Si	+5.48015*Ti	-4.12103*Ni
CLS*N	-1.22102*Si <sup>2</sup>	-14.0669*Ti <sup>2</sup>	-0.849752*Ni <sup>2</sup>
-449.443*N <sup>2</sup>	+22.8353*Ti <sup>3</sup>	CLS*N <sup>3</sup>	CLS*B <sup>3</sup>
+0.00000843852*Fe <sup>3</sup>	CLS*C*S	-20.2257*C*Ni	+87.9968*C*Nb
-2678.61*C*B	+446.323*S*Al	-708.703*S*Nb	+1033.32*S*N
+1.03954*Mn*Cu	+1.25468*Mn*Mo	-6.07283*Si*Ti	-2.93412*Si*Mo
-0.273412*Si*Cr	+2.37037*Si*Ni	CLS*Si*N	-0.685823*Cu*Co
-9.79902*Cu*Ti	+6.26836*Cu*Mo	-21.3804*Cu*Nb	CLS*Cu*N
CLS*Co*Mo	CLS*Co*Al	+1.51008*Ti*Mo	+2.20947*Ti*Ni
-35.3195*Ti*Al	-35.7064*Ti*Nb	+84.9319*Ti*N	-12.0076*Mo*V
+0.297646*Cr*Ni	+0.322343*Cr*V	+9.06333*Ni*Nb	-11.37*Ni*N
+190.226*Ni*B	+34.9798*Nb*N	CLS*V*N	-0.005526*C/Ti
-508.07*C/Cr	+0.000557*Cr/C	-0.006649*C/Nb	CLS*Nb/C
-0.090395*V/C	CLS*S/Mn	-570.968*N/Cr	

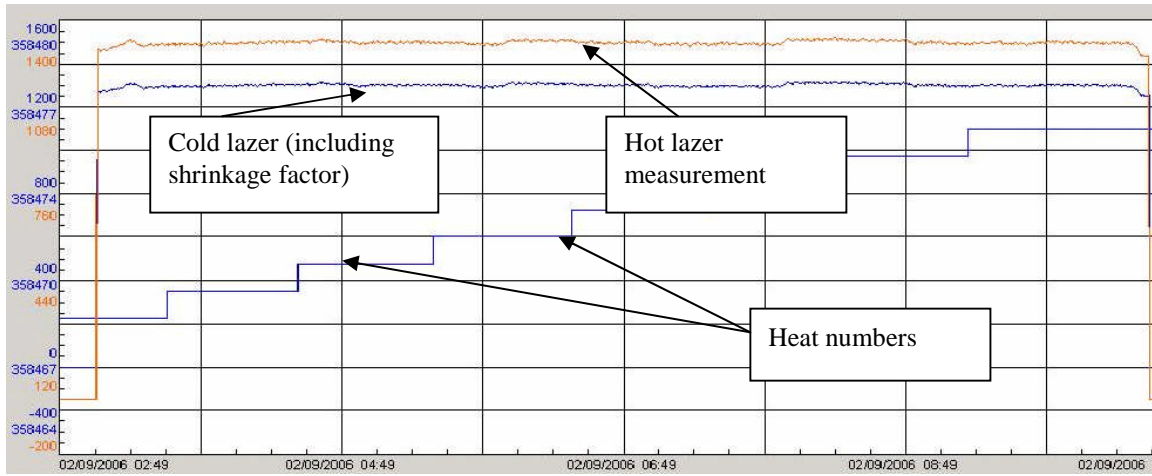
$$LNK = -34.3412$$

-1.20867*Mn	+19.2604*Ti	+0.164882*Cr	-6.5598*Ni
-597.722*N	+0.407547*Fe	+60.0472*C <sup>2</sup>	-3.70847*Si <sup>2</sup>
-15.1266*Ti <sup>2</sup>	+17.4978*V <sup>2</sup>	+735.331*N <sup>2</sup>	-107.111*C <sup>3</sup>
CLS*Mn <sup>3</sup>	+1.00523*Si <sup>3</sup>	CLS*N <sup>3</sup>	-42.3335*C*Si
-77.5978*C*Ti	+59.9413*C*Mo	+24.548*C*Ni	-141.505*S*Mo
+17.7342*Mn*Al	-8.85472*Si*Ti	+0.470194*Si*Cr	+2.27374*Si*Ni
+5.55689*Si*Nb	CLS*Cu*Ti	-21.2405*Cu*Mo	CLS*Co*Al
-202.522*Ti*N	+0.195079*Mo*Cr	CLS*Cr*N	-8.34154*Ni*Nb
+127.008*Al*V	+0.009702*C/Ti	+12.3751*S/Mn	+3310.95*N/Cr
-0.467526*Al/N	+0.009897*N/Nb		

$$CR95 = \exp((\ln(-\ln(1-0.95))-LNK)/N).....(C5)$$

## Appendix D

### Example of ASPEN graph representing laser measurements



## Appendix E

### Response surface equations

#### Definitions:

- A = AC1
- B = Gamma max
- C = CR95
- D = Amax
- E = Ferrite factor
- F = Carbon + Nitrogen

#### Linear and interaction terms combined

Width error =

$$1388.7-0.66297*A+8.0113*B+58.783*C+40.334*D-502.73*E-24053*F-0.015812*A*B-0.047074*A*C-0.057829*A*D+0.5316*A*E+32.114*A*F-0.067733*B*C-0.011128*B*D+0.51737*B*E+1.6934*B*F-0.015042*C*D-1.0978*C*E-8.0292*C*F+0.52708*D*E+30.635*D*F-438.26*E*F.....(E1)$$

#### Linear and Quadratic terms combined

Width error =

$$12553 +0.018996*(A^2)+0.0027562*(B^2)-0.07477*(C^2)+0.02058*(D^2)-1.4754*(E^2)+15450*(F^2)-30.797*A-0.6422*B+1.6815*C-4.0243*D+32.207*E-1323.2*F.....(E2)$$

#### Full Quadratic response surface

Width error =

$$-44.93A - 43.15B + 9.1744 C + 188.22D + 444.27E - 75797F +0.050142AB + 0.011877*A*C-0.11559*A*D-0.19633*A*E+77.154*A*F-0.22263*B*C-0.082837*B*D+0.3561*B*E+5.7516*B*F+0.31193*C*D-1.7269*C*E-84.5*C*F-3.3303*D*E+244.49*D*F-333.05*E*F+0.030135*(A^2)+0.028397*(B^2)-0.06548*(C^2)-0.40724*(D^2)-2.2585*(E^2)-29994*(F^2)+11876.....(E3)$$



## Appendix F

### Polynomials for fuzzy logic model

for n=1:6

```
p = polyfit(x,y,5);
Y = polyval(p,x);
R = corrcoef(Y,y);
Correlation = (R(1,2));
Corre(n) = Correlation;
```

end

WE = Width Error

$$\begin{aligned} \text{WE} &= -4.724*(AC1)^3 + 8.012*(AC1)^2 - 2.9784*AC1 + 0.5078 \dots\dots\dots (F1) \\ \text{WE} &= 2.9147*(Amax)^3 - 5.5702*(Amax)^2 + 2.4045*Amax + 0.4055 \dots\dots\dots (F2) \\ \text{WE} &= 1.3266*(CR95)^3 - 2.7794*(CR95)^2 + 2.0942*CR95 + 0.1805 \dots\dots\dots (F3) \\ \text{WE} &= 4.2285*(Gmax)^3 - 6.1782*(Gmax)^2 + 1.4921*Gmax + 0.6988 \dots\dots\dots (F4) \\ \text{WE} &= 0.6271*(FF)^2 + 0.2055*FF + 0.0753 \dots\dots\dots (F5) \\ \text{WE} &= 34.3452*(C+N)^5 - 94.048*(C+N)^4 + 92.8363*(C+N)^3 - 39.1632*(C+N)^2 + \\ &5.8697*(C+N) + 0.4508 \dots\dots\dots (F6) \end{aligned}$$

## Appendix G: Raw laser data gathering and representation

Steps for storing laser width per slab:

1. Strand is cast continuously.
2. Strand width is measured continuously with lasers
3. Strand is cut into sections (slabs) as it exits the casting machine
4. The continuous width measurement taken on each slab is averaged and stored on a database. This average value is the value used for the width of a slab.

**Figure G1:** Schematic overview of lazer measurement at CCM

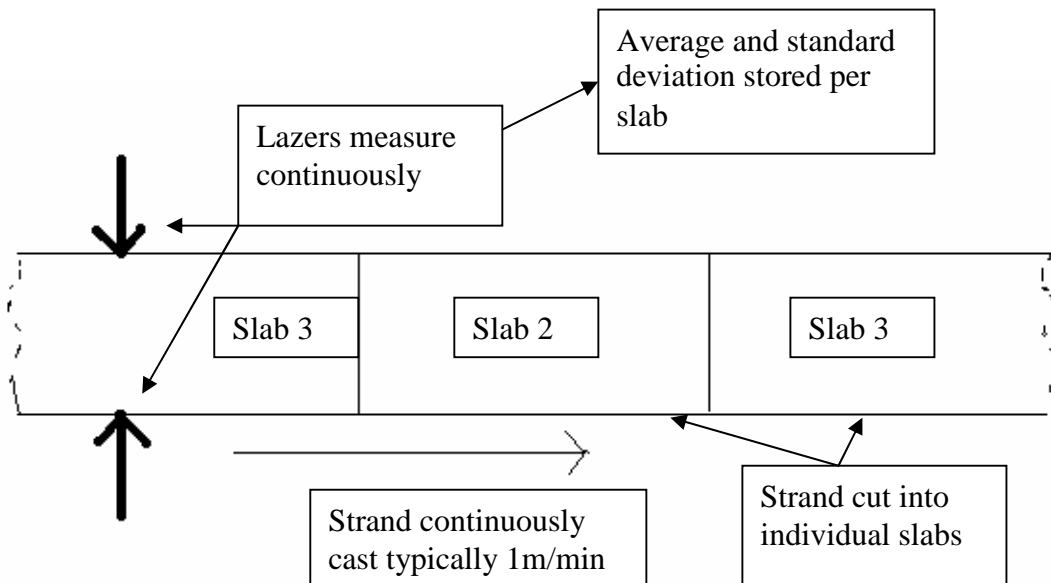
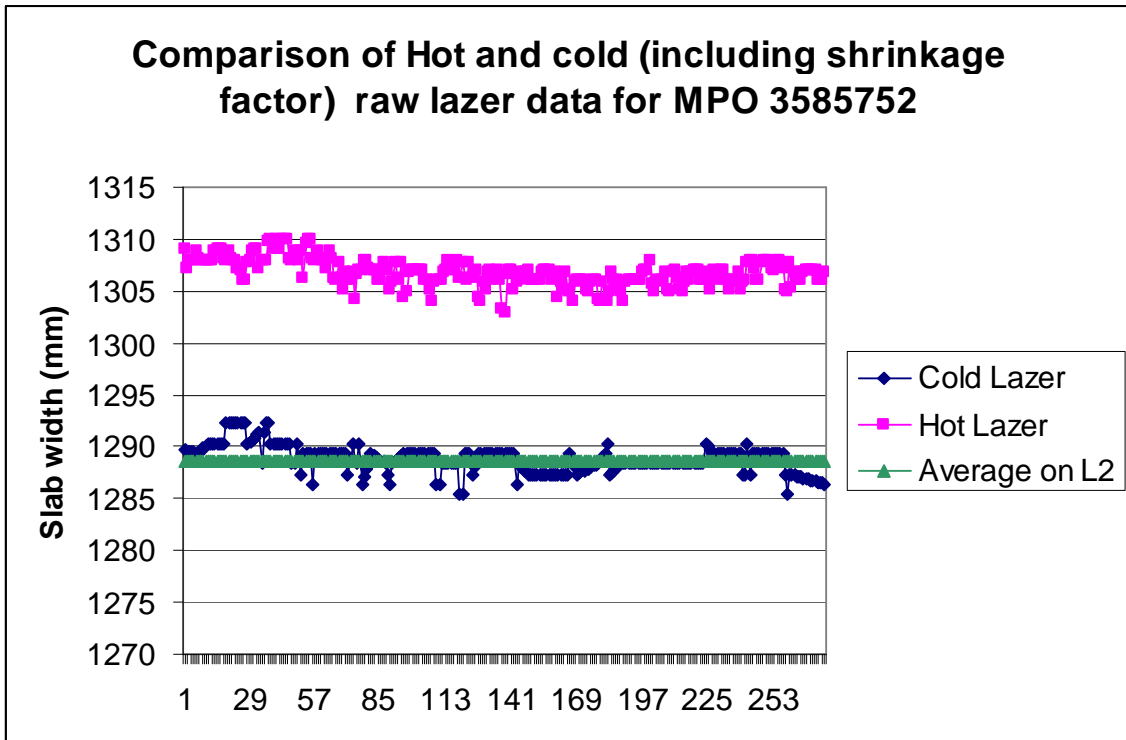


Figure G2 indicates the actual hot laser measurements and the converted laser measurements (shrinkage factor) representing the cold width of a specific slab. The laser measurements are stored every second and for this specific slab there were 275 measurements. The value that is used as representative of the width of the slab is the average of the 275 logged measurements.

**Figure G2:** Comparison of Hot and Cold (including shrinkage factor) raw lazer data for MPO 3585752.



**Table G1:** Average and standard deviation of actual lazer measurements for MPO 3585752

Stdev	Average
1.31	1288.82

## Appendix H: Raw data for the Training and Validation set

### Training set

Record #	AC1	Amax	CR95	Gmax	FF	C+N	Width Error
1	805	93.5	0.1	83.7	9.43	0.0345	-18.60
2	820	82.9	2.6	76.4	11.28	0.0405	-18.38
3	802	91.4	0.2	80.9	9.67	0.037	-17.15
4	806	98.6	0.4	79.1	8.91	0.0365	-14.93
5	815	87.5	1.4	77.2	10.86	0.0378	-13.15
6	804	98.9	0.4	79.3	8.69	0.0412	-12.98
7	812	96.3	0.4	79	8.88	0.035	-12.73
8	820	80.7	2.3	75.7	11.31	0.035	-12.56
9	824	78.7	4.3	73.6	11.33	0.0309	-11.56
10	799	89	1.1	75.7	10.52	0.035	-11.56
11	814	92	1.3	77.9	10.49	0.0375	-11.42
12	822	85.7	3.4	75.9	11.15	0.0362	-11.14
13	820	85.6	2.1	78.3	11.17	0.0461	-10.25
14	806	97.7	0.1	88.3	9.26	0.0453	-10.17
15	815	87.1	1.3	75.3	10.66	0.0335	-9.97
16	820	83.9	1.9	76.1	10.85	0.0335	-9.79
17	822	82.3	4.1	73.6	11.35	0.0333	-9.73
18	821	83	2.2	71.9	11.06	0.0295	-9.60
19	814	89.1	2.9	73.4	10.97	0.0314	-9.58
20	815	90	2	77.9	10.53	0.0405	-9.52
21	812	88.4	1.8	77.9	10.85	0.0504	-9.47
22	823	85.5	4	72.2	11.18	0.0295	-9.34
23	807	91.7	0.4	81.9	10.11	0.0385	-9.15
24	813	93.8	2.3	73.8	10.34	0.0369	-9.12
25	816	87.3	2.6	75.2	10.93	0.033	-9.07
26	818	83.5	2.6	74.1	11.13	0.0375	-8.90
27	816	84.2	2.5	73	11.32	0.037	-8.50
28	819	89	2.2	73.3	10.82	0.031	-8.39
29	821	81.9	2.6	75.4	11.16	0.0362	-7.93
30	811	90.9	1.4	76.5	10.76	0.0376	-7.84
31	819	84.7	4.7	72.8	11.22	0.0327	-7.75
32	827	78	4.5	72.4	11.76	0.0399	-7.60
33	825	75.3	4.2	69.1	11.8	0.03	-7.42
34	816	84.5	3.4	71.6	11.19	0.0336	-7.39
35	828	78.7	3.5	72.3	11.29	0.0297	-6.19
36	819	79.4	4.4	72.2	11.48	0.0355	-6.05
37	823	81.7	3.9	73.6	11.41	0.032	-5.92
38	823	81.8	1.7	77.2	11.39	0.0464	-5.67
39	822	81.2	5.1	71.2	11.21	0.028	-5.19
40	823	85.5	3.3	76.4	11.34	0.0428	-5.00
41	822	85.6	1.7	75	10.93	0.0305	-4.78
42	811	91.5	1.7	76.3	10.74	0.035	-4.57
43	823	84.5	2.1	73.9	11.21	0.029	-4.54
44	820	80.4	3.4	72.6	11.54	0.034	-4.40
45	822	79.7	4.4	73.7	11.48	0.0325	-4.29
46	818	83.3	1.4	74	11.16	0.033	-4.28
47	818	84	2.7	75.9	11.14	0.0385	-4.14

48	789	93.2	0.2	87.3	9.58	0.0485	-4.07
49	817	85	2.2	75.5	11.09	0.0349	-4.07
50	824	74	4	69.3	11.76	0.0305	-4.05
51	822	85.1	2.9	74.1	11.25	0.035	-4.02
52	818	88.2	2.1	74.3	11.09	0.0335	-3.96
53	819	82.9	3.5	72.5	11.37	0.039	-3.93
54	817	88.1	2.2	74.2	11	0.0338	-3.92
55	825	82	2.7	73.9	11.36	0.0317	-3.89
56	817	95	0.4	78	9.5	0.033	-3.89
57	816	91.6	1.2	76.9	10.59	0.0367	-3.88
58	820	81.3	2.3	73.9	11.3	0.0345	-3.87
59	800	97.1	0.8	80	10.31	0.0444	-3.80
60	826	80.4	3.6	77.3	11.45	0.045	-3.80
61	812	91.5	1.1	77.9	10.67	0.0375	-3.79
62	823	77.2	8.6	70.7	11.77	0.0278	-3.71
63	821	83.3	4.6	74.3	11.18	0.0332	-3.31
64	822	80.8	2.8	72.2	11.47	0.027	-3.17
65	808	93.5	0.9	76.2	9.9	0.0374	-3.17
66	820	83.6	3.6	74.2	11.14	0.0328	-2.89
67	828	82	4.1	75.1	11.64	0.0417	-2.80
68	816	86.5	4.4	75.5	11.08	0.0368	-2.80
69	831	82	2.8	70.7	11.19	0.023	-2.74
70	821	83.3	2.9	73	11.29	0.031	-2.63
71	816	94.9	0.4	76.7	9.38	0.0308	-2.27
72	824	82.1	3.2	71.1	11.23	0.0275	-2.01
73	820	82.9	2.2	77.8	11.38	0.048	-2.00
74	826	82.9	5.8	72.2	11.29	0.0279	-1.40
75	815	87.2	3.6	75.4	11.17	0.0388	-0.71
76	822	82	3.1	75.4	11	0.0372	-0.67
77	824	81.6	3.5	74.2	11.22	0.0325	-0.50
78	811	61.1	0.9	81.9	10.18	0.0244	-0.25
79	823	79.8	5.1	73.6	11.43	0.0328	-0.25
80	822	87.6	2.2	75.7	10.65	0.0323	0.20
81	816	90	1.2	76.1	10.83	0.0387	0.50
82	818	88.6	1.5	76.2	10.78	0.0357	0.50
83	832	80.3	5.4	77	11.53	0.0421	0.67
84	808	93	1.1	78.9	10.32	0.0434	1.00
85	826	83.7	3.9	75.2	11.59	0.0419	1.20
86	820	86.7	2.7	75.4	10.92	0.034	1.50
87	820	83.6	2.6	74.3	11.16	0.0351	2.20
88	828	78.1	6.6	70.4	11.67	0.0268	3.25
89	827	80.8	4.1	72.9	11.18	0.0303	3.80
90	829	81.7	4.5	75.2	11.75	0.0388	4.00
91	821	85.1	2	72.7	11.12	0.0353	4.14
92	818	92.6	1.1	77.4	10.39	0.035	4.17
93	816	84.3	3.3	72.9	11.16	0.03	4.40
94	819	87.8	1.9	73.8	11.13	0.0335	4.43
95	819	79.9	2.4	69.6	11.36	0.0343	4.67
96	820	88.2	2.3	78.2	11.2	0.0474	4.83
97	814	89.1	1.6	79	10.54	0.0435	5.50
98	825	73.2	8.8	66.5	12.41	0.0302	5.75
99	819	81.5	2.5	79.1	11.18	0.0498	5.80

100	824	84.2	2.6	75.2	10.93	0.0335	6.00
101	821	84.3	2.6	73.7	10.96	0.0327	6.00
102	823	79.6	4.7	72.1	11.43	0.0312	6.00
103	823	85.5	3.1	75	11.42	0.041	6.25
104	810	90.7	1.6	76.7	10.68	0.0375	6.75
105	821	83.6	3.6	73.6	11.39	0.0323	6.83
106	829	84.7	4.5	75.6	10.98	0.0292	7.00
107	824	85.5	4.8	73.4	11.11	0.0294	7.00
108	820	84.5	3	73.8	11.24	0.0344	7.20
109	823	84.5	3.3	75.1	11.34	0.0422	7.33
110	828	74.9	5.6	72.7	11.89	0.0416	8.00
111	832	76.6	7.8	72.5	11.8	0.0394	8.00
112	833	78.5	8.5	71.2	11.68	0.0271	8.00
113	826	79.9	2.2	73.2	11.19	0.0317	8.17
114	819	84.7	2.3	73.7	11.17	0.0327	8.67
115	825	83.3	4	72.9	11.36	0.0295	8.80
116	827	84	4.1	73	11.15	0.0277	8.81
117	811	60.8	0.7	77.4	10.4	0.0231	8.86
118	819	79.4	3.5	71.9	11.32	0.0305	9.00
119	799	91.5	3.5	71.8	10.88	0.0227	9.25
120	819	85.8	2.8	73.6	11.03	0.0326	9.63
121	819	73.6	3.2	70.7	11.65	0.0382	9.67
122	826	81.5	6.3	73.2	11.22	0.0292	10.00
123	827	80.4	4.4	74.1	11.7	0.0431	10.11
124	823	78.6	3.2	72.1	11.62	0.037	10.20
125	831	77.8	10	71	11.79	0.0282	10.50
126	817	86.5	1.4	74	10.93	0.0341	10.80
127	825	80.8	5.3	72.7	11.46	0.0306	11.20
128	824	80.4	5.3	73.3	11.58	0.0337	11.33
129	827	81.1	3.1	71.1	11.32	0.0286	11.40
130	824	80.3	6.4	72.6	11.46	0.0318	11.50
131	825	81.7	4.1	73	11.16	0.0297	11.60
132	816	87.2	2.5	73.4	10.98	0.0329	11.63
133	833	78.9	10.5	72.4	11.89	0.0377	12.00
134	829	78	4.9	71.8	11.9	0.0363	12.20
135	830	78.1	6.1	71.6	11.83	0.0385	12.40
136	829	77.8	6.2	72	11.47	0.0288	13.20
137	825	78.4	3.7	71.1	11.44	0.0295	13.40
138	822	78.6	3.3	73	11.44	0.035	13.50
139	827	78.6	3.5	71.8	11.97	0.0423	13.50
140	823	80.9	3.5	72.1	11.19	0.0309	13.60
141	819	85.9	2.1	75.3	11.24	0.0428	13.80
142	832	79.3	8.5	70.4	11.68	0.0251	14.33
143	829	81.6	5.2	71	11.48	0.027	14.50
144	810	89.8	1.8	73.5	10.59	0.0344	14.91
145	819	81.9	3.3	74	11.22	0.0344	15.25
146	826	83.3	4.2	74	11.5	0.0363	15.40
147	825	80.3	5.4	72.7	11.64	0.0386	15.60
148	830	79.1	8.1	72.3	11.83	0.0372	15.63
149	830	77.9	9.6	71.9	11.89	0.0278	15.80
150	829	79.4	6.1	73.6	11.99	0.04	16.00
151	819	87.5	2.5	75.5	11.18	0.0435	16.25

152	833	78.2	5.3	69.6	11.56	0.0262	16.60
153	827	73.3	8.5	67	12.04	0.0273	16.60
154	821	84.1	2.6	72.5	11.24	0.0302	16.67
155	837	73.7	9	72.5	12.57	0.0371	17.50
156	828	77.6	9.4	71.5	11.79	0.0373	17.75
157	831	77.2	7.3	70.6	11.99	0.0345	17.80
158	833	76.6	6.7	71.8	12.07	0.0378	18.00
159	835	77.2	8.8	71.3	12.06	0.0354	18.00
160	822	84.2	2.6	75.1	11.36	0.0336	18.33
161	827	81.9	4.1	74.7	11.46	0.0388	18.33
162	826	81.6	3.7	74.4	11.52	0.0415	18.40
163	821	74.9	3.2	65	11.78	0.0283	18.60
164	825	81.3	4.1	70.9	11.36	0.0295	20.03
165	825	76.7	6.5	74.2	11.26	0.0333	20.17
166	818	86.3	7.3	70	11.59	0.0254	20.31
167	828	80.3	8.1	73.6	11.71	0.0368	20.50
168	827	76.3	7.5	73.4	11.49	0.0317	20.51
169	831	76.9	8	71.1	11.6	0.0268	20.54
170	831	80.5	7	69.7	11.63	0.0248	20.57
171	825	84.5	3.9	69.6	11.45	0.0261	20.80
172	833	72.2	14.8	69	12.05	0.0286	20.83
173	824	77.3	5.7	71.8	11.56	0.0323	20.84
174	827	73.9	6.1	72.5	11.64	0.037	20.89
175	822	84.9	2.2	74	11.34	0.037	20.90
176	823	74.8	4	72.6	11.51	0.037	21.04
177	825	75.5	6.1	72.8	11.59	0.033	21.06
178	836	83.5	4.4	72.9	11.53	0.0275	21.16
179	823	77.7	4.7	71.4	11.57	0.033	21.26
180	828	80.5	4.1	71.3	11.51	0.0294	21.33
181	829	80.5	5.3	69.8	11.67	0.0264	21.33
182	819	84	2.1	75.6	11.06	0.0345	21.35
183	825	85.1	3.5	70.8	10.94	0.0242	21.43
184	831	81.5	3.8	68.9	11.48	0.0249	21.50
185	815	87.3	2.4	76.3	10.95	0.039	21.57
186	830	77.9	9.6	71.9	11.89	0.0278	21.59
187	792	86.7	0.4	75.9	10.4	0.03	21.60
188	826	81	7.1	69.8	11.45	0.025	21.60
189	828	80	5	71.4	11.49	0.0284	21.75
190	837	72.4	8.8	67.5	12.01	0.0245	21.85
191	828	79.6	4.6	69.7	11.65	0.026	21.99
192	822	81.7	3.2	73.1	11.37	0.0294	22.00
193	820	84.9	4.5	70.8	11.29	0.0281	22.10
194	821	80.7	4.2	73.7	11.29	0.0335	22.38
195	826	82.6	4.7	71.4	11.38	0.0293	22.40
196	822	84.3	4.9	71.1	11.42	0.0292	23.01
197	829	70.4	11.4	68.9	11.93	0.0285	23.05
198	821	82.6	3.1	76	11.29	0.0391	23.20
199	823	83.7	3.2	76	11.32	0.0386	23.25
200	832	76	6.2	68.9	11.76	0.0245	23.52
201	826	76.6	6.2	70.9	11.55	0.0294	23.67
202	827	75.1	2.2	61	11.99	0.0223	23.82
203	829	80.5	5.3	69.8	11.67	0.0264	23.84

204	822	81.7	3.2	73.1	11.37	0.0294	23.95
205	822	76.8	6.4	70	11.82	0.0304	24.27
206	829	80	4.8	70.4	11.4	0.0254	24.29
207	840	79.1	12.4	71.4	11.9	0.0324	24.40
208	832	79.3	8.5	70.4	11.68	0.0251	24.50
209	821	81.5	3.4	74.6	11.12	0.0339	24.63
210	824	79.2	4.2	74.1	11.34	0.0324	24.80
211	818	84	2.1	74	11.03	0.0351	24.98
212	829	75.4	5	67.9	11.87	0.028	25.00
213	833	83.7	12.2	72	11.28	0.0285	25.05
214	831	74.8	8.8	72.5	11.65	0.0305	25.10
215	823	80.7	5.2	72	11.43	0.0295	25.13
216	827	73.7	6.6	71.3	11.69	0.0295	25.25
217	825	84.5	3.9	69.6	11.45	0.0261	25.39
218	845	82.6	5.9	72.1	11.31	0.0235	25.44
219	837	71.5	9.7	70.1	11.95	0.0275	25.44
220	832	80.9	3.3	70.8	11.46	0.0286	25.54
221	828	79.2	5.5	71.3	11.59	0.029	25.62
222	821	84.5	4.2	72.3	11.29	0.0308	25.91
223	828	82.4	4	70.4	11.49	0.027	26.06
224	829	78.2	4.8	73	11.49	0.031	26.07
225	831	80.5	7	69.7	11.63	0.0248	26.40
226	825	80.8	6	72.7	11.39	0.0293	26.49
227	831	73.1	8.9	70.5	11.89	0.0305	26.79
228	824	82.9	1.9	75.7	10.97	0.0314	27.18
229	825	79.2	4.1	72.7	11.54	0.0301	27.80
230	830	78.3	7.6	73	11.85	0.0381	27.80
231	834	76.9	7.1	71.1	11.53	0.0275	28.20
232	829	79.5	5.9	72.5	11.41	0.0274	28.26
233	835	83.6	3.9	70.8	11.25	0.0255	28.66
234	828	78	4.9	69.7	11.66	0.0285	28.71
235	822	81.8	2.3	73.1	11.26	0.0305	28.77
236	831	76.5	6.2	69.4	11.8	0.026	29.04
237	831	72.8	13.6	72	11.79	0.0299	29.08
238	830	81.4	5.1	71.4	11.47	0.0264	29.17
239	830	80.5	4.5	70.3	11.53	0.026	29.91
240	827	79.2	8.7	70.3	11.64	0.0259	29.93
241	831	76.9	4.8	69.9	11.79	0.0285	30.18
242	834	74.1	17.5	68.6	11.9	0.024	30.72
243	830	81.4	5.1	71.4	11.47	0.0264	31.00
244	827	78.8	4.5	70	11.72	0.028	33.05
245	822	78.6	4.8	70.2	11.56	0.0293	33.48
246	852	78.9	0.2	73.6	11.15	0.0302	34.60
247	830	74.1	6.4	68.4	11.7	0.0283	35.03
248	826	79.1	4	69.8	11.69	0.028	36.43
249	830	75	9.9	67.7	11.98	0.0275	38.00



### Validation Set

Record #	AC1	Amax	CR95	Gmax	FF	C+N	Width Error
1	825	82.2	2.12	74.6	11.34	0.0295	-18.97
2	825	80.5	2.7	74.6	11.31	0.0365	-16.58
3	833	74.2	6.52	72.7	11.61	0.0305	-13.30
4	821	78.9	2.37	77	11.28	0.0405	-10.33
5	826	80	5.1	73	11.36	0.0285	-8.18
6	828	80.4	4.4	73.5	11.44	0.03	-6.84
7	825	82	3.2	75.6	11.21	0.036	-6.75
8	823	79.4	3.6	73.3	11.38	0.034	-6.23
9	822	84.2	3.11	74.3	11.28	0.0345	-6.03
10	829	76.6	6.6	73.1	11.61	0.032	-5.94
11	819	85.4	2.5	75.2	11.07	0.0355	-5.66
12	832	76.5	7.72	71.9	11.57	0.027	-5.38
13	831	78.2	4.2	74.1	11.36	0.0325	-5.30
14	819	84.5	2.7	75.4	11.02	0.037	-5.08
15	822	82.7	2.5	75.3	10.9	0.037	-5.07
16	819	85.9	2.2	75.2	11.06	0.037	-5.07
17	823	77.4	3.6	71.5	11.54	0.0315	-4.86
18	817	86	2.19	77.4	10.96	0.041	-4.66
19	813	83.4	1.8	73.1	11.29	0.031	-4.65
20	822	79.1	2.87	73	11.43	0.0325	-4.26
21	814	86.3	2	74	11.28	0.04	-4.14
22	819	83.3	2.6	74.6	11.31	0.0385	-3.88
23	819	86.2	1.6	76.2	10.48	0.033	-3.86
24	831	76.7	4.79	71.8	11.5	0.0255	-3.00
25	822	81.3	3.7	74.8	11.3	0.0345	-2.98
26	820	83.2	2.3	76.6	11.12	0.04	-2.59
27	822	84.4	3.4	73.6	11.15	0.031	-2.42
28	821	82.4	3.6	74.6	11.4	0.035	-1.82
29	821	78.7	2.3	75.2	11.17	0.0325	-1.30
30	819	87.9	1.43	72	10.55	0.0275	-1.12
31	822	82.5	4.2	74.1	11.44	0.038	-1.00
32	831	75.9	4.53	73.8	11.51	0.032	-0.84
33	818	87.7	2.1	75.6	10.87	0.036	-0.68
34	824	81	3.4	73.5	11.41	0.033	-0.68
35	819	85	2.9	73.5	11.18	0.036	-0.52
36	823	74.4	4.24	73.2	11.78	0.0375	-0.19
37	821	81.1	3.1	76.3	11.3	0.0385	-0.04
38	824	81.2	3.5	73.5	11.32	0.031	-0.03
39	816	85.2	2.4	73.9	11.01	0.035	0.09
40	822	83.1	2.5	75.3	11.32	0.0345	0.22
41	823	79.9	3.3	73.6	11.43	0.0365	0.24
42	819	81.4	3	75.4	11.31	0.0385	0.30
43	820	77.7	2.6	70.2	11.59	0.033	0.40
44	828	76.2	6.6	70.8	11.76	0.0275	0.48
45	822	78.7	3.8	74.5	11.36	0.0355	0.57
46	820	78.1	3.72	74.3	11.45	0.039	1.11
47	822	83.5	3.5	75.6	11.31	0.037	1.25
48	818	76.4	2.4	70.5	11.72	0.038	1.35
49	822	80.4	3.3	73.7	11.46	0.0335	1.55

50	823	81.1	2.7	74.4	11.31	0.0365	1.57
51	825	72.1	2.61	71.6	11.7	0.037	1.62
52	834	70.9	6.64	73	11.65	0.032	1.67
53	825	77.7	4.22	73	11.39	0.031	1.78
54	819	82.5	3	74	11.29	0.038	1.81
55	820	83.3	3.1	74.9	11.2	0.035	1.99
56	825	81.4	0.93	76.9	10.85	0.0315	2.00
57	817	78.3	2.5	72.6	11.38	0.0295	2.09
58	826	81.9	2.31	73.2	11.07	0.029	2.40
59	828	78.5	5.1	72.7	11.52	0.031	2.43
60	826	82.1	3.6	72.8	11.17	0.0295	2.57
61	821	82.9	2.5	75.6	11.15	0.033	3.07
62	812	80.9	1.6	73	11.37	0.0265	3.10
63	829	79.5	4.9	73.4	11.57	0.0315	3.14
64	826	75.6	1.9	73.1	11.56	0.0315	3.14
65	825	75	1.26	72.1	11.37	0.031	3.16
66	825	78.4	2.5	72.9	11.38	0.0345	3.22
67	825	80.5	3.8	74	11.29	0.032	3.33
68	828	78	3.78	75.4	11.19	0.035	3.36
69	822	80.1	5.4	73.6	11.48	0.0355	3.61
70	826	78.5	4.1	73.8	11.5	0.0325	3.62
71	827	80.1	2.84	75.8	11.36	0.036	3.68
72	827	71.6	6.9	71.4	12.46	0.031	3.74
73	828	79.3	3.4	72.5	11.47	0.036	3.75
74	822	75	3.4	74.7	12.56	0.0415	3.90
75	825	78.6	2.95	74	11.27	0.033	4.00
76	823	80.7	0.7	74.7	10.86	0.0295	4.08
77	817	83	2.3	74.6	10.96	0.036	4.17
78	824	77.6	3.5	73.7	11.56	0.0345	4.34
79	816	86.4	2.2	74.3	10.83	0.0345	4.47
80	819	82.4	2.31	75.5	11.14	0.0355	4.60
81	816	83	3.15	74.6	11.22	0.0375	4.75
82	823	76.9	2.08	72.3	11.45	0.0325	4.93
83	819	80.6	3.2	73.4	11.45	0.035	4.94
84	827	80.4	5.1	70.7	11.62	0.0285	5.28
85	822	85.5	2.8	75.1	10.96	0.0305	5.71
86	827	79.9	5.41	73.3	11.49	0.0315	5.84
87	827	78.9	3.76	73.1	11.51	0.0335	5.90
88	821	82.9	3.25	71.5	11.41	0.0285	6.13
89	823	81.2	2.4	74.4	11.02	0.031	6.15
90	822	85.6	2.7	75.8	11.05	0.0325	6.21
91	825	83.8	3.1	73.5	11.44	0.0335	6.59
92	825	80.3	3.88	74	11.31	0.032	6.81
93	824	79.9	2.8	74.8	11.33	0.0355	7.44
94	821	81.9	3.5	74.5	11.44	0.0365	7.48
95	833	73	3.11	69.6	11.65	0.0265	7.59
96	822	83.6	3.6	74.8	11.46	0.0375	7.83
97	824	80.3	3.18	75.3	11.24	0.034	7.92
98	824	76.1	4.27	74.1	11.46	0.0345	8.07
99	825	76.9	4.79	74.7	11.47	0.036	8.63
100	824	75	3.5	69.7	11.73	0.0335	9.40
101	822	82.8	3.6	74.3	11.29	0.037	9.41

102	823	84.1	2.87	73.1	11.08	0.03	9.81
103	824	79.3	10.1	70.6	11.74	0.028	10.43
104	819	77.2	3.5	74.2	11.57	0.0405	10.56
105	823	78.4	3.9	73	11.52	0.0355	10.64
106	832	74.6	6.9	71.2	11.63	0.028	10.73
107	824	79.1	4	73	11.49	0.0325	11.10
108	837	75.4	8.4	70.2	11.7	0.026	11.42
109	820	81.8	3.5	72.2	11.31	0.033	11.75
110	816	83.5	2.7	73.7	10.97	0.0335	12.16
111	826	77.3	3.1	68.8	11.74	0.031	12.52
112	831	75.2	10.06	69.5	11.77	0.0275	13.19
113	826	78.5	4.53	73.9	11.24	0.0315	14.13
114	823	84.4	1.78	72.8	11.03	0.03	14.77
115	810	75.7	2.1	73.3	11.51	0.049	15.32
116	830	78.7	6	71.6	11.31	0.028	15.90
117	829	75.1	4.63	70.5	11.73	0.027	16.72
118	810	88.8	0.97	75.9	10.39	0.032	18.15
119	824	78.6	3.1	74.9	11.25	0.034	18.42
120	829	79	6.8	69.3	11.66	0.0225	20.96
121	843	72.8	32.77	68.6	11.76	0.02	21.25
122	837	74.6	10.87	69.4	11.9	0.0251	21.54
123	822	85.2	2.6	75.4	10.81	0.0395	22.22
124	827	77.6	6.2	73.4	11.52	0.0315	26.73
125	824	81.1	3.4	73.8	11.36	0.033	28.28
126	820	81.5	2.56	71.8	11.27	0.028	29.67