STEEL SLAB SURFACE QUALITY PREDICTION USING NEURAL NETWORKS

M.J. ACKERMAN, Pr.Eng., B.Eng.

Dissertation submitted in partial fulfilment of the requirements for the degree Magister Ingeneriae in the School of Mechanical and Materials Engineering at the Potchefstroomse Universiteit vir Christelike Hoër Onderwys

Supervisor: **Prof. C.G. Du Toit**

Assistant Supervisor: **Prof. D.A. De Waal**

November 2003

Potchefstroom

ABSTRACT

Columbus Stainless grinds the majority of the steel slabs that are produced to improve the surface quality. However, the surface quality of some slabs is good enough not to be ground. If a reliable method can be found to identify these slabs, the production costs associated with grinding can be saved.

Initially slabs were selected manually based on knowledge of the process parameters that affect the steel surface quality. This was not successful and may have been due to the interaction between variables and non-linear effects that were not taken into account. A neural network approach was therefore considered.

A multilayer perceptron neural network was used for defect prediction. The neural network is trained by repeatedly attempting to match input data to the corresponding output data. Linear regression and decision tree models were also trained for comparison.

The neural networks performed the best. The effectiveness of the models was tested using a test data set (data not used during the training of the model) and the neural networks gave high levels of accuracy (greater than 75% for both defect and no-defect cases). A committee of models was also trained, but this did not improve the prediction accuracy.

Neural networks provided a powerful tool to predict the slab surface quality. This has enabled Columbus Stainless to limit the deterioration in the steel quality associated with non-grinding of slabs.

OPSOMMING

Die meeste platblokke by Columbus Stainless word geslyp om die beste oppervlak kwaliteit te verseker. Dit is 'n duur proses en daar is gevind dat baie van die platblokke se oppervlak kwaliteit aanvaarbaar is en dat die slyp daarvan onnodig is. Indien 'n geskikte metode gevind kan word om te besluit watter platblokke nie geslyp hoef te word nie kan die maatskappy baie spaar.

Aanvanklik is van die kennis van die veranderlikes wat staal oppervlak kwaliteit beïnvloed gebuik gemaak om platblokke te kies. Dit was egter onsuksesvol en moontlike redes hiervoor kan die interaksie tussen veranderlikes wees asook nie-liniêre veranderlikes wat nie in ag geneem is nie. Daar is besluit om die geskiktheid van 'n neurale netwerk vir voorspelling te ondersoek.

'n Multilaag perseptron neurale netwerk is gebruik om defekte te voorspel. Die neurale netwerk is opgelei deur die voorspellings vanaf insette met die werklike uitsette te vergelyk en die fout te minimeer. Liniêre regressie en besluitnemingsbome is ook opgelei en met die neurale netwerke vergelyk.

Die neurale netwerke het die beste resultate gelewer. Die effektiwiteit van die modelle is getoets deur van 'n toets datastel (data wat nie tydens die opleiding gebruik is nie) gebruik te maak. Die neurale netwerke was in staat om beide defekte en platblokke met geen defekte met 'n akkuraatheid van meer as 75% te voorspel. 'n Komitee van netwerke is ook opgelei, maar het nie die voorspellings akkuraatheid verbeter nie.

Daar is gevind dat neurale netwerke 'n kragtige metode bied om platblok oppervlak kwaliteit te voorspel. Dit het Columbus Stainless instaat gestel om die aantal ongeslypte platblokke te verhoog sonder dat die staal kwaliteit noemenswaardig verswak het.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my supervisor, Prof. C.G. Du Toit, and assistant supervisor, Prof. D.A. De Waal, for their guidance during this study.

I would also like to thank my wife and children for their support and patience.

TABLE OF CONTENTS

NOMENCLATURE

$\mathbf{a}^{m+1}$	Output from neural network nodes
$\mathbf{f}^{m+1}$	Transfer function
$\mathbf{W}^{m+1}$	Input weights to the node
$\mathbf{a}^{m}$	Input from the previous layer
$\mathbf{b}$	Bias
M	Number of layers in the network
$\mathbf{t}$	Expected output
$\mathbf{a}$	Predicted output
$\mathbf{s}^{M}$	Errors calculated for the output layer
$\mathbf{F}(\mathbf{n}^{M})$, $\mathbf{F}^{m}(\mathbf{n}^{m})$	Derivative of the node output to the input weights
$\mathbf{s}^{m}$	Errors calculated for layer
$\mathbf{s}^{m+1}$	Errors of the next layer
$(\mathbf{W}^{m+1})^{T}$	Transpose of the output weight vector
K	Iteration number
$\mathbf{W}^{m}$	Weight vector
$\mathbf{b}^{m}$	Bias vector
α	Learning rate
γ	Momentum coefficient
$\mathbf{J}$	Jacobian matrix
μ_k	Constant that controls the learning rate
$\mathbf{x}_k$	Parameter vector that contains the weights and biases
$\mathbf{I}$	Identity matrix
$\mathbf{v}$	Vector of the calculated errors

LIST OF TABLES

TABLE OF FIGURES

CHAPTER 1 INTRODUCTION

1.1 Background

Stainless steel surface quality is important for the client and the majority of the steel produced at Columbus Stainless is therefore ground. Grinding refers to the removal of the outer millimetre of the slab surface. This grinding adds to production costs and also results in product yield losses. The saving from not grinding is significant and it is therefore important to find methods to reduce grinding.

Metallurgical defects are caused by process deviations in the steel plant. These defects include inclusions, line inclusions and skin laminations. Defects are only detected once the steel is processed at the finishing lines. This feedback is available from approximately four weeks after the steel is produced in the steel plant. At that stage the material with serious defects is scrapped or reclassified to a lower grade. It is therefore important that, when a slab is not ground, every effort is made to ensure that the slab does not have serious surface defects.

Various steelmaking parameters are known to influence the final quality of the steel. These include amongst others the steel chemistry, mould cooling, heat transients during casting, casting speed, mould level fluctuations, steel superheat, mould flow and powder lubrication (Capotosti *et al.*,1994:201). In many studies only the effect of mould heat transfer is considered to indicate slab surface defects (Capotosti *et al.*,1994:201 and Bellemo *et al.*,1998:199). In this study the contribution of other steelmaking process parameters is assessed.

The effect of all the processes in the steel plant on quality is considered. The scope of the study includes the converter, the rinse station and the continuous caster. The only unit that is excluded is the electric arc furnace because it is far removed from the casting process and is not believed to influence the final product quality. A few variables at the converter are considered. These include the final silicon content of the steel and whether the heat was reblown or not.

The rinse station operation is important for quality control. Excessive stirring may result in slag entrainment. Too little stirring may cause poor mixing and dissolution of material added to the steel. These factors are known to contribute to poor quality. After stirring, the period of time that elapses determines the amount of thermal stratification that occurs. If this period is too short or too long, temperature control in the tundish may be difficult (Chakaborty & Sahai,1992:140). As the temperature decreases, solubility products are formed and this contributes to submerged entry nozzle clogging and quality problems.

Many operating conditions at the continuous caster are important for quality. Two of the critical areas are the flow in the mould and the heat transfer. The velocity of the steel in the mould can cause mould flux entrainment. Several researchers have developed criteria for the critical velocities in the mould above which entrainment can be expected (Jimbo *et al.*,1991:119 & Kubota *et al.*,1990:359). Since the casting speed is changed often during a cast, the possibility exists that these critical levels may be exceeded. The actual critical velocities are not known but the effect thereof has to be included in the model to be developed.

The importance of the heat transfer in the mould has already been mentioned. The mould powder that is used for lubrication in the mould controls the heat transfer. The powder should also have the capacity to remove inclusions from the steel (Mills,1991:121) and this is primarily controlled by the composition of the powder. The combination of the composition and viscosity controls the thermodynamics and the kinetics of oxide absorption into the mould powder. The rate of inclusion removal is designed to be rapid, indicating that if contact between the inclusions and mould powder can be established, inclusions will be removed.

Several statistical studies have been done at Columbus Stainless and although these studies gave some indication of parameters that affect quality, the results were not repeatable and no general applicable models could be constructed. One exception is a solubility product model that was developed at Columbus Stainless (Nunnington & Sutcliffe, 2001:24). This model provides an indication of steel cleanliness (that is the levels of the oxides and nitrides in the steel). A good correlation between solubility product levels and final product quality was found in previous studies at Columbus Stainless.

A manual approach for slab selection using knowledge of the steel plant process was not successful in the past. This may have been due to the interaction between parameters or non-linear dependencies that were not taken into account. A non-linear (neural network) modelling approach is therefore considered.

Neural networks have been successfully applied to various fields in the metallurgical industry (Van Deventer *et al.*,1994:793). These models are especially useful where the relationship between dependent and independent variables are not known or where it is not easy to determine it analytically. Neural networks have been successfully applied in the pyrometallurgical industry (gold recovery and the recovery of lead and zinc from flue dust (see Van Deventer *et al.*,1994:793)), steel plate processing (Singh *et al.*,1998:355) and surface quality prediction using mould thermocouples (Bellomo *et al.*,1998:199). In all these cases the models were able to generalise, that is correctly predict outputs for data not seen before by the model.

The most frequently used architecture in neural network development is the multilayer perceptron due to its flexibility in dealing with different types of data. In some cases cluster analysis has been used in combination with neural networks to improve accuracy and the rate of training. The prediction accuracy can also be improved by using a committee of models (Freeman & Skapura,1991:109). These approaches are considered to find an appropriate model for quality prediction.

Network stability is important during the training process (Hagan *et al.*,1996:17-1). Convergence may be difficult when the process that is modelled is complex. In this study, the initial model included up to fifty input variables. As the number of variables increases, there is a corresponding increase in the volume of data required. Every effort should therefore be made to ensure that only significant variables are included in the model.

The size of the network is also important. No guidelines for the optimum size of the network exist, but three layers are normally sufficient (Freeman & Skapura,1991:104). The units in the hidden layer should be as few as possible to speed up the training of the model. It is envisaged that the model will have to be retrained as the conditions in the steel plant change. It is therefore essential that the network is as simple as possible.

1.2 Problem statement

A manual approach in slab selection for non-grinding was not very successful. There is a need for a model to predict slab surface quality that can be used for selection of slabs for non-grinding.

1.3 Goals

The goal of this study is to develop a non-linear (neural network) model that can be used to predict slab surface quality. The non-linear model is compared with a linear model to determine whether there is any gain in using the more complex model. Another goal is to determine whether a committee of models improves the prediction accuracy.

1.4 Method of investigation

The variables in the steel plant that affect the slab surface quality were determined using a literature study. Some of these variables are not measurable and possible indirect indicators have to be identified in these cases. Where variables cannot be included, these exclusions have to be taken into account during interpretation of the model results.

Data for a period of approximately six months were used in the analysis. The data were filtered to remove all the records with missing and anomalous values. This resulted in a substantial reduction in the available data.

A correlation matrix can be used to reduce the number of variables. This form of analysis indicates the interdependence between variables and therefore variables that can be excluded without losing prediction accuracy.

Different neural network training algorithms were considered. Decision tree and regression models were trained using the same data set and results were compared. The prediction accuracy of the models was assessed using a test data set (10% of the original data set). All the models were implemented in SAS Enterprise Miner.

1.5 Contribution

Columbus Stainless will be the main beneficiary of this study. Through the selective grinding of slabs an annual saving of R24 million is possible (that is if half of the current slabs produced is not ground). Savings from both production costs and improved yields are possible. By decreasing the number of slabs that needs grinding, the loading on the grinders will decrease and this may result in the postponement of future capital expenditure for the erection of additional capacity.

The importance of having an accurate method to predict the slab surface quality is highlighted by the costs involved with the incorrect selection of a slab. The loss when scrapping material is in the order of R3000/ton. This would mean that if only 2.5% more material is scrapped due to non-grinding, no benefit will be achieved from non-grinding. An accurate prediction model is therefore required.

1.6 Overview of document

The information gathered in the literature study is first discussed. The steelmaking process and factors that influence slab quality are then considered in detail. Sections on the application of neural networks in the process industry and the data mining process are also included in this chapter.

Neural networks are discussed in more detail in the next chapter. Various training algorithms are considered. The application of neural networks and other supervised learning techniques in SAS Enterprise Miner are also discussed in detail. Decision trees, linear regression and a committee of models are also considered.

Before the construction of the final model, data are filtered and the variable list is finalised. This is followed by the analysis of the model results, comparison of the different modelling techniques and final conclusions.

CHAPTER 2 LITERATURE STUDY

In the first section of this chapter the steelmaking process is discussed with special emphasis on the factors that influence the steel quality. In the second section an introduction to the data mining process is given.

2.1 Factors affecting steel surface quality

One of the goals during the steelmaking process is to produce quality steel. Figure 1 gives a layout of the Columbus Stainless steel plant. Scrap and ferrochrome are melted in the electric arc furnace and the liquid steel is then transferred to the CLU converter where carbon is removed by blowing mixtures of oxygen and argon. Final chemistry adjustments are done at the rinse station. When the steel reaches the target temperature, the ladle is transferred to the continuous caster. The steel moves through a tundish and into a copper mould where the steel starts to solidify. Complete solidification occurs in the secondary cooling zone of the caster. The steel slabs are ground at the grinders where the outer steel layer is removed. The slabs are then reheated in the reheat furnace and hot rolled.

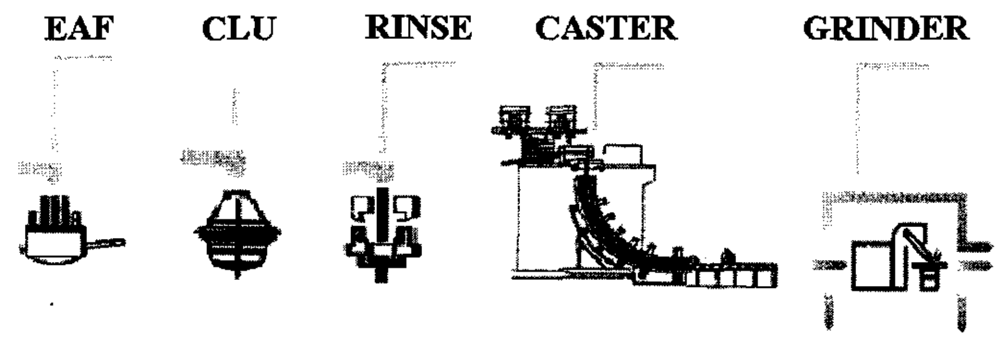


Figure 1 Steel plant layout

The variables that influence steel quality will be discussed next. Some of these variables are not directly related to metallurgical defects but may provide an indication of abnormal conditions that contribute to defect formation (mould heat flux for example).

2.1.1 Reduction efficiency

The control of the steel's oxygen activity in the converter is important for steel quality (Nunnington & Sutcliffe, 2001:15). This control is performed by adding ferrosilicon at the end of the carbon removal process. High silicon levels ensure that the silicon controls the equilibrium oxygen level in the steel and therefore limits the formation of alumina and titania inclusions in the steel. The formation of these oxides is discussed in section 2.3.

2.1.2 Rinse station practice

Inclusions may form at the rinse station due to deoxidation or reoxidation reactions, slag entrapment, slag-metal reactions and slag-refractory reactions (Dyson *et al.*,1998:279). The removal of these inclusions is a complex process of agglomeration by collision, flotation and attachment to bubbles (Miki *et al.*,32). The rate of the reaction at the steel-slag interface also influences the removal of the inclusions. For inclusion removal a minimum stirring-rate is required but if the utilized rate is too high, slag entrapment can occur.

The mixing time and dissolution of material additions at the rinse station are also important. Mixing depends on the intensity of agitation (Zhu *et al*,1995:473) and sufficient time should be allowed for mixing and the dissolution of additions into the steel. The rinse time after final additions is therefore important (Austin *et al.*, 1992:198).

2.1.3 Steel cleanliness

A model was developed at Columbus Stainless to predict the probability of solubility products forming at different stages in the steelmaking process (Nunnington & Sutcliffe, 2001:24). This model assumes equilibrium conditions and is a good approximation for conditions in the ladle and tundish. The model was adapted to calculate the precipitation of alumina, titanium oxide, silica and titanium nitride as a function of temperature. It was assumed that all the oxides and nitrides with the potential to precipitate at a particular temperature did precipitate. A mass balance was done at every temperature step and the steel chemistry was continually updated as the oxides / nitrides precipitated. Typical results of precipitation as a function of temperature can be seen in Figure 2.

Figure 2 compares the solubility products precipitating from a titanium-stabilised steel. Figure 2 (a) shows levels at start-up (high oxygen and nitrogen levels) just after ladle open. Figure 2 (b) shows lower levels during steady state casting (calculations from tundish steel temperature down to the solidus temperature). The higher oxygen and nitrogen levels were probably due to air ingress during start-up. The model predicted the formation of 25% more solubility products for the first case.

The effect of air entrainment had a significant effect on the steel cleanliness. Although some of the inclusions associated with the entrained air were probably removed, there was still a significant amount flowing into the mould. Note that these calculations were done using the total oxygen measurements next to the stopper rod.

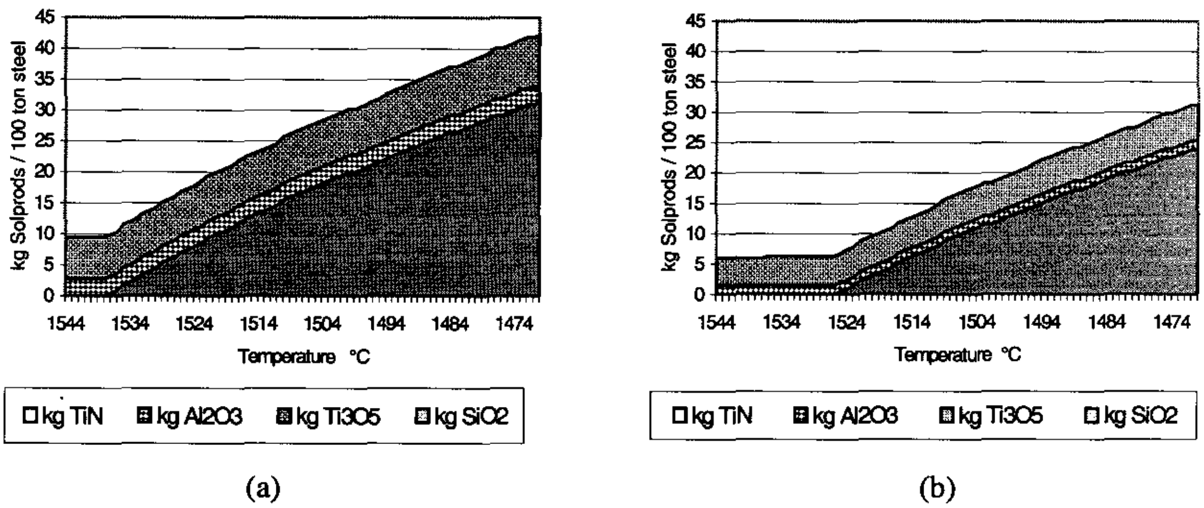


Figure 2 Predicted solubility products precipitating during transfer from tundish to mould for steel type 409 (Figure (a) for start-up conditions and Figure (b) for steady state)

As a rule the solubility product levels for steel type 304 is lower than that of titanium stabilised steels and therefore also the inclusion levels. Problems like air ingress may however result in high alumina levels in these steels. This will contribute to clogging of the submerged entry nozzle and poor slab surface quality.

2.1.4 Temperature control

The other critical variable that controls the precipitation of oxides in the steel is the steel temperature. When steel is tapped from the converter into a ladle the temperature drops due to convection and radiation heat losses. Heat is also continuously removed in the ladle through the refractory lining and the slag layer on top of the steel. This heat loss through the sidewalls is a function of the initial refractory heat content. At the rinse station the steel temperature decreases further due to additions and gas stirring.

The time between rinse end and start of cast is difficult to control due to scheduling. A ladle can stand on the floor for a period between five minutes and two hours. During this time the heat loss through the ladle lining and the slag cover result in natural convection currents in the steel. The temperature of the ladle exit stream is therefore a function of the time between rinse end and start of cast, the ladle refractory heat content, the thickness of the slag cover and the casting speed.

Various researchers have shown that by controlling the amount of stratification (temperature layering of the steel) that occurs in the ladle, the exit steel temperature can be controlled (Hlinka & Miller,1970:133 and Chakaborty & Sahai,1992:149). Heat loss causes the steel close to the ladle wall to cool down, increase in density and flow to the ladle bottom (Olika *et al.*,1996:363). The colder steel collects at the ladle bottom and a convection current is induced. Steel flow upwards through the centre of the ladle develops.

Chakaborty and Sahai (1992:147) found, contrary to expectation, that a longer time between end of rinse and start of cast was better for steel temperature control. The outlet temperature varied by 4 °C when the holding time was 20 minutes and 8 °C when the holding time was 5 minutes before ladle opening. From these results it was concluded that a longer holding time was beneficial for temperature control.

Both water and mathematical models have been used to study ladle stratification. These models have been validated using plant tracer trial studies. Chakaborty and Sahai (1992:136) used a two dimensional mathematical model to study temperature stratification. Their results indicated that the slag layer thickness played an important role in the process. With an insulated top the teeming temperature remained almost constant for a 45 minute cast. With no slag cover, the teeming temperature drop was significant. Their results confirmed the findings of Hlinka and Miller (1970:130). Other researchers claimed that slag thickness had only a small effect (Austin *et al.*,1992:199). They concluded that this was the case because conductive heat transfer is an order of magnitude lower than radiation. Their results also showed that a lid did not reduce temperature stratification significantly if the steel was covered with slag. A thin layer of slag should therefore provide enough insulation and a lid acts in the same manner as the slag. Hlinka *et al.* (1970:124) disagreed and stated that a slag layer of 200 mm was necessary to suppress stratification. At Columbus Stainless a slag cover is added and a lid is placed on top of the ladle to limit heat loss during casting.

The temperature stratification increased as the initial refractory heat content decreased (Grip *et al.*,1997:1088). Deb *et al.* (2001:205) found that a 100 °C variation in the ladle preheat temperature (measured on the inside of the ladle) resulted in a 6 °C change in the steel temperature. Better preheating ensured that a ladle reached steady state conditions sooner. It does, however, still take at least six heats for a newly rebuilt ladle to reach saturation (Austin *et al.*,1992:198).

Refractory wear increases the heat loss from the ladle. A small amount of wear had a small effect, but heat loss increased substantially as the wear increased (Austin *et al.*,1992:198). The type of brick and its high temperature thermal properties also influence the steel heat loss during a cast. A ± 0.4W/mK change in thermal conductivity leads to a ± 1.5 °C variation in the steel temperature (Deb *et al.*, 2001:205).

Plant tracer trials have shown that there are pockets of steel that only leaves the ladle towards the end of teeming (Grip *et al.*,1997:1087). This is also seen at Columbus where the steel temperature decreases towards the end of teeming. These tracer trials also confirmed that steel flows not only from the bottom but also from higher levels in the ladle during initial teeming.

No direct measurements of temperature stratification in the ladle are available in the plant. Heat loss through the ladle refractory and slag layers are also unknown. It was however shown that the time between end of rinse and start of cast provides a good indication of temperature control during a cast (Chakaborty & Sahai, 1992:147).

While at the rinse station, the steel temperature is measured. This information as well as the argon stir rate provide a good indication of heat loss and slag surface instability in the ladle. These two conditions were combined in one parameter for the rinse station: temperature decrease divided by the argon stir rate.

The steel temperature during casting has a significant influence on the steel quality. The superheat (temperature above the steel's liquidus temperature) affects the thickness of the chill, columnar and equiaxed zones during solidification (Laing *et al.*,1997:521). The size of the columnar and equiaxed zones is important for the steel's internal quality. Columnar dendrites are more susceptible to the formation of internal cracks and a big columnar zone can increase the severity of centreline segregation and porosity.

The steel chemistry and the temperature can be used to control inclusion levels. The solubility level of oxygen and nitrogen in steel decreases with temperature and higher inclusion levels are found at lower temperatures. Low temperature and the associated inclusion levels contribute to SEN clogging (Nunnington & Sutcliffe,2001:8). In extreme cases the clogging may cause a cast to be aborted.

From steel breakout shell investigations it was seen that most of the superheat is removed on the off-corner wide sides and the mould narrow faces. This agrees with findings in literature (Thomas *et al.*,1990:131). Most of the superheat is removed from a distance 0.4 m below the meniscus (evidence from breakout shells). This position correlates with the steel jet impact point on the shell. Most of the superheat is dissipated through shell thinning and only a small portion contributes to an increased heat flux. A too high superheat may therefore result in a breakout because of excessive shell thinning (Flint, 1990:481).

According to Flint (1990:483) a change in SEN depth has a small influence on the temperature at the meniscus. The steel superheat has a much greater effect. Flint predicted an increase of 4 - 8 °C at the meniscus for a 16 - 32 °C increase in incoming steel superheat. The length of time over which the total superheat was removed increased from 2.25 m to 3.5 m down the slab length as the superheat was increased from 16 to 32 °C.

Steel temperature should therefore be included in the model since it not only influences solubility product levels but also meniscus temperature in the mould as well as the solidification structure of the steel.

2.1.5 Air ingress and inclusion removal from the tundish

Two major sources of inclusions are found in the tundish, namely re-oxidation and slag entrainment. Casting without a shroud is a source of re-oxidation and nitrogen pickup. This can also occur during ladle changes or when the ladle tap hole has to be opened with a poke or oxygen lance.

Tundish furniture is used to remove as many of the inclusions as possible. The furniture increases the residence time of steel in the tundish (and therefore time for inclusions to float out) and provides surface directed flow to improve inclusion removal. The problem is that conditions in the tundish are seldom at steady state. The tundish temperature varies during a heat and ladle change brings further instability into the system. During ladle change the tundish level drops. Some researchers have shown that flow reversal and short-circuiting can occur when colder steel flows into a hotter tundish bath. Other researchers maintain that these periods are short and do not affect the overall effectiveness of the furniture (Morales *et al*, 2000:981). In a study done at Columbus Stainless plant trials indicated no benefit from using tundish furniture in non-titanium stabilised casts (Ackerman & Orban, 2001:417).

Air ingress cannot be measured directly, but steel chemistry and the corresponding solubility product levels should provide a good indication. The tundish weight can also be used to give an indication of unstable conditions in the tundish. The steady state operating tundish weight is approximately 23 tons.

2.1.6 Mould heat transfer

Heat is removed through the copper mould as the steel solidifies. The level of the heat flux during a cast is one of the useful variables that may indicate surface quality problems. The ideal would be to use temperatures measured by thermocouples mounted in the copper plates, but this falls outside the scope of the current study. The overall heat transfer will however be used in this study.

Figure 3 shows the factors that contribute to the mould heat transfer. These factors can be divided into three main areas, namely caster operating conditions, mould powder properties and the steel composition. These factors, as well as their combined effect on mould heat transfer, will be discussed.

2.1.6.1 Caster operating conditions

Different operating conditions are used for casting different steel grades. Some of the operating conditions such as mould oscillator conditions and mould cooling-water flow are fixed and similar for all grades. Others such as superheat, casting speed and mould taper are a function of the steel grade.

Casting speed has the biggest influence on the heat transfer. According to Emling and Dawson (1991:198) a 0.1 m/min increase resulted in a 5% increase in the overall heat transfer. This was confirmed by other researchers (Ho, 1992:6 and Mahapatra *et al.*,1991:879). This increase is due to a decrease in mould powder consumption.

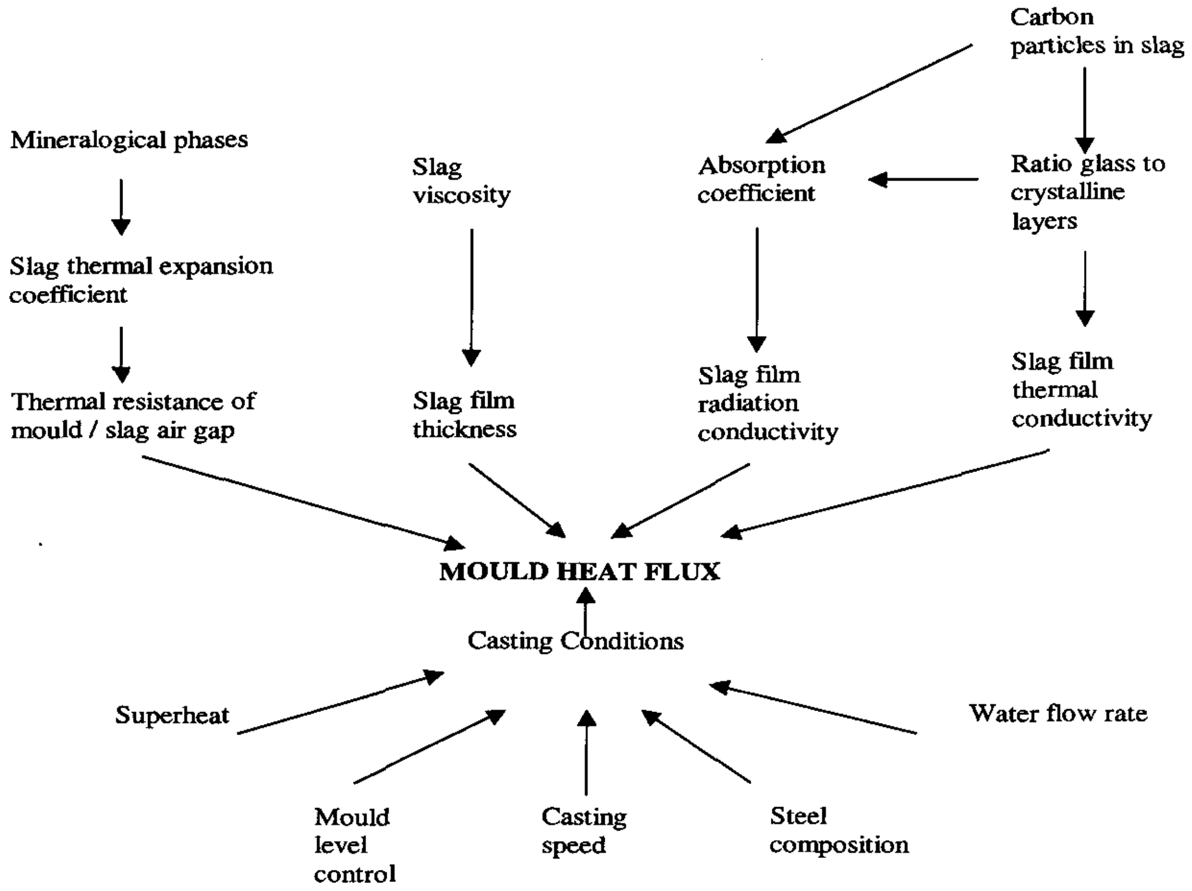


Figure 3 Factors affecting the heat transfer in the mould (from Mills (1991:124))

Lu *et al.* (1995:363) concluded from a study done on a slab caster mould that the effect of water flow, water temperature and copper mould thickness were masked by the contribution of superheat, casting speed and interface heat resistance. Wolf (1980:714) states that as the water velocity in the mould slots is increased above 6 m/s the heat flux is decreased. Mahapatra *et al.* (1991:883) conclude that the velocity of the cooling water affects the slag rim thickness. An increase in the water velocity increases the water heat transfer coefficient and lowers the copper hot face temperature. This results in a thicker slag rim that restricts heat flow.

In September 1996 the water flow in the Columbus Stainless plant mould wide side was increased from 3200 l/min to 4000 l/min. The effect of this step change was investigated statistically. The effect of superheat, casting speed and water inlet temperature was taken into account. The change resulted in a 38% water temperature decrease. The net result of the 25% increase in the mould water flow was an 8.7% decrease in mould heat flux (see Table 1).

Water flow l/min	Delta Temperature °C	Heat flux MW/m^2	Heat flux % change
3200	5.75	1.229	
4000	4.2	1.122	-8.7%

Table 1 Effect of mould water flow on mould heat flux

2.1.6.2 Mould powder properties

The use of a mould powder is critical to produce a good quality slab. Mould powder is continuously fed at the top of the mould during casting. The functions of the mould powder include the following: to prevent steel oxidation, to provide thermal insulation, to lubricate the gap between the steel and the copper, to provide uniform heat transfer at the meniscus and to absorb inclusions from the steel (Mills,1991:121).

Due to the heat removal through the water-cooled copper plates, the mould flux solidifies as a glassy layer next to the copper wall. A crystalline layer can form next to this layer depending on the slag chemistry. This layer is important since it controls the heat transfer to the mould. The crystals scatter the radiation and reduce the heat transfer. The mould flux crystallization temperature can therefore be used to control the heat removal in the mould.

A low mould powder consumption rate can prevent liquid lubrication and increase the friction in the mould. The factors affecting the consumption are the viscosity, the solidification temperature of the mould powder, the heat removal and the oscillation characteristics (Mills,1997:36). Ogibayashi *et al.* (1987:4) showed that the infiltration of mould flux varied with the viscosity of the powder. The powder film thickness, heat transfer and mould temperature variation is at a minimum when the viscosity times casting speed are in the range of 1 to 3.5.

The mould powder consumption also depends on the oscillating practice used. Wolf (1997:254) suggests that the negative strip time (time that the mould moves downward faster than the casting speed) controls mould powder consumption, but that the drag due to the upward movement of the mould may pull back a portion of the mould flux (mould flux refers to melted mould powder). The oscillation practice also influences oscillating mark formation on the slab. The oscillation marks are filled with mould flux and represent the major part of the powder consumption.

The depth of the slag pool on top of the liquid steel is important to ensure constant infiltration of mould flux into the steel-copper gap. The mould level behaviour is also critical since a standing wave of major mould level fluctuations may cut off the supply to the steel mould cavity. The depth of the slag pool was determined using "slag dip" tests at Columbus Stainless. Figure 4 shows an example of these measurements. The slag thickness varied with position in the mould. These measurements show an area of low liquid powder thickness towards the narrow side (due to the standing wave as mentioned before) and may have contributed to powder feeding problems.

A good mould flux has the capacity to remove inclusions from the steel. If the solubility level for alumina and titania is exceeded, mould lubrication breaks down and may result in a breakout. Different mould powders have different capacities for oxides that may influence the eventual quality of the steel.

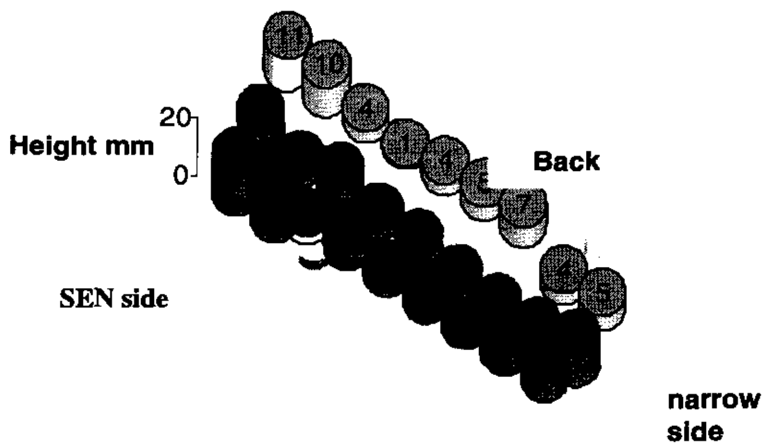


Figure 4 Mould flux pool depth across half of the mould length

It is not possible to measure mould flux behaviour in the mould directly. The mould heat flux is one of the best indicators of the mould powder performance. By adding the type of powder used to the input parameters, the differences between powders are taken into account.

2.1.6.3 Steel chemistry

The steel chemistry determines how the steel solidifies. Austenitic stainless steels are more prone to form depressions due to the ferrite-austenite transformation. Slow cooling is proposed to limit the contraction at the meniscus (Wolf,1997:256). This transformation does not occur during solidification of ferritic steels and more cooling - that is less crystalline powders - can be used.

2.1.6.4 Slab surface defects and its influence on heat transfer

In this section the various slab surface defects will be discussed. Although many of these fall outside the scope of this study, the possibility exists that these defects may be confused with metallurgical defects during inspection of the final product. The parameters used for the prediction of the metallurgical defects are also affected by these defects.

Longitudinal cracks are related to the heat flux at the meniscus and are most common in depression sensitive grades (Mills,1991:124). Decreasing the heat flux by using a more crystalline powder at the meniscus is one of the ways to prevent this defect. Uneven and excessive heat transfer can cause stresses (due to delta ferrite to austenite transformation) that contribute to longitudinal cracks (Emling & Dawson,1991:198). The irregularities in shell thickness cause strains in the longitudinal direction that are relieved by surface cracking. These defects can form if the liquid pool thickness is too small and the powder melting rate should therefore be sufficient to ensure homogeneous supply of mould flux around the mould perimeter (Ogibayashi *et al.*,1987:3).

Oscillation marks per se are not seen as defects, but it is believed that defects such as transverse cracks are associated with them. The trend is therefore to reduce oscillating mark dimensions by increasing the oscillating frequency or reducing the stroke (Mills,1991:100).

The properties of the mould flux also contribute to transverse corner cracking (Mills,1991:127). Transverse cracks are caused by slag entrapment in depressions or due to segregation during solidification. Transverse cracking is also attributed to insufficient powder consumption.

Transverse depressions are common on austenitic steel grades (Wang *et al.*,1999:449). Local overcooling at the meniscus in combination with the delta ferrite to austenite transformation is the cause of depression formation. It is important to maintain even shell growth on austenitic stainless steels for good quality product. Wolf (1997:256) proposed that a more crystalline powder be used to cast these grades.

Laps and bleeds are influenced by the interaction of mould level fluctuations, mould flux feeding and delta ferrite to austenite phase transformation in austenitic stainless steels (Wang *et al.*,1999:449). The volume contraction associated with this transformation can cause an air gap at the meniscus. High superheat will exacerbate the formation of air gaps. Metal level fluctuations cause a delay in shell growth and a non-uniform shell. The air gap in combination with the other issues causes tearing of the shell and bleeds and laps will form.

2.1.7 Flow in the mould

The design of the submerged entry nozzle (SEN) determines the flow in the mould. Most slab producers make use of a bifurcated nozzle design where the flow is diverted towards the narrow sides of the mould. This results in a double roll flow condition where some of the steel is diverted towards the top surface and the rest flows downwards. This flow condition is present in the Columbus Stainless mould as seen in water and mathematical modelling of the mould. The steel jet impact point on the narrow side and the extent of this upward flow influence mould level fluctuations. The SEN depth (that is depth of the top of the SEN port below the meniscus) also affects mould level fluctuations because it changes the steel jet point of impact.

Optimised flow conditions in the mould are essential to produce a good quality product. The fluid flow behaviour in the mould influences factors such as shell growth, superheat dissipation, inclusion flotation, mould flux entrapment, distribution of mould flux, meniscus freezing and surface turbulence. Low surface velocities and low downward velocities next to the submerged entry nozzle (SEN) are necessary to produce good quality steel. High surface temperatures and SEN jet impact not too low down on the narrow side of the mould are the preferred conditions.

Poor inclusion flotation and mould flux entrapment are two of the concerns associated with too high surface velocities in the mould. Two models are proposed in literature for mould flux entrapment. The one model is based on an energy balance on a slag droplet entrained in the steel (Jimbo *et al.*,1991:119). For entrainment the energy of the slag droplet must exceed the sum of the slag droplet buoyancy force and the interfacial tension between the steel and slag. Using this model the minimum velocity for entrainment in the mould is between 0.6 and 0.7 m/s.

Jimbo *et al.* (1991:120) proposed that entrainment is caused by the shear stress as the steel flow past the interfacial waves at the steel-flux interface. These interfacial waves are formed due to mould oscillation, mould level control, SEN design and turbulence in the mould. Another cause of interfacial waves is the differential velocities of two fluid layers. Since this velocity is lower (0.45 m/s) than the entrainment velocity, the conclusion is that the presence of surface waves on the mould surface does not necessarily indicate flux entrainment (Jimbo *et al.*,1991:120).

Kubota *et al.* (1990:359) found that the critical velocity for entrainment in their mould is 0.3 m/s (mould width 1250 mm; casting speed 2.1 m/min and SEN type -45°). This is the minimum surface steel velocity that results in the formation of vortices of size 0.8 to 2 mm and causes entrapment of similar size mould flux particles.

There are different views on the existence or not of vortices in the mould. Herbertson *et al.* (1991:142) state that vortexing will only occur when asymmetrical flows are present in the mould. Asymmetrical flows are caused by an off-centred SEN, clogging or erosion of the SEN and slide gate throttling. Gupta and Lahiri (1996:363) did cold model studies on a 640 x 80 mm mould to determine slag entrainment velocities. Vortex entrainment was shown to occur at much lower velocities than the entrainment of flux by surface velocity. They found that vortex entrainment is dependent on the mould width, port diameter, well depth of the nozzle, the properties of steel and slag (density and viscosity) and is independent of SEN depth.

McDavid and Thomas (1996:678) studied a 1400 mm by 230 mm mould to determine the influence of the steel flow on the mould flux layer on the steel surface. They found that as the steel flow rate increased, the mould flux is dragged from the narrow side of the mould towards the SEN. As flux is consumed on the narrow sides, a separation point is formed in the flux. The flux thickness is a minimum in this area.

Localised shell remelting occurs due to the stream impingement on the narrow side in the mould. The pressure of this impingement is a function of casting speed and SEN design parameters such as port divergence angle, bore size, port length and well depth (Herbertson *et al.*,1991:138). Herbertson *et al.* found that the change in the position of impact on the steel shell varies directly with a change in SEN depth. This factor is of great importance because a too low submergence depth can cause insufficient shell solidification when the steel leaves the mould. This may lead to a steel breakout.

As seen in the previous sections, direct measurements of for instance steel velocities at the meniscus and impingement point are difficult to measure. The steel flow rate and superheat will however provide good indications.

2.1.8 Interaction between mould parameters at the meniscus

The meniscus is not only the most important area in the caster, but also the most difficult to understand. There are three major contributors to heat transfer at the meniscus including the mould flux behaviour (slag rim and mould powder consumption), steel solidification and the mould level. These factors have been discussed in previous sections, but it is also important to analyse the combined effect on heat transfer.

The mould level is controlled with an eddy current sensor at a level of 90 mm below the top of the mould. The mould level provides the setting for the stopper rod that controls the flow of steel from the tundish. The aim of the mould level controller is to control the mould level within ± 5 mm. SEN clogging and bulging make the mould level more difficult to control. SEN clogging may lead to biased flow, non-uniform heat removal and thermal stresses in the shell and uneven shell growth (Emling & Dawson, 1991:198). SEN clogging is a good indicator of steel cleanliness.

The mould level fluctuation found close to the narrow sides of the mould is greater than the fluctuation at a quarter width position at a casting speed of 2.1 m/min. The mould level controller is situated close to the quarter width position. A correlation was found between defect levels and the variation of the long period wave of the off-corner level fluctuation. This correlation showed that when the fluctuation is in this area of the mould was controlled between certain limits, defect occurrence was minimised.

The mould level affects the feeding of mould powder into the steel-copper gap. If a standing wave (higher mould level towards the narrow sides of the mould) exists, this may impair the feeding of mould flux in these regions of the mould. High mould level fluctuations may also temporarily prevent lubrication. This lack of lubrication may result in stickers and other defects on the steel shell. The mechanism for the formation of various defects is discussed in the next section.

2.1.9 Summary of variables

The steelmaking process was discussed and several parameters that influence steel quality were identified. A list of these parameters is given in Table 2. In the steelmaking process the control of the steel oxygen level is important. More defects are also associated with the reblow of the steel in the converter.

At the rinse station care should be taken to prevent steel / air contact, to limit slag entrainment and to facilitate inclusion removal. The steel temperature control is important. The time between end of rinse and start of cast will influence the amount of thermal stratification and therefore temperature control in the tundish. Ladle refractory heat content also affects stratification, but the heat content is not known at the time of casting.

The residence time of the steel in the tundish determines the time available for inclusions to float out. The combination of the casting speed and tundish mass provides an indication of this residence time. Steady state casting conditions are most beneficial for quality. Poor quality is related to start-up conditions since air ingress and slag entrainment are associated with these conditions. Ladle transition (ladle change during a cast) can also contribute to poor quality (air ingress and lower tundish levels).

Steelmaking parameters	**Continuous casting parameters**	**Continuous casting parameters (continued)**	**Other parameters**
CLU final Silicon	Tundish mass	Mould level fluctuation	Steel chemistry
Reblow / Return to CLU	Stopper rod position	Mould level fluctuation close to mould corner	Solubility products
Opened steel eye due to stirring	Tundish temperature Control	Oscillation frequency	
Rinse station stirring rate	Air ingress	SEN depth	
Rinse after final additions	Heat in cast sequence	Mould heat flux	
Metal addition mass	Start-up conditions	Mould water flow	
Ladle refractory heat content	Steel superheat		
Rinse station temperature	Casting speed		
Time between rinse end and start of cast	Casting speed change		
Ladle transition	Mould powder used		

Table 2 Steel plant parameters that affect quality

Mould heat transfer was shown to give a good indication of steel solidification. The mould water flows and the mould powder used control the heat transfer. Instabilities in the casting speed and mould level are also deleterious to quality and should be avoided.

A few additional parameters should be considered. The steel chemistry is important. Related to the chemistry is the amount of oxides (and therefore inclusions) present in the steel. The solubility products provide a quantitative measure.

The critical variables were determined and it is now important to consider techniques to predict steel quality. In the next section the use of neural networks in the process industry and the data mining process is discussed.

2.2 Application of neural networks in the process industry

Neural networks are based on the functions of the human brain and are attempts to replicate them in a technical environment. It can be described as numerous computational structures of simple process units connected on a parallel scale (Aldrich,1997:6). Input is received from the input layer (see Figure 5) and transmitted through hidden neurons to the outputs. Non-linear transfer functions are used in the hidden neurons and this give neural networks the ability to model non-linear functions.

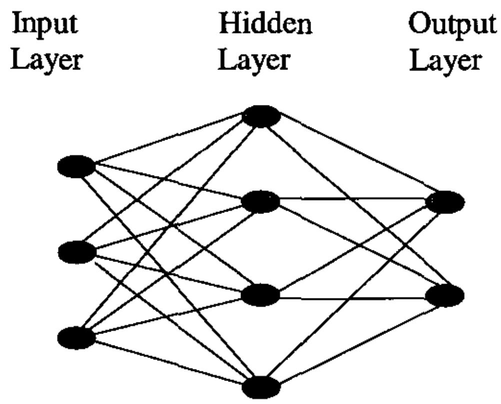


Figure 5 Structure of a neural network

Neural networks can be used to solve problems that are difficult to model analytically. This includes problems that are of a non-linear nature. Several successful applications are found in practice. The multiplayer perceptron neural network trained using error backpropagation is the most frequently used architecture in process engineering (Aldrich,1997:13).

Babu and Hanratty (1993: 1951) used multilayer perceptrons with backpropagation of the error to model a chemical batch process. Secondary measurements were used to monitor and control the process since the quality was only available after a batch was finished. These intermediate measurements provided an indirect way to assess the effect of certain disturbances on the process. In this study the neural network predictions were compared to quadratic regression models and it was concluded that the neural network was superior in terms of both accuracy and tolerance for noise (Babu & Hanratty,1993:1960).

Other applications include the modelling of a grinding plant and blast furnaces (Flament *et al.*,1992:235 and Zuo & Björkman,2001:115 respectively). A neural network was used as the controller for the cyclone overflow fineness as part of the grinding plant simulation. They also used a neural network to determine the inverse of the process dynamics (Flament *et al.*,1992:242). In the blast furnace the occurrence of channelling was detected using neural networks. The factors affecting channelling was known, but the thresholds for the control of these factors could not be determined using analytical models. Indirect measurements were used as input to a neural network to predict the onset of the channelling.

Neural networks have also been successfully applied in the steelmaking industry. Behzadipour *et al.* (2002:19) used neural networks to estimate the power during steel rolling. Some parameters were difficult to determine accurately and indirect measures were included in the model to account for them. Output from empirical models were also included as input to the neural network (Behzadipour *et al.*,2002:20). The model predictions were compared with data collected from various rolling mills and it was found that the model predicted the power within 8% of the actual value in 95% of the cases (Behzadipour *et al.*,2002:26).

Neural networks were also successfully applied to the prediction of stickers in the continuous caster mould. Many factors affected the heat transfer as measured with thermocouples that were embedded in the copper moulds and many false signals were generated (due to noisy temperature data). Neural networks reduced the number of false signals significantly (Bellomo *et al.*,1995:345).

In another example where a neural network was applied to predict submerged entry nozzle (SEN) clogging (Saarelainen *et al.*,1999:89). The SEN is a refractory tube used between the tundish and mould to transfer steel into the mould. Clogging occurs when the steel temperature is low and when the steel contains high levels of oxides and nitrides. An expert in their steel plant was used for selection of variables to predict clogging. Their initial set of 58 variables was reduced to 8 in the final model (Saarelainen *et al.*,1999:92). A test data set was used to determine the reliability of the model and the model's prediction accuracy was 86%.

From these examples it can be seen that neural networks have been applied successfully in the process industry. Where direct measurements were not available, indirect measures were used to develop models of acceptable accuracy. Supervised learning was used to develop these models and models were evaluated by using test data sets.

In the next section the data mining process will be discussed. This process was applied to develop the quality prediction model.

2.3 Data mining

Data mining is an iterative process used to discover trends in data (Westphal & Blaxton,1998:6). With the development of computer networks an abundance of process data is available. These data are only useful to the business if meaningful information can be extracted. Although the focus of this study is on the use of neural networks, the data mining methodology is used.

The data mining process includes data selection, data preparation, data transformation and finally information extraction (Cabena *et al.*,1998:42). Berry and Linoff (1997:22) describe the process in broader terms and state that the problem should first be identified, followed by analysis, then implementation and finally evaluation of the outcome. The most time-consuming step is the data preparation. Raw data is seldom in the right format for the analysis and often more than one data source with the same data exist. The best data sources have to be determined.

The quality of the data will determine the accuracy of the model and is therefore critical. Visualisation and statistical techniques can be used to assess the quality of the data. Charts can be used to determine both outliers and missing values. Missing values can either be dropped or substituted with an acceptable estimate. This decision is influenced by the availability of data and the sensitivity of the output to the parameters with missing values.

In the information extraction step algorithms are applied to the data to produce a model. The model predictions are then analysed to determine the accuracy. Techniques to measure the accuracy will be discussed later.

Over-training is one of the common problems. A model is over-trained if too much of the noise in the training data set is modelled and the model does not generalise when confronted with new observations. Over-training can be prevented by using a validation data set during the training process (Berry & Linoff,1997:79). An additional test data set is also used to assess generalisation after the model is trained. A good model will perform in a similar fashion on the training, validation and test data sets.

2.3.1 Data mining techniques

Supervised and unsupervised modelling techniques can be used to model the process. In unsupervised techniques there is no output variable to use during the training of the model. Clustering is an example of an unsupervised learning algorithm and can be used to determine groups of variables that act together. Once clusters are identified other methods are used to determine the meaning of the clusters.

Decision trees and neural networks are both supervised learning methods. Their advantage over statistical techniques is that they can be applied to determine local surfaces whereas statistical techniques determine conditions over the whole population (Berry & Linoff, 1997:115).

A decision tree builds a model in the form of a tree with branches. As a record moves from the root down to the branches, if-then questions are repeatedly asked. Categorial variables are split between the groups of values while continuous variables are split at a derived threshold. Current decision tree algorithms are not optimal since the decision on where to split is never revised once the decision has been made (Cabena *et al.*,1998:69). There is no backtracking (that is rules are not changed once established) as found in neural networks. The handling of missing values can have serious effects on the outcome of a decision tree. However, nodes can be created to accommodate the missing values. The decision trees also suffer from fragmentation. When the leaves have very few points learning is difficult. This problem is limited by pruning the tree. The effect on over-fitting can also be prevented by pruning.

The major advantage of decision trees over neural networks is that it provides rules (Berry & Linoff, 1997:243). These rules can be more important than the model itself since they provide guidelines through which the process can be improved. If only an accurate prediction is required, neural networks may provide better results.

In neural networks the nodes are linked by weights. Hidden layers can be included between inputs and outputs. Backpropagation (a supervised learning technique) is the most widely used training technique for multiplayer perceptron models (Cabena *et al.*,1998:73). The name refers to the way in which the errors are propagated back from the output layer to the input layer during training. This method is very versatile and works for many problems.

According to Berry and Linoff (1997:309) missing values don't cause serious problems in neural networks since each data point is weighted during the training process. If the number of missing values is significant it can have a serious effect. In this case it would be best to impute the missing values. Over-fitting is also common and convergence problems may occur if data contain impure records or if the problem is too complex. Neural networks are black box models and the method of decision-making is therefore uncertain. This can be overcome by doing a sensitivity analysis.

Regression models are also supervised learning models. The disadvantage of regression is that only linear dependencies between parameters can be accounted for (in this study only linear models were considered). Determining how the model made a decision is, however, simple.

2.3.2 Measuring model effectiveness

Determining the accuracy and the benefit from the model are important steps. Two methods can be used to determine the effectiveness of the model. The first is a confusion matrix (Cabena *et al.*,1998:75-76). An example of a confusion matrix is shown in Table 3 and shows the number of records that were correctly predicted (either Yes or No) and those that were predicted wrongly. The coverage of the model is the number of "Yes" values that were predicted correctly (400/600 = 66.7%) while the accuracy is given by the number of "Yes" and "No" values that were correctly predicted (8400/8700 = 96%).

Actual	Predicted Yes	Predicted No	Total
Yes	400	200	600
No	100	8000	8100
Total	500	8200	8700

Table 3 Confusion matrix

Another method that is used to evaluate models is lift charts. This method is useful to compare the results from several models (Berry & Linoff, 1997:107-109). The difference in the number predicted correctly in a sample versus that of the general population is called the lift:

$$\text{Lift} = \text{Probability}\ (\text{Class}(i) \mid \text{sample})\ /\ \text{Probability}\ (\text{Class}(i) \mid \text{population})$$

Class(i) = class variable to be predicted (for example the class "Yes" in Table 3).

The lift chart is constructed from a scored data set by sorting the predicted probabilities in descending order (one to zero for a binary target). The data are then grouped into deciles and plotted cumulatively. Assume for example a data set consists of 10% records with defects and the rest with no defects. This data are scored with the model that was developed and the data are sorted in descending order based on the predicted defect probabilities. If 10% of the data with the highest predicted defect probability is selected, the percentage of defects in this group should be more than 10% (probability of a defect in the population) for a good model. The higher this percentage, the better the model.

Although the lift chart provides a good indicator for comparison of models it does not answer the question whether the modelling process was worth the effort. The impact of the model has to be measured. A return on investment is required and is determined by comparing the cost of model development with the savings/income derived from the application of the model.

2.4 Summary

The steel plant operation was discussed in detail with special emphasis on factors that affect the steel slab quality. Variables to describe the steel quality were identified and will be used in the model development.

Several approaches can be used to model the quality. Since the problem under consideration is a complex one, most likely non-linear, and because analytical modelling is highly complex, it was decided to use a neural network. Neural networks have been applied successfully in many industries including the steel industry. In the majority of cases the multiplayer perceptron with backpropagation of the error is used.

In the last section of this chapter the data mining process was discussed. This is a process used to uncover trends in data to aid decision-making. This process of identifying variables, data filtering, pre-analysis of data, modelling and testing was applied in this study. Special emphasis was placed on the testing of the model to ensure good generalisation. Various modelling approaches were discussed and some of these will be used for comparison with neural networks.

The benefits and drawbacks of the different modelling techniques were discussed and are summarized in Table 4. Although linear regression models are simple, only linear dependencies are taken into account. The non-linear models should provide more accurate results but over-fitting is a concern. Over-fitting can be assessed using a test data set after training. The major benefit of decision trees is that the rules used for decisions are known.

	Linear regression	Decision tree	Neural network
Linear / Non-linear	Linear	Non-linear	Non-linear
Missing values	Negative	No problem (add nodes for missing values)	Negative (delete records or impute missing values)
Over-fitting	Not if linear	Possible	Possible
Interpretation of decisions	Simple	Simple depending on number of splits	Difficult

Table 4 Comparison of modelling techniques

In the next chapter the neural networks and various training algorithms are discussed in more detail.

CHAPTER 3 THEORETICAL BACKGROUND

The multilayer perceptron with backpropagation of the error is one of the most popular neural networks and training techniques used in industry. This network and some alternative algorithms to update the weights are discussed.

3.1 Backpropagation

Backpropagation is a generalization of the least mean squares algorithm. These networks learn by repeatedly attempting to match predicted and target values. The weight matrix of the network nodes is repeatedly adjusted to minimize the mean square error between the calculated and actual outputs (Aldrich,1997:13).

A neural network usually consists of an input layer, one or more hidden layers and an output layer. The transfer functions used in the majority of feed forward neural nets are sigmoidal functions (Aldrich,1997:14). The sigmoidal function can have the form $f(x)=1/(1+b^{-bx})$ that produces an output between zero and one or it can be a S-shaped function. The hyperbolic function $(f(x)=(1-e^{-x})*(1+e^{-x}))$ is an example of and S-shaped function and produces an output between –1 and 1. It is these transfer functions that give neural networks their non-linear properties. A linear transfer function is usually used in the output layer.

In the first step the input is propagated forward through the network (Hagan *et al.*,1996:11-7):

$$\mathbf{a}^{m+1} = \mathbf{f}^{m+1}(\mathbf{W}^{m+1}.\mathbf{a}^{m} + \mathbf{b}^{m+1}) \quad \text{for } m = 1,\ldots,M-1 \tag{1}$$

where $\mathbf{a}^{m+1}$ is the output of the nodes,

$\mathbf{f}^{m+1}$ refers to the transfer function (linear, sigmoidal, etc.),

$\mathbf{W}^{m+1}$ is the input weights to the node,

$\mathbf{a}^{m}$ is the input from the previous layer,

$\mathbf{b}$ is the bias and

M is the number of layers in the network.

The output of one layer therefore becomes the input to the next layer. $\mathbf{a}^M$ is the output of the neural network.

The next step is to propagate the errors backward through the network. The error for the nodes in the output layer can be calculated as follows:

$$\mathbf{s}^M = \mathbf{F}^M(\mathbf{n}^M)(\mathbf{t} - \mathbf{a}) \qquad (2)$$

where $\mathbf{t}$ is the expected output, $\mathbf{a}$ the predicted output and $\mathbf{s}^M$ the errors calculated for the output layer. $\mathbf{F}(\mathbf{n}^M)$ is the derivative of the node output to the input weights.

This output layer error has to be propagated backward through the network so that the weights in each layer can be adjusted appropriately:

$$\mathbf{s}^m = \mathbf{F}^m(\mathbf{n}^m)(\mathbf{W}^{m+1})^T\mathbf{s}^{m+1} \qquad \text{for } m = M\text{-}1, \ldots, 2, 1 \qquad (3)$$

where $\mathbf{s}^m$ is the errors calculated for this layer and is a function of $\mathbf{s}^{m+1}$,
the errors of the next layer,
$(\mathbf{W}^{m+1})^T$ is the transpose of the output weight vector and
$\mathbf{F}^m(\mathbf{n}^m)$ is the similar to that used in equation (2).

All that remains is to update the weights and biases using the approximate steepest descent method:

$$\mathbf{W}^m(k+1) = \mathbf{W}^m(k) - \alpha.\mathbf{s}^m(\mathbf{a}^{m-1})^T \qquad (4)$$

$$\mathbf{b}^m(k+1) = \mathbf{b}^m(k+1) - \alpha.\mathbf{s}^m \qquad (5)$$

where k is the iteration number,
$\mathbf{W}^m$ and $\mathbf{b}^m$ are the weight and bias vectors and
α is the learning rate.

The learning rate (value between zero and one) determines the size of the change to the weights and biases.

3.1.1 Limitations of backpropagation method

The steepest descent backpropagation is a generalisation of the least mean squares method where the mean squared error is minimised (Hagan *et al.*,1996:12-3). This algorithm always converges for a single layer network, but convergence is not guaranteed for multilayer networks and convergence may be slow. This is because the surface for a multilayer network can have several local minima and the curvature can vary widely in different regions of the solution space. The difficulty is to find the global minimum.

The error surface for multilayer networks varies over the parameter space and it is difficult to choose a learning rate for the steepest descent algorithm (Hagan *et al.*,1996:12-5). The surface may be flat in areas and have high curvature in other areas. On the flat surfaces a large learning rate is feasible while a small learning rate is required for high curvature surfaces. The learning rate can also lead to oscillatory behaviour if the rate chosen is close to one.

Several algorithms were developed to counter the limitations of the standard backpropagation algorithm. These algorithms include Momentum, Variable Learning Rate and two numerical optimization techniques namely conjugate gradient and Levenberg-Marquardt.

3.1.2 Momentum

Convergence can be improved by smoothing the oscillations between iterations. For this purpose a momentum coefficient (γ) is used to filter the severity of oscillations between iterations (Hagan *et al.*,1996:12-10). The weights and biases are adjusted as follows:

$$\mathbf{W}^m(k+1) = \gamma\ \mathbf{W}^m(k) - (1-\gamma)\alpha\ \mathbf{s}^m(\mathbf{a}^{m-1})^T \qquad (6)$$

$$\mathbf{b}(k+1) = \gamma\ \mathbf{b}(k) - (1-\gamma)\alpha\ \mathbf{s}^m \qquad (7)$$

3.1.3 Variable Learning Rate

As mentioned previously, for convergence it will be beneficial to increase the learning rate on flat surfaces and to decrease it where the slope increases (Hagan *et al.*,1996:12-12). This section deals with the method used to determine when to change the learning rate and by how much. The rules for the variable learning rate backpropagation algorithm are (Hagan *et al.*,1996:12-12):

1. If the squared error over the entire model increases by more than a certain percentage (1-5%), discard the new weights, multiply the learning rate by a factor between zero and one and set the momentum coefficient to zero. This procedure will limit oscillatory behaviour on high curvature surfaces.
2. If the error increase by a value less than the percentage in (1) (that is 1-5%), accept the weight update and keep the learning rate unchanged. If the momentum was previously set to zero it should be restored to the original value.
3. If the squared error after an iteration decreases, accept the weights and multiply the learning rate by a factor greater than one. This will increase the learning in flat areas of the error curve. If the momentum is zero, restore it to its previous value.

3.1.4 Conjugate gradient

The steps in the conjugate gradient algorithm can be summarized as follows: first the search direction is determined as the negative of the gradient. The weights and biases are then updated to minimize the error using a specific learning rate. The next search direction is chosen to decrease the gradient further. These steps are repeated until a converged solution is found.

Two problems have to be addressed to implement this algorithm in neural networks (Hagan *et al.*,1996:12-15). The first is to find a method to determine the minimum of a function in a specific direction. The function is evaluated at two points and then evaluated at new points, successively doubling the distance between the points. This is repeated until the value between two consecutive evaluations increases. A process of interval reduction is then used to move closer to the minimum.

3.1.5 Levenberg-Marquardt

The Levenberg-Marquardt algorithm is a variation of Newton's method. (Hagan *et al.*,1996:12-19). The algorithm can be summarized by the following equation:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - [\mathbf{J}^T(\mathbf{x}_k)\mathbf{J}(\mathbf{x}_k) + \mu_k \mathbf{I}]^{-1} \mathbf{J}^T(\mathbf{x}_k)\mathbf{v}(\mathbf{x}_k) \tag{8}$$

where **J** is the Jacobian matrix,

$\mathbf{x}_k$ is the parameter vector that contains the weights and biases,

I is an identity matrix,

v is the vector of the calculated errors and

μ_k is a constant that controls the learning rate.

As μ_k is increased it approaches the steepest descent algorithm with a small learning rate and as μ_k is decreased to zero the algorithm becomes the Gauss-Newton algorithm. μ_k is initially set to a small value and if the error does not decrease, μ_k is multiplied by a number greater than one (say η) in the next step. Smaller steps are taken and eventually the error should decrease. If the next step does produce a smaller error, μ_k is divided by the factor η and should provide faster convergence.

3.2 Summary

In this chapter the backpropagation algorithm was discussed. The convergence problem was highlighted. There is always the risk that the algorithm will get stuck in a local minimum and not converge to the global minimum. The methods that were proposed to counter the limitations of standard backpropagation were the addition of a momentum term and the use of a variable learning rate. The convergence can also be improved by using numerical optimization techniques such as conjugate gradient and Levenberg-Marquardt to determine the optimum direction to search for the global minimum.

CHAPTER 4 MODEL DEVELOPMENT

In this chapter the implementation of models in SAS Enterprise Miner is discussed. Enterprise Miner provides a platform for exploratory data analysis as well as model development. The implementation of neural networks, regression, decision tree models as well as combining these model results are considered.

4.1 Introduction

The SEMMA approach is used in SAS Enterprise Miner to uncover information from data (Wielenga *et al.*,1999:5). SEMMA is an abbreviation for sample, explore, modify, model and assess. Samples of data are used for data exploration using various statistical and graphical tools. Before developing the models, the data are filtered to remove unwanted and inaccurate data. Once the data are in an acceptable format various models are constructed. These models are compared using tools such as lift charts and confusion matrices.

In the training of all the models the data set is divided into 65% training data, 25% validation data and 10% test data. The validation data set is used to prevent over-training while the test data set is used for final testing of the models.

4.2 Implementation of neural networks in SAS Enterprise Miner

The input variables are automatically separated into input groups for interval, nominal and ordinal variables. Interval variables are standardised by subtracting the minimum and dividing by the variable range. This ensures that results are not negatively influenced by the differences in variable order of magnitude. Steel temperature values are for instance 10000 times that of converter silicon and can lead to the suppression of the effect of the silicon if variables are not standardised.

Multilayer perceptron networks is the most commonly used approach in literature and are also the least sensitive to the dimension of the problem when compared to radial based function networks (Aldrich,1997:13). It was therefore decided to use this type of network.

One or more hidden layers can be specified with one or more neurons per layer. It is difficult to determine the optimum number of neurons per hidden layer beforehand. Numerous networks with different numbers of hidden units have to be trained to determine the optimum number of neurons. In this study one hidden layer with five, ten, fifteen and twenty neurons respectively was used.

A variety of training algorithms are available for training the network. Some of these are compared in Table 5. The three discussed in detail in the previous chapter were considered for use in this study. The conjugate gradient method was selected because this method was also available for linear regression and made comparison easier.

These training algorithms are not guaranteed to find the global minimum (Potts, 2000:128). Changing the starting point usually gives a different solution. Preliminary training can be done to find the best starting point. A number of random starting points are used to find the best final value after a small number of iterations. This value is then used for the training of the network.

In this study early stopping was used to find the best network. The weights corresponding to the minimum error for the validation data set were selected and used for scoring the network. For a fast training method like Levenberg-Marquardt over-training is possible on the first iteration. It is therefore better to use one of the slower training techniques such as Backpropagation, Quickprop and Rprop (Anon., 2001:47). Preliminary training is not available when using early stopping.

Before training, a criterion for choosing the best solution must be specified. An error function that measures the difference between the calculated and actual values of the output is specified. This error function is minimized during training. In this study the ordinary least squares method was used to find the best solution.

Training algorithm	Description	Characteristics
Standard backpropagation	Learning rate and momentum to be specified.	Slow. Tune learning rate manually.
Quickprop	Newton-like method	Faster than standard backpropagation. More reliable.
Rprop	Different learning rate for each weight and rates are variable.	More stable compared to Standard Backprop and Quickprop.
Levenberg-Marquardt	Variation of Newton's method.	Very fast for small networks, but memory intensive.
Quasi-Newton	Approximates Hessian matrix with matrix of first derivatives.	Good for medium sized problems. Needs more iterations than Levenberg-Marquardt.
Conjugate gradient	Negative gradient search.	Very little storage and therefore better for large networks. More iterations required compared to Quasi-Newton techniques and Levenberg-Marquardt.

Table 5 Comparison of training algorithms

4.3 Regression model in SAS Enterprise Miner

A logistic regression model is used to predict the probability of a binary target having a specific value. Logistic regression is similar to a multilayer perceptron with no hidden layers and a logistic output activation function (Wielanga *et al.*, 1999,102). The conjugate gradient method is specified as the optimisation method, that is the same method that is used to determine the neural network weights. The use of a slower optimisation method is also a prerequisite for using early stopping during training. The regression model is kept simple and no interaction between parameters is considered. All the variables are included in the final model.

4.4 Decision trees in SAS Enterprise Miner

Decision trees can be used to model complex non-linear surfaces and have the advantage that results are easy to interpret. This is the major benefit of decision trees compared to neural networks.

Both binary and multi-way splits are considered. Binary and multi-way splits were compared during preliminary runs where it was found that a five way split gave a smaller mean squared error. The multi-split model was however difficult to interpret. Since the goal of the study is to determine whether neural networks perform better than other modelling techniques - and not how the model made decisions - it was decided to use five-way splits.

If the multi-way split is used, an additional consolidation phase is necessary to group the levels of inputs. A set of candidate splits is determined and a splitting criterion is used to determine the best split. When the split gives pure children no further splitting is done.

The Chi-squared test is used to judge the worth of a split. This method tests whether the class proportions are the same for each child node. It measures the difference between the observed node counts and what would be expected if the branches and target classes were independent (Potts, 2001:50).

A two-step process is followed where the best split of each input variable is selected and then the best of all the input variables. The problem is that the Chi-square statistic favours the selection of the split with the most branches (Potts, 2001:47). The P-value (Chi-square statistic) is therefore adjusted by taking the degrees of freedom into account. In the same way adjustment is required when the splits of the different variables are compared. If no adjustment is made, the variables with the larger number of splits are favoured.

A similar problem is found as the depth of the tree increases. The P-value for a split several levels below the root node is lower than that of a branch higher up in the tree. The split lower down would therefore always be favoured if no compensation is made. To adjust for this, a depth multiplier is used. This depth adjustment limits tree growth by decreasing the required P-value for a split low down in the tree. For a branch six levels below the root to split, the P-value has to be 64 times smaller than that of the root node for the split to occur.

4.5 Committee of models

The results from the different models are combined by averaging the posterior probabilities from the individual models. The results from the various neural networks, the decision tree and the regression model are combined. The combined model will however only give more accurate results if the individual models disagree with one another. The results from all the models are therefore compared to determine whether the committee of models performed better.

4.6 Model assessment

The assessment node provides a tool for the comparison of various models. A lift chart is used to compare alternatives. The lift chart is constructed as discussed in Chapter 2.

As part of the assessment, additional data sets can be scored (using the scoring node) for further validation of the model. The trained model was used to score a data set that was not used during training. The aim was to determine whether the model was able to generalize.

4.7 Summary

The application of three types of models in SAS Enterprise Miner was discussed. These included neural networks, regression models and decision trees. The facility for combining the models was also discussed.

It was decided to train both the neural networks and the regression model using the conjugate gradient optimisation method. A decision tree with a five-way split was selected for comparison and all the model results were combined into a committee of models to try and improve generalisation further.

The most time-consuming step in model development is the sourcing and filtering of noisy plant data. This process and some preliminary statistical analysis will be discussed in the next chapter.

CHAPTER 5 ANALYSIS OF STEEL PLANT DATA

The accuracy of the data is important to ensure a good model. This chapter deals with the filtering of the data. This is followed by a preliminary data analysis that was done to remove unnecessary variables and to visualise the effects of the variables on defect occurrence.

5.1 Data preparation

The screening of data is an important step in the modelling process. Records with missing values were deleted and where parameters exceeded certain limits, records were also discarded. Plant experts supplied the limits for variables. The SAS program used for data filtering is included in Appendix A.

Only material with a final gauge smaller than 3.1 mm is used in the analysis. The detection of metallurgical defects during inspection on thicker gauges is not as effective. All the metallurgical quality information (that was severity and frequency of defects) was included in the data file. Although this information can be used to filter the lighter defects, it was not used in this study.

Steel types are grouped based on their composition and properties. As discussed before, titanium stabilised steels are more prone to defects and are excluded from this study. Steel types that are included are austenitics 304, 309, 310, 316 and ferritics 412 and 430. The quality of titanium stabilised steels is worse than that of other steels and as a precaution these steels are always ground.

Slabs are grouped by width since the width distribution is not continuous. Three major width groups are found, namely 1100 mm, 1300 mm and 1500 mm.

The output from empirical models was also included as inputs. The solubility product model was discussed in detail in Chapter 2. The predicted levels of alumina, titania and titanium nitrides are used as input variables to the models.

Two parameters included in the analysis were the Ferrite Factor and Chrome-Nickel Equivalence. These are empirical equations from literature that are used to predict hot ductility under hot rolling conditions (see Suutala, 1983:53 and Norstrom & Malm, 1979:31). These parameters do not relate to metallurgical defects, but have to be included because hot ductility defects are at times wrongly booked as inclusions or line inclusions. Laboratory work and data analysis showed that this defect is related to steel composition and the cooling rate in the mould. Since inclusions are seldom reported on steel with a gauge less than 3 mm and the fact that an increase in inclusion levels is associated with the hot ductility defect (as evidenced by laboratory studies) this defect was excluded from the models.

Since the percentage of defect occurrence in the original database is low, it is important to balance the data sets to improve prediction accuracy. The original database included only 6% slabs with line inclusions and 16% with skin laminations. Copies of the defect data were added to make the defect/no-defect ratio more balanced. The Prior Vector in the Target Profile was set to the original percentages of defects in the database. This was necessary to compensate for the over sampling of the defect data. Separate balanced data sets were created for the two metallurgical defects (skin laminations and line inclusions) respectively. The number of defects per data set was 50% and 41% for the skin-lamination and line-inclusion defects respectively. The line-inclusion defect data set had 5517 records and the skin-lamination data set 5957 records. The number of records in the training, validation and test data sets is shown in Table 6.

	Skin-lamination models	Line-inclusion models
Training data	3872	3586
Validation data	1489	1379
Test data	596	552

Table 6 Records in each of the data sets

5.2 Correlation

A correlation matrix was used to identify redundant variables. Table 7 shows the variables (only part of the cross-correlation matrix) that were highly correlated to other input variables. Variables with a high correlation (0.7 or –0.7 for inverse correlation) were removed from the data set (highlighted variables). Most of the influence of these variables is included in the other variables.

Variable	Correlated variables	Correlation value
Width	Wide/Narrow	0.8
	DeltaT3	0.59
	DeltaT4	0.64
	Oscillation	-0.83
Level	Level difference	0.96
	Speed Difference	0.7
Speed	DeltaT1	0.57
	DeltaT2	0.52
	Level	-0.74
	Level Difference	-0.77
	Speed Difference	-0.79
	Wide /Narrow	-0.56
Oscillation	Wide/Narrow	-0.69
Level difference	Speed difference	0.74
V1 Silicon	LASi mass	-0.7
B211	Metal mass	0.51
Total argon	Bubbling time	0.55
Water Loose Wide	Water Fixed Wide	1
	Water Left Narrow	0.94
	Water Right Narrow	0.96
	Delta T1	-0.69
	Delta T2	-0.72
	Delta T3	-0.53
	Delta T4	-0.46
Water Right Narrow	Water Fixed Wide	0.96
	Water Loose Wide	0.96
	Water Left Narrow	0.99
	Delta T1	-0.72
	Delta T2	-0.75
	Delta T3	-0.52
	Delta T4	-0.5

Table 7 Correlation values for selected parameters

Variable	Correlated variables	Correlation value
Delta T1	Delta T2	0.97
Delta T3	Delta T4	0.97
Al2O3	Ti3O5	0.51
	Oxygen	0.6
TiN	Titanium	0.8
	Aluminium	0.6
	Nitrogen	-0.56
Cr-Ni equivalence	Ferrite factor	0.67
	Rinse temperature	0.89
	C	-0.61
	Mn	-0.88
	V	-0.54
	Co	-0.66
	Cr	-0.82
	Ni	-0.86
	N	-0.56
Ferrite factor	Mn	-0.7
	Co	-0.54
	Ni	-0.72
	Nb	0.59
	B	0.67
	Rinse temperature	0.67
	Cr-Ni equivalence	0.67

Table 7 · Correlation values for selected parameters (continued)

The Wide / Narrow parameter refers to the ratio of the heat removed from the wide sides relative to the narrow sides. This parameter was highly correlated to the cast width, casting speed and oscillation frequency. A strong correlation was also seen between the mould level, mould level fluctuation and speed variation. A speed change is in general accompanied by mould level instability and is confirmed by the high correlation between these parameters.

As was expected, there was a high correlation between the four mould water flow rates. The heat removed on the two wide and narrow sides were also correlated (Delta T3 and Delta T4 on the wide sides and Delta T1 and Delta T2 on the narrow sides).

The majority of the chemistry effects are included in the Chrome-Nickel equivalence and the Ferrite factor (calculations discussed in the previous section). There was therefore a high correlation between these parameters and the steel chemistry. Although the correlation between the Chrome-Nickel Equivalence and the Ferrite factor was high, it was decided to keep both to ensure that all the chemical elements were included. The silicon addition (LASi) at the Rinse Station was inversely proportional to the silicon content in the converter (V1Si) and the silicon additions variable was therefore discarded.

Three other variables were also excluded from the data set since they were univariate. These variables were poke open, NCR-lumps and boiling in the mould.

The final set of variables that were used in the models is shown in Table 8. A total of thirty-nine variables were left. The average values of these variables per slab were used to predict the surface quality of the steel.

5.3 Data visualisation

In the literature study the relationship between the metallurgical defects and operating variables was discussed. Data visualisation methods are used to assess the presence of these relationships in the data. The SAS Multi-plot facility was used to draw histograms of the incidence of defects for each variable. Data were separated into bins for each independent variable and the frequency of the defect and no defect records for each bin was shown. When some of the bins contain a higher percentage of defects this indicates operating ranges to avoid. A few examples are included in Appendix B (see Figures 1 to 15).

Figure 1 of Appendix B shows the histogram of the frequency of skin laminations and no skin laminations for each bin of the heat removed from the mould narrow sides. For the slabs where the mould narrow side heat removal was below 4 °C there was a high incidence of skin laminations. Other variables where the dependence was evident included the heat in the cast (Figures 3 and 4, Appendix B), mould level fluctuation (Figure 5, Appendix B), steel titanium nitride content (Figure 9, Appendix B) and total argon consumed at the rinse station (Figure 10, Appendix B). The incidence of both skin laminations and line inclusions were higher for the first heat compared to the other heats. More skin laminations were also associated with high mould level fluctuations (> 0.035) and high titanium nitride levels (>17.5%). All the line-inclusion defects were removed when the total argon used at the rinse station exceeded 18 m^3.

Steelmaking parameters	**Continuous casting parameters**	**Continuous casting parameters (continued)**	**Other parameters**
CLU final Silicon (V1Si)	Tundish mass	Mould level fluctuation	Steel chemistry: • Phosphorus • Sulphur
Reblow	Stopper rod movement	Mould oscillation frequency	Solubility products: • Al_2O_3 • Ti_3O_5 • TiN
Return to CLU	Ladle transition	Mould water temperature	Chrome-Nickel equivalence
Opened steel eye during stirring	Steel superheat	Mould water flow: • Wide side (waterFW) • Narrow side (waterRN)	Ferrite factor
Rinse station stir time	Unshrouded	Mould powder used	Final gauge
Rinse station stirring rate (Nm^3/min/°C)	Burn open	Mould narrow side heat removed (DeltaT1)	Steel type
Total argon consumed	Heat in cast	Mould wide side heat removed (DeltaT3)	Slab width
Rinse time after final additions			Grinding loss
Fluorspar additions			Time in reheat furnace
Mild steel addition (B211)			
Metal addition mass			
Time between rinse end and start of cast			

Table 8 Final list of independent variables used in models

In some of the cases, the correlation between the defect and the variable was not as expected. For example, slightly more line inclusions were found for a tundish weight of 24.5 tons (see Figure 13, Appendix B). This is not as expected since a higher tundish mass increases the steel residence time and should assist with inclusion removal. This higher level may be due to other variables that have an effect on inclusion formation at a high tundish mass.

There were other variables where the results did not agree with the findings of the literature study as well. Figure 11 in Appendix B indicates that approximately 20% of non-transition slabs (that is not the first slab of a heat) had skin laminations compared to approximately 5% for transition slabs. Results for the line-inclusion defect were similar (see Figure 12 in Appendix B). The high percentage associated with non-transition slabs may be due to other variables that are not taken into account when looking at each variable individually.

For other variables there was no clear indication of dependencies. Figures 2, 6, 7, 8, 14 and 15 in Appendix B are examples. The assessment of these figures is included in Appendix B.

It is worth noting that the dependencies between defects and the operating variables were more evident for the skin-lamination defect. This may be explained by the fact that individual variables have a major effect on the incidence of skin laminations. In the case of line inclusions the combined effect of variables may be of greater importance. The results from the visual inspection of the data may assist with the interpretation of the modelling results in the next chapter.

5.4 Summary

Raw data have to be filtered before any modelling is done since wrong measurements and missing values can influence results. Records with missing values were removed. The data set was also balanced by adding more defect data. This was necessary to improve the model's prediction accuracy.

A correlation matrix was used to reduce the number of variables. Correlated variables were removed since the effect of these variables was already explained by other variables. The remaining variables were studied by using histograms of the dependent variables for all the independent variables. For the skin-lamination defect the dependencies were evident for a number of variables. Fewer dependencies were evident for the line-inclusion defect, probably because interaction between variables was more important.

Although there are three defects that can be attributed to steel plant operation, only two are considered in this study. This is because inclusions are seldom reported on steel with a gauge less than 3 mm and since an increase in inclusion levels is associated with the hot ductility defect. Therefore only models for skin laminations and line inclusions are constructed.

CHAPTER 6 RESULTS

Separate models were trained for the two types of defects (line inclusions and skin laminations). The results from the various models (neural network, decision tree, regression and combined models) are discussed in this chapter.

6.1 Skin-lamination models

Several techniques are used to compare the models. These include the analysis of errors, lift charts and confusion matrices. In Figures 6 and 7 the errors of the models are compared. The root squared error and the misclassification rates are shown for the training, validation and test data sets. The errors for the training data sets are the smallest while only a small difference is seen for the validation and test data sets. The neural network with twenty hidden neurons performed the best with the minimum root squared error and the lowest misclassification rate.

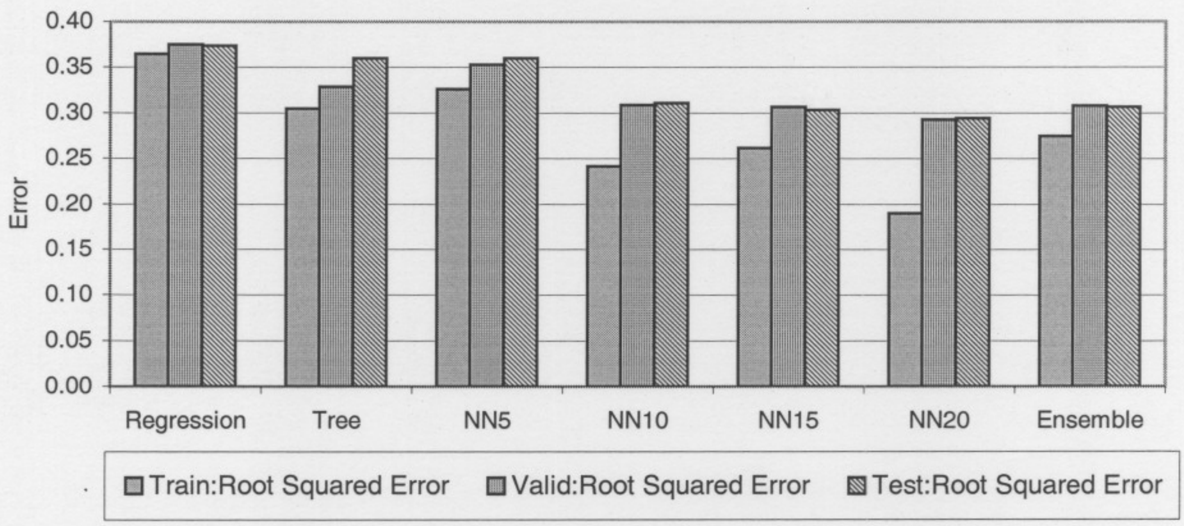


Figure 6 Comparison of skin-lamination model errors for training, validation and test data sets - Root Squared Error

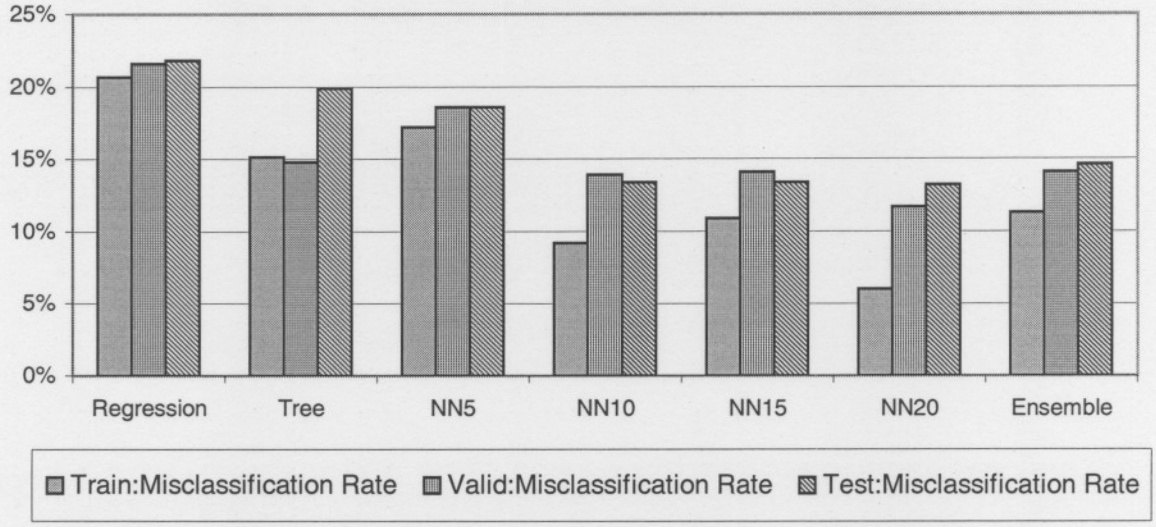
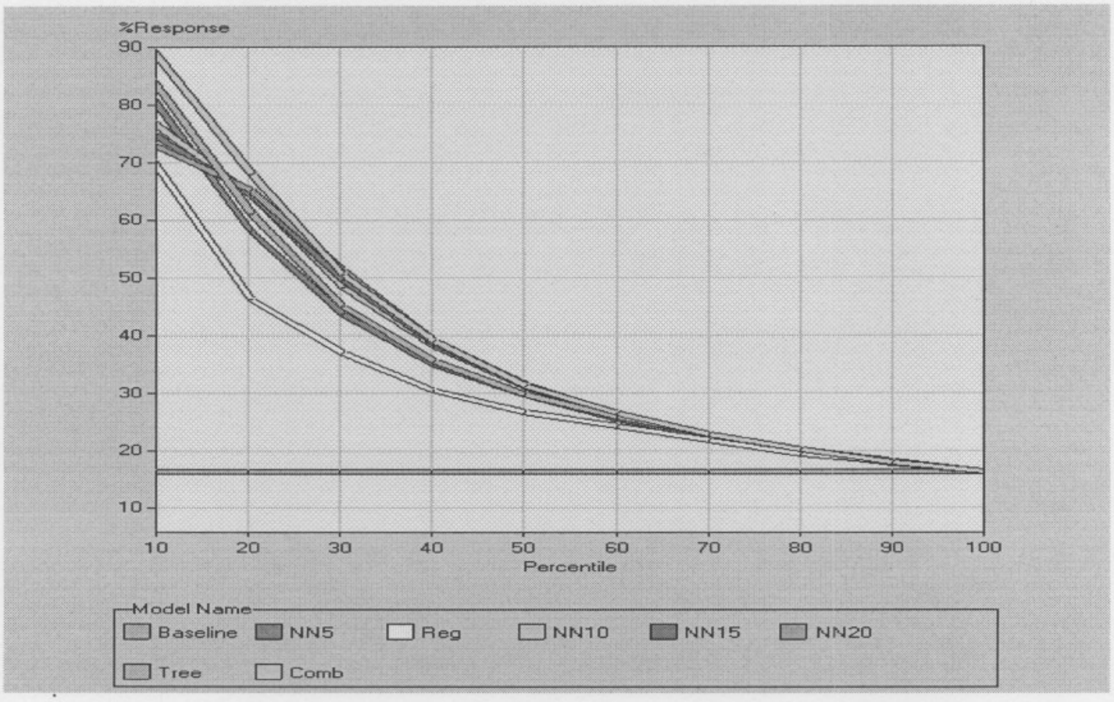


Figure 7 . Comparison of skin-lamination model errors for training, validation and test data sets – Misclassification Rate

Figure 8 shows the lift chart for the skin-lamination models. The ensemble model performed the best with the neural network models with ten, fifteen and twenty neurons not much worse. This figure was constructed by sorting the predicted defect probabilities in descending order. When 20% of the records with the highest defect probability are selected, 70% of the defects are isolated using the combined model. For most of the other models 60 to 67% of the defects are isolated. As the percentage is increased, the difference between the various models becomes smaller. When the upper 50% of the probability records are selected the difference between the best and worst model is less than 5%.

The results from the confusion matrices confirmed the observations from the error and lift charts. All the models performed well on the prediction of no skin laminations, but not on the slabs with skin laminations (see Table 9). The threshold refers to the selection criterion that is used to decide whether a defect is present or not. When the model predicts a probability in excess of 0.5 (probability between zero and one), a defect is assumed. The neural network with twenty hidden neurons achieved the best classification of both slabs with skin laminations and slabs with no skin laminations.



Legend:	NN5 - 5 neurons	NN10 – 10 neurons
	NN15 – 15 neurons	NN20 – 20 neurons
	Reg – Regression	Tree – Decision tree
	Comb – Ensemble of models	

Figure 8 Lift chart for skin-lamination models

Model	Skin laminations correctly predicted	No skin laminations correctly predicted
Regression	43.2%	96.8%
Decision tree	62.3%	97.2%
Neural network with 5 neurons	52.1%	96.6%
Neural network with 10 neurons	70.8%	94.2%
Neural network with 15 neurons	70.0%	94.3%
Neural network with 20 neurons	80.4%	92.4%
Committee of networks	64.6%	97.2%

Table 9 Prediction accuracy of skin-lamination models using the validation data set (threshold 50%)

The results for the test data set are shown in Table 10. These values indicate a small deterioration in the prediction accuracy when compared with the validation data set results. Since this difference is not substantial it is concluded that the models generalise well.

Model	Skin laminations correctly predicted	No skin laminations correctly predicted
Regression	41.8%	97.4%
Decision Tree	54.9%	93.3%
Neural network with 5 neurons	51.9%	96.9%
Neural network with 10 neurons	74.8%	92.8%
Neural network with 15 neurons	70.4%	95.1%
Neural network with 20 neurons	77.2%	91.8%
Committee of networks	62.1%	97.7%

Table 10 Prediction accuracy of skin-lamination models using the test data set (threshold 50%)

The results from the different modelling techniques are discussed in more detail in the next sections.

6.1.1 Neural networks

Four neural network models were trained using different numbers of neurons in the hidden layer (five to twenty). The figures of the average errors that were observed during training are included in Appendix C, Figures 1 to 4. Note that these figures only show the last 500 iterations. A total of 3000 iterations were performed for each model with various restarts from different initial conditions. This was done in an effort to decrease the errors further. The final model in each case was selected where the validation set error was at a minimum. In general, the average error decreased as neurons were added to the model.

Since the neural network with twenty neurons in the hidden layer gave the best results, this network was studied in more detail. Table 11 shows the confusion matrix (cross tabulation of actual and predicted values) for the test data set at a threshold of 50%. This threshold was the default and was used to decide whether a defect was present or not. Of the total of 16% slabs with skin laminations, 77.2% were correctly predicted. Of the slabs with no defects 91.8% were correctly predicted. The accuracy (number of yes and no correct) is therefore 89.5% (91.8% x 84% + 77.2% x 16%). It is also important to determine whether the model will select an acceptable number of slabs for non-grinding. Based on the data from Table 9, 80.8% of the slabs will be selected (91.8% x 84% + 22.8% x 16%).

Actual	Predicted: No	Predicted: Yes
No	91.8%	8.2%
Yes	22.8%	77.2%

Table 11 Confusion matrix for skin-lamination prediction using a neural network with twenty neurons in the hidden layer

6.1.2 Regression

Although the Regression model performed the worst (see Tables 9 and 10 and Figures 6 and 7), the results could be used to identify important variables. At a 50% threshold level, 41.8% of skin laminations are correctly predicted and 97.4 % of no skin lamination cases (for the test data set). This model's performance is improved by using a 15% threshold level (compared to the 50% level initially assumed). With this setting 69 % of skin laminations are correctly predicted and 80% of the no skin lamination cases. The use of this setting is more conservative since more of the skin laminations are correctly predicted. The penalty is however an increase of 17% in the no skin-lamination cases that is predicted incorrectly.

The variables with the biggest influence on skin laminations are mould level fluctuation, rinse station eye size, grinding loss, mould wide and narrow side heat removal, tundish mass and the oscillation frequency. These variables are identified using the T-values in Table 12. Variables with high absolute values are highly correlated with the dependent variable.

Effect Name	Parameter estimate	Effect T-scores	Effect name	Parameter estimate	Effect T-scores
LEVELDIF	31.8038	7.7	WaterFW	-0.0033	-3.1
DeltaT3	-1.0024	-5.3	nm3_min_C	4.5490	2.9
EYE	0.0069	4.5	TiN	0.0630	2.8
Grnd_Loss	0.3777	4.4	MouldH2OT	-0.0508	-2.5
WIDTH1060	-1.1836	-4.2	HeatInCast2	-0.4672	-2.4
DeltaT1	-0.6369	-4.1	GAUGE	-0.1469	-2.4
OSILA	0.0346	4.0	RET	-0.0866	-2.4
Tund_mass	0.0563	3.8	v1SI	-0.7474	-2.4
P	61.6761	3.5	WIDTH1335	-1.0162	-2.2
SuperHeat2	0.0242	3.3	HeatInCast7	-0.9720	-2.2
META_MASS	-0.0002	-3.2	REBLOWNO	-0.2797	-2.1
CLU_RETURNNO	0.6824	3.2	POWDE5	2.5622	2.0
B211	0.0003	3.2	POWDE4	2.2546	2.0

Table 12 Regression variables in order of importance

Mould level fluctuation is the most important variable. High levels of fluctuation cause mould flux entrapment. The importance of this variable was also evident during the visual inspection of the data (see in Appendix B, Figure 5). The eye size provides a similar indicator for the rinse station. A big eye size correlates with a high stirring intensity and possible slag entrapment.

The grinding loss is a function of the quality of the steel. If the slab surface quality is poor, more grinding is required. The importance of this parameter is therefore as expected.

The width was treated as an ordinal variable. The least number of skin-lamination defects is associated with a width of 1060 mm according to Table 12. This was not evident from the histograms as discussed in the previous chapter.

The mould heat flux provides an indirect indicator of conditions in the mould (casting speed and slab width). As shown in section 5.2 there was a reasonable correlation (> 0.5) between the narrow side heat flux and the casting speed. In the correlation analysis it was seen that the wide side heat removal was correlated with the slab width. Figure 1 in Appendix B indicates that low levels of heat flux should be avoided.

The tundish mass is an indicator of continuous caster stability. Another variable, the oscillation frequency influences mould powder feeding and oscillation mark formation. An increase in skin laminations is predicted for an increase in either of these two variables (see Table 12). This is counter-intuitive since higher values of these two variables should actually lead to improved quality. From the histograms of the tundish mass variable slightly higher levels of defects were evident at the higher tonnages. For the oscillation frequency, a few slabs were cast at a very high frequency and all had defects. The high oscillation frequency is related to a higher casting speed and may have corresponded with a low superheat. This may explain the defects at high frequencies. For both variables, variable interaction may therefore be responsible for the higher defect levels. Both these variables are therefore probably indirect indicators of other defect sources.

High superheat was also predicted to result in more skin laminations. This was not evident from the histogram analysis. This may be due to a to lower viscosity mould flux and tundish cover associated with the higher temperature and therefore an increase in the chances of entrapment.

6.1.3 Decision tree

Figure 9 shows that the best model was one with 115 leaves where the fraction correctly specified was 0.95 for the training set and 0.92 for the validation set.

There was good agreement between the importance of variables in the regression and decision tree models (compare Tables 12 and 13). As with the regression model the mould level and mould heat transfer were most important. Other variables such as the number of the heats in the cast and the steel type were also shown to be important. No attempt was made to interpret the decision tree results because the five-way split made results difficult to understand.

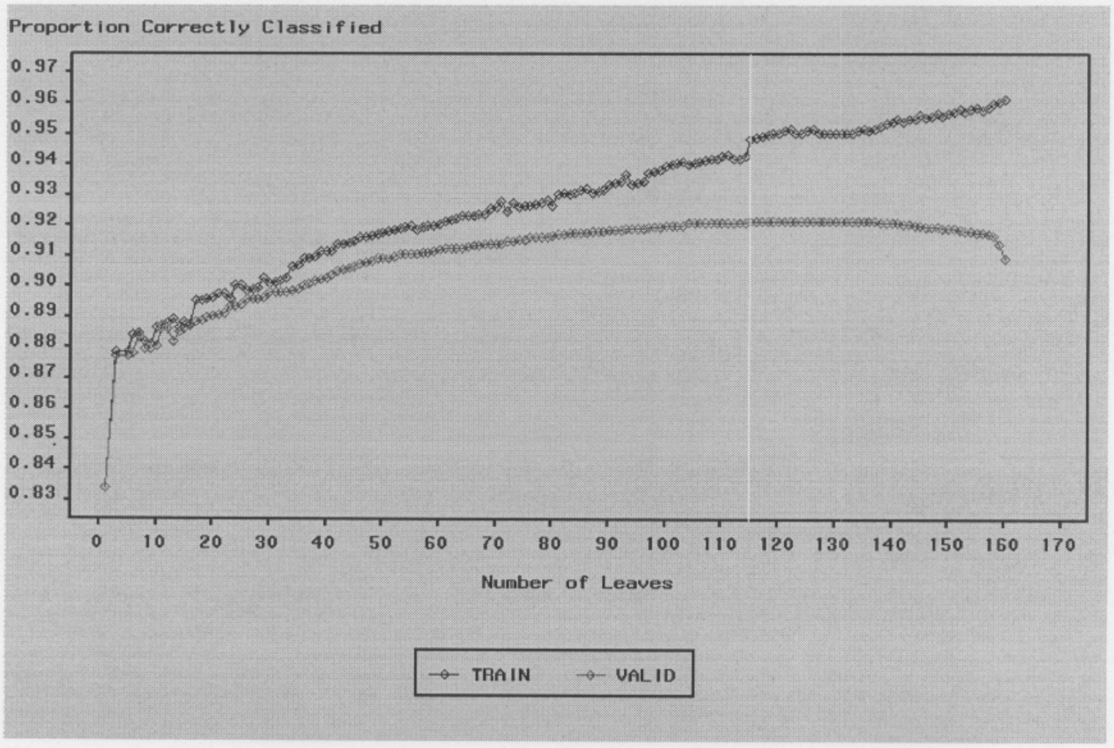


Figure 9 Decision tree proportion of skin laminations correctly classified

Name	Importance	Name	Importance
LEVELDIF	1.00	WATERFW	0.19
HEATINCAST	0.67	MOULDH2OT	0.19
DELTAT3	0.61	SECH2OT	0.19
STEELTYPE	0.60	RET	0.16
TUND_MASS	0.53	META_MASS	0.16
DELTAT1	0.51	V1SI	0.15
GRND_LOSS	0.36	TRANSITION	0.15
CONTACT_TIME	0.32	SUPERHEAT2	0.14
TI3O5	0.30	NM3_MIN_C	0.09
BUBB_TIME	0.22	S	0.07
AL2O3	0.21	POWDE	0.07
B211	0.21	TOT_AR	0.07
STOPDIF	0.21		

Table 13 Variables included in the skin-lamination decision tree

6.1.4 Committee of models

The combination of models did not improve the prediction accuracy. The lift chart shows that the committee of models performed well, but results from the confusion matrices suggests that the combined model did not improve the prediction accuracy.

6.2 Line-inclusion models

The same models were trained for the line-inclusion defect. The neural networks gave the best results as was the case for the skin-lamination defect. The neural network with fifteen neurons in the hidden layer gave the lowest errors and misclassification rates (see Figures 10 and 11). The errors for the validation and test data sets for the neural networks with ten, fifteen and twenty hidden neurons and the combined model were very similar. In Figure 11 it is seen that the misclassification rate for the neural network with fifteen hidden neurons is significantly lower than the other models. This was the case for all three the data sets. The committee of models did not improve the accuracy of the prediction compared to the fifteen-neuron neural network model.

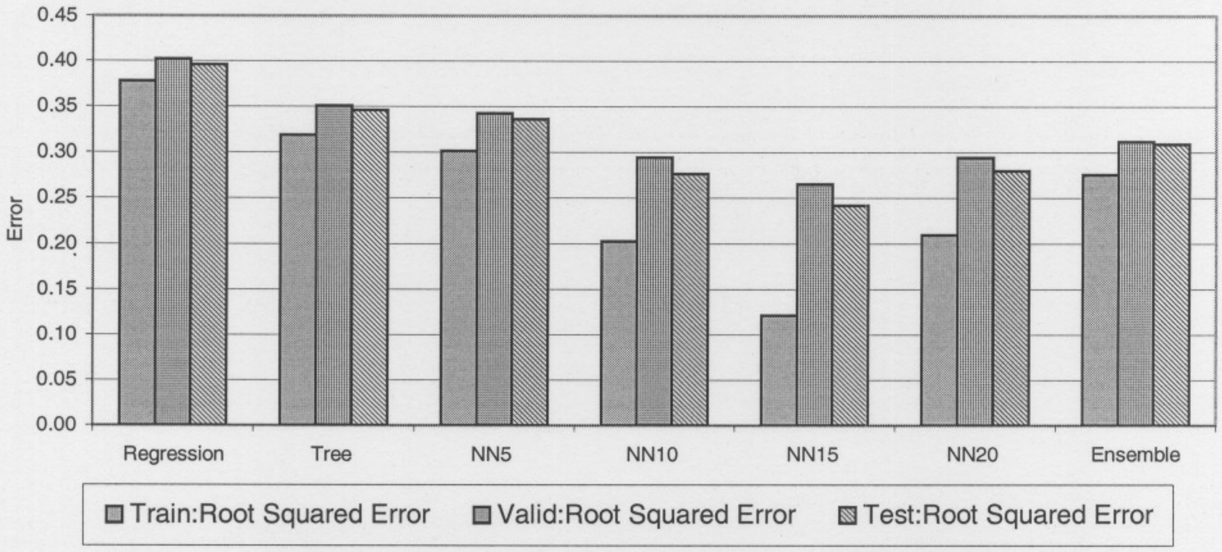


Figure 10 Comparison of line-inclusion model errors for training, validation and test data sets – Root Squared Error

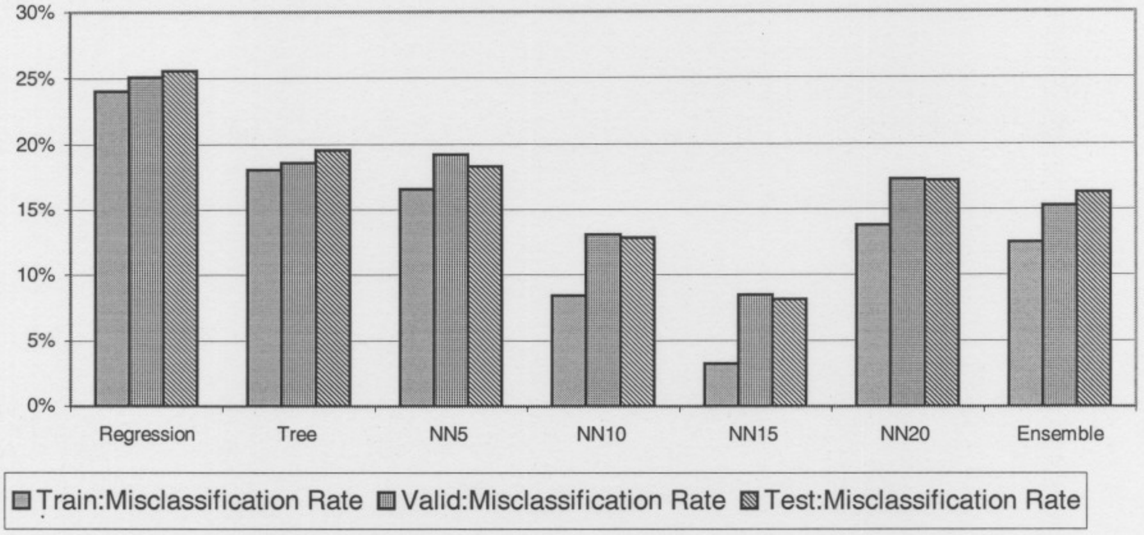
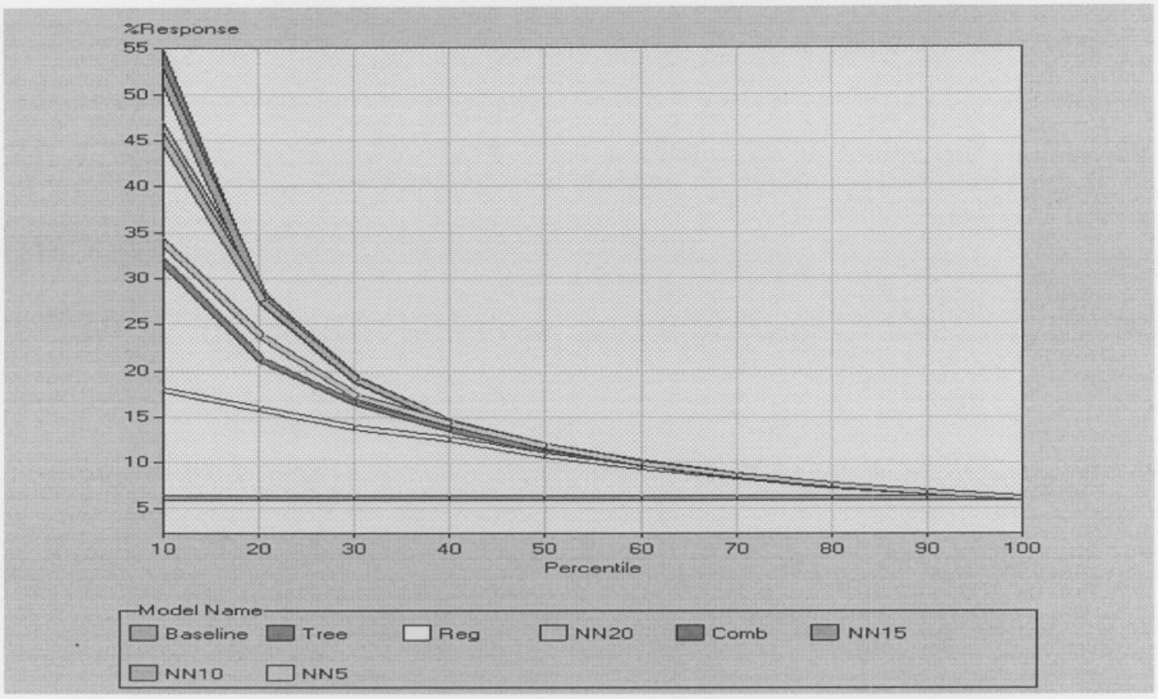


Figure 11 Comparison of line-inclusion model errors for training, validation and test data sets – Misclassification Rate

The results from the lift chart confirmed the observations made from the model errors (see Figure 12). The difference in the performance of the various line-inclusion models was larger compared to the skin-lamination results. The regression and decision tree performed the worst and little benefit would be achieved by using these models compared to a random selection of slabs. The increase in the number of hidden layer neurons improved the performance of the neural network.

The evaluation of the prediction accuracy showed that the committee of models' overall performance was worse compared to most of the neural network models. This model predicted only 40.7% of line inclusions correctly (see Table 14). The neural network with fifteen neurons in the hidden layer performed the best with 81.1% of line inclusions predicted correctly and 95% of the no line-inclusion cases (for the validation data set).



Legend: NN5 - 15 neurons NN10 – 10 neurons
NN15 – 15 neurons NN20 – 20 neurons
Reg – Regression Tree – Decision tree
Comb – Committee of models

Figure 12 Lift Chart for line-inclusion models

Model	Line inclusions correctly predicted	No line inclusions correctly predicted
Regression	2.0%	98.9%
Decision Tree	28.5%	99.0%
Neural network with 5 neurons	33.3%	92%
Neural network with 10 neurons	64%	94.6%
Neural network with 15 neurons	81.1%	95%
Neural network with 20 neurons	39.2%	97.1%
Committee of Networks	40.7%	99.3%

Table 14 Prediction accuracy of line-inclusion models for the validation data set (threshold 50%)

As was the case with the skin-lamination models, there was a slight deterioration in the performance for the test data set (compare Tables 14 and 15). The performance was still acceptable and it was concluded the models were able to generalise.

Model	Line inclusions correctly predicted	No line inclusions correctly predicted
Regression	2.2%	98.6%
Decision tree	25.4%	98.8%
Neural network with 5 neurons	31.1%	97.3%
Neural network with 10 neurons	59.4%	96.4%
Neural network with 15 neurons	80.4%	95.7%
Neural network with 20 neurons	37.0%	98.1%
Committee of networks	36.2%	99.5%

Table 15 Prediction accuracy of line-inclusion models for the test data set (threshold 50%)

6.2.1 Neural networks

The average errors for the neural networks that were observed during training can be seen in Appendix C, Figures 5 to 8. The model with the minimum error for the validation set was selected in each case. The neural network with fifteen hidden neurons performed the best.

Table 16 shows the confusion matrix for the test data set for the neural network with fifteen neurons in the hidden layer at a threshold of 50%. Of the total of 6% of slabs with line inclusions, 80.4% were correctly predicted. Of the slabs with no defects 95.7% were correctly predicted. The accuracy (number of yes and no correct) is therefore 94.8% (95.7% x 94% + 80.4% x 6%). It is also important to determine whether the model will select an acceptable number of slabs for non-grinding. Based on the data from Table 16, 91.1% of the slabs will be selected (95.7% x 94% + 19.6% x 6%). Assuming that the two defects are mutually exclusive (both defects do not occur on the same slab), the average number of slabs selected will be 86%.

Actual	Predicted: No	Predicted: Yes
No	95.7%	4.3%
Yes	19.6%	80.4%

Table 16 Confusion matrix for line-inclusion prediction using a neural network with twenty neurons in the hidden layer

6.2.2 Regression

The regression model's performance was poor. From Table 14 it is seen that only 2% of the defects were predicted correctly. The model was however still useful to identify important variables. These variables and their importance are shown in Table 17.

The final gauge was the most important variable. This was probably due to the difficulty of detecting these small defects on thicker gauge material. The contact time, that is the time between end of rinse and start of cast, was also important. The longer this time, the lower the probability of a line-inclusion defect.

High phosphorus levels can contribute to hot ductility problems and cause defects that may be confused with line inclusions. The coefficient indicates that an increase in phosphorus should decrease the probability of line inclusions. This is counter-intuitive.

As was the case with skin laminations, a lower probability of defects was predicted for 1060 mm wide slabs. The stopper difference is an indicator of instability between the tundish and the mould and higher values were associated with a higher incidence of line inclusions.

If the heat was not reblown, the probability of defects was less. However, if the heat was not returned to the converter, the data indicated a higher probability of line inclusions. This was contrary to expectations and may once again indicate the effect of variable interaction that was not taken into account with the linear regression model. The effect of titanium nitride was also contrary to expectation and may be due to the interaction with the superheat. The same contradiction was seen for the time between last additions and the end of rinse.

As was indicated, several of the variables predicted the opposite behaviour to that which was expected. This may explain why the linear model did not give good predictions. Since the non-linear effects and interaction between variables were not taken into account, the linear model performed poorly. This is where the neural network has the major benefit.

Effect name	Parameter estimate	Effect T-scores	Effect name	Parameter estimate	Effect T-scores
GAUGE	-0.6575	-10.7	WIDTH990	2.2093	3.1
CONTACT_TIME	-4.0740	-4.8	WIDTH1335	-1.1543	-3.0
P	-93.8299	-4.6	v1SI	0.8289	2.7
WIDTH1060	-1.2207	-4.5	B211	0.0002	2.6
STOPDIF	3.2951	4.2	UnshroudedNO	-0.9032	-2.5
REBLOWNO	-0.4575	-3.8	DeltaT3	-0.4720	-2.4
TiN	-0.1039	-3.7	OSILA	0.0197	2.4
CLU_RETURNNO	0.7674	3.6	POWDE2	2.3267	2.3
ADD_TO_GAS_END	1.8735	3.5	FF	0.1553	2.1
Tund_mass	-0.0447	-3.3	Bubb_time	-0.0207	-2.0
RET	-0.1289	-3.2	SteeltypeT413	3.8019	2.0

Table 17 Regression variables for line inclusions in order of importance

6.2.3 Decision tree

The decision tree with 52 leaves gave the most correct predictions for the validation data set (see Figure 13). There was a good agreement between the important variables in the regression and decision tree (compare Table 17 and 18).

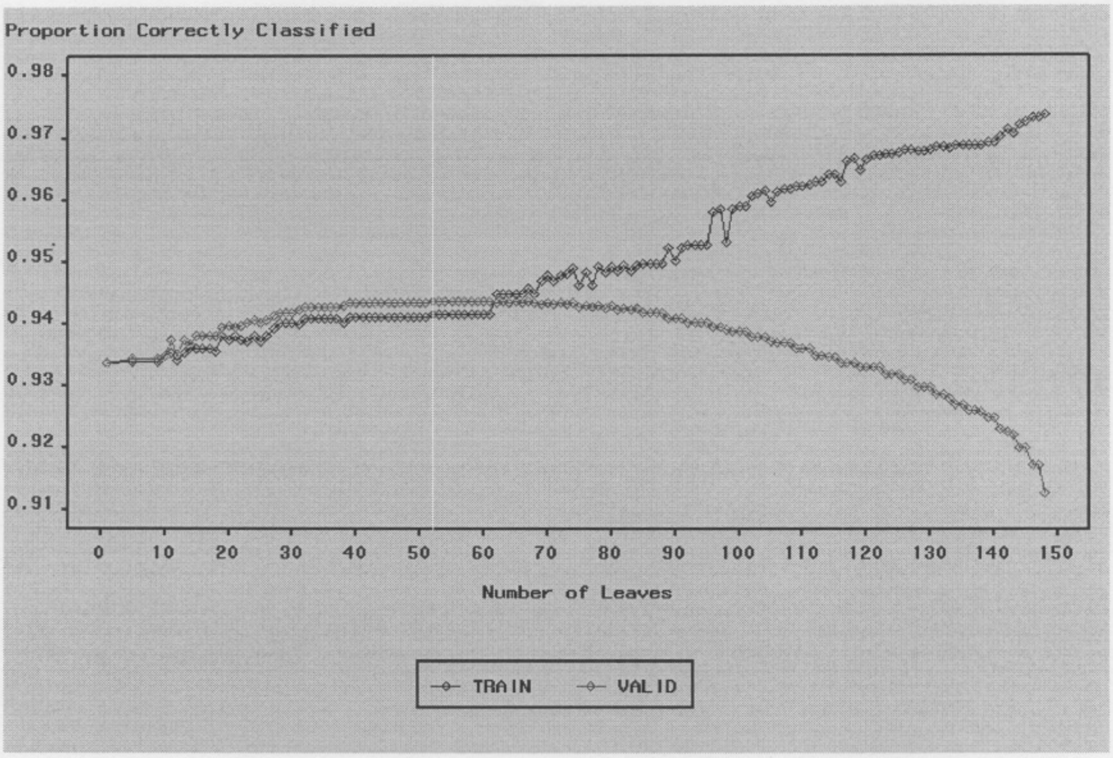


Figure 13 Decision tree proportion line inclusions correctly classified

Name	Importance
GAUGE	1.00
FF	0.77
BUBB_TIME	0.44
DELTAT1	0.35
GRND_LOSS	0.32
AL2O3	0.29
DELTAT3	0.29
META_MASS	0.20
TUND_MASS	0.16
TI3O5	0.15
OSILA	0.15
CONTACT_TIME	0.15

Table 18 Variables included in the line-inclusion decision tree

6.2.4 Committee of models

The committee of models again did not improve on the performance of the individual neural network models. Only the model with five hidden neurons performed worse.

6.3 Summary

The results for the skin-lamination and line-inclusion models were discussed. The non-linear models performed better compared to the linear regression models. The neural networks gave the best overall prediction accuracy. A neural network with fifteen neurons in the hidden layer performed the best for the line-inclusion defect while twenty neurons were necessary for the skin-lamination defect. When the models were applied to the test data sets, there was only a small deterioration in prediction accuracy and the models can be said to generalise well.

Although the regression and decision tree models performed worse compared to the neural networks, these models provided some indication of the variables with the biggest influence on the independent variables. The regression model for the skin-lamination defect was the most useful and the majority of coefficients were as expected. Many of the coefficients for the variables in the line-inclusion model were not as expected. This was probably due to non-linear effects of variables and interaction between variables that was not taken into account in the linear regression model.

CHAPTER 7 CONCLUSIONS

7.1 Identification of variables

Various factors affect the slab surface quality of the steel. A literature study was done to identify the variables to be included in the model. These included operating variables from the converter to the continuous caster.

A few of the variables that are deemed critical for the control of quality are not measured in the plant. In these cases indirect measurements are identified that can be used as an indication. This is especially important where the variable is critical for surface quality prediction. Examples of the variables that are not measured include the steel temperature stratification in the ladle, ladle-, tundish- and mould slag entrapment, steel solidification structure, liquid steel temperature at the meniscus, mould powder solidification, mould powder consumption, oscillation mark formation, the standing wave in the mould and flow in the mould. The indirect measurements that are used include the standing time between end of rinse and start of cast, tundish weight, casting speed, steel temperature and the mould heat flux.

Where variables are not measured and no indirect measurements are available, fundamental models can be used. One such a model is a solubility product model that was developed at Columbus Stainless. This model makes predictions of the internal cleanliness of the steel. These predictions have been shown in the past to correlate with the slab surface cleanliness.

The amount of data that is required for modelling increases as the number of independent variables increase. It is therefore important to remove variables that are redundant before modelling is done. A correlation matrix was used to reduce the number of variables. Variables that were highly correlated (greater than 0.7 or less than –0.7 for an inverse correlation) were removed. In this way the possible adverse effect of correlated variables was also removed.

Charting was used to do a data pre-analysis. For some of the variables the effect on the incidence of defects was obvious. In other cases the cause-effect relationship was not that strong and with others no effect was evident. This pre-analysis was important and assisted with the interpretation of the modelling results.

7.2 Modelling results

Neural network, decision tree and linear regression models were used to predict slab surface quality. A major benefit from models such as neural networks and decision trees is that non-linear dependencies between dependent and independent variables can be modelled. Decision trees have the added advantage that the rules used for decision making are known after modelling. This is not the case for neural networks. Over-fitting is a concern for both non-linear modelling techniques while linear regression models do not have this problem.

Multilayer perceptron neural network models are the most popular models used in literature. Several training algorithms were evaluated in this study and included error backpropagation, the conjugate gradient method and more advanced techniques like Levenberg-Marquardt. It was decided to use the conjugate gradient method since this method was available for training both the neural network and regression models and made comparison simpler.

Early stopping was used to prevent over-training. In all the cases the models that were selected corresponded with the minimum in the validation set error. An increase in this error indicated over-training (that is modelling of the noise in the data).

Several techniques were used to select the best model. This included the use of model error values, misclassification rates, lift charts and confusion matrices. The neural network models performed the best for both types of defects. This was obvious from both lift chart and confusion matrix results. The prediction accuracy for both the training and validation data sets were improved by adding neurons to the hidden layer. The twenty-neuron neural network model gave the best results for the skin-lamination defect while a fifteen-neuron model performed the best for the line-inclusion defect.

Model generalisation was tested using a test data set that consisted of data that was not used during the training of the models. Although the accuracy of predictions deteriorated between the validation and test data sets, this was small and it was concluded that the models were able to generalise. The line-inclusion and skin-lamination defect models performed similarly with both these models being able to predict approximately 80% of the defect occurrence. Regular testing of the models should be done to assess the deterioration of predictions over time. This can be used to determine retrain intervals.

The committee of models did not improve the prediction accuracy. This may have been due to the fact that there was not enough difference between the various model predictions and also because the best models already gave excellent results.

Although the regression models performed the worst, the results could still be used to identify the variables that influenced the quality. The regression model for the skin-lamination defect performed significantly better compared to the line-inclusion model. This may be explained by the size of the defect (line inclusions are smaller than skin laminations and more difficult to observe), but may also indicate that with the line-inclusion defect the interaction between variables and non-linear effects are important.

Although the decision tree also took non-linear dependencies into account, this model did not perform as well as the neural networks. It should however be borne in mind that only five-way splits were considered and an increase in the number of splits might have improved the prediction accuracy.

In this study neural networks were developed to predict slab surface quality and it was shown that these models gave superior results compared to both linear regression and decision tree models. The models that were developed can be used to select at least 75% of slabs with an improved accuracy compared to random selection of slabs. It was also concluded that a committee of models did not improve the prediction accuracy compared to the individual models.

7.3 Future work

Several data problems were highlighted in this study. The average values that were used to capture the effect of processing parameters on steel surface quality were not as effective as initially expected. Another problem is the human effect associated with the collection of defect data at the final processing lines. The first problem can be addressed by using time series data measurements. Although this will be a more tedious exercise, the effect of transient conditions can be taken into account. The second problem can be addressed by taking the severity of the defects into account. Light defects can be ignored and may improve the prediction accuracy of the models. The quality of defect data may also be improved by using inspection data collected by automatic systems.

In this study neural networks and decision trees were considered to model non-linear effects. Alternative methods that could be considered are fuzzy logic and neuro-fuzzy systems. Both are non-linear techniques with the added benefit that the decision-making process is known after model development. Neuro-fuzzy models combine the benefits from both neural networks and fuzzy systems. In this approach the fuzzy rules are derived using neural network training methods.

The neural networks performed the best in this study and are definitely a viable alternative for defect prediction. It should however be borne in mind that operating practices in a steel plant is not stagnant and this will necessitate frequent retraining of the models. More work is required to automatically retrain the models on a regular basis. Model prediction accuracy can also be improved further if more of the variables that are not currently measured can be included.

REFERENCES

ACKERMAN, M.J. & ORBAN, W.G., 2002. A perspective of practical issues relating to tundish development. (*In* Iron and Steel Society, Inc. Steelmaking Conference Proceedings Vol. 85. Warrendale, USA: Iron and Steel Society, Inc. p. 409-418.)

ALDRICH, C. 1997. Neural networks for the process industries –course notes. Stellenbosch. 112p.

ANON. 2001. Neural network node: Reference. Cary, USA: SAS Institute Inc. 50p.

AUSTIN, P.R., CAMPLIN, J.M., HERBERTSON, J. & TAGGERT, J. 1992. Mathematical modelling of thermal stratification and drainage of steel ladles. *ISIJ International*, 32 (2): 196-202.

BABU, J. & HANRATTY, F.W. 1993. Predictive control of quality in a batch manufacturing process using artificial neural network models. *Ind. Eng. Chem.* Res. 32: 1951-1961.

BEHZADIPOUR, S., KHAJEPOUR, A. & LENARD, J. 2002. Power estimation in shape rolling using neural networks. *Iron & Steelmaking,* 19-27, October.

BELLOMO, P., PALCHETTI, M., SANTA MARIA, E. & SALVEMINI, G. 1995. Neural Network utilization for break-outs monitoring. (*In* Iron and Steel Society, Inc. Steelmaking Conference Proceedings Vol. 78. Warrendale, USA: Iron and Steel Society, Inc. p. 345-349.)

BELLEMO, P., BRASCUGLI, G. & MILONE, M. 1998. Surface quality prediction of stainless steel C.C. slabs by artificial neural networks. (*In* Iron and Steel Society, Inc. Steelmaking Conference Proceedings Vol. 81. Warrendale, USA: Iron and Steel Society, Inc. p. 199-206.)

BERRY, M.J.A. & LINOFF, G., 1997. Data mining techniques for marketing, sales, and customer support. New York: John Wiley & Sons, Inc. 454p.

CABENA, P., HADJINIAN, P., STADLER, R., VERHEES, J. & ZANASI, A. 1998. Discovering data mining: from concept to implementation. New Jersey: Prentice-Hall, Inc. 195p.

CAPOTOSTI, L., CAPOTOSTI, R., MACCI, F, MILONE, M. & SPACCAROTELLA, A. 1994. Computerised surface quality prediction in the continuous casting of stainless steel slabs. (*In* Second European Continuous casting Conference. Vol. 1. p. 201-208.)

CHAKABORTY, S. & SAHAI, Y. 1992. Effect of slag cover on heat loss and liquid steel flow in ladles before and during teeming to a continuous casting tundish. *Metallurgical Transactions B*, 23 (B): 135-151.

DEB, P., MUKHOPADHYAY, A., GHOSH, A. & BASU, B. 2001. Operational experience with a mathematical model for temperature prediction in secondary steelmaking. *Steel Research*, 72 (5,6): 200-207.

DYSON, D.J., ROSE, A.J. & WHITWOOD, M.M. 1998. Studies in development of clean steels Part 2 Use of chemical analysis. *Ironmaking and Steelmaking*, 25 (4): 279-286.

EMLING, W.H. & DAWSON, S. 1991. Mold instrumentation for breakout detection and control. (*In* Iron and Steel Society, Inc. Steelmaking Conference Proceedings Vol. 74. Warrendale, USA: Iron and Steel Society, Inc. p. 197-217.)

FLAMENT, F., THIBAULT, J. & HODOUIN, D. 1993. Neural network based control of mineral grinding plants. *Minerals Engineering,* 6(3): 235-249.

FLINT, P.J., 1990. A three-dimensional finite difference model of heat transfer, fluid flow and solidification in the continuous slab cast. (*In* Iron and Steel Society, Inc. Steelmaking Conference Proceedings Vol. 73. Warrendale, USA: Iron and Steel Society, Inc. p. 481- 490.)

FREEMAN, J.A. & SKAPURA, D.M., 1991. Neural networks: Algorithms, Applications, and Programming Techniques. New York: Addison-Wesley Publishing Company. 401p.

GRIP, C.E., JONSSON, L. & JONSSON, P.G. 1997. Prediction of emptying flows in ladles and verification with data from trace element plant trials. *ISIJ International*, 37 (11): 1081-1090.

GUPTA, D. & LAHIRI, A.K., 1996. Cold model study of slag entrainment into liquid steel in continuous slab caster. *Ironmaking and Steelmaking*, 23 (4): 361-363.

HAGAN, M.T., DEMUTH, H.B. & BEALE, M. 1996. Neural network design. Boston: PWS Publishing Company. 19-21p.

HERBERTSON, J., HE, Q.L., FLINT, P.J & MAHAPATRA,R.B. 1991. Modelling of metal delivery to continuous caster moulds. (*In* Iron and Steel Society, Inc. Mold operation for quality and productivity. Warrendale, USA: Iron and steel society, Inc., p.135-149.)

HLINKA, J.W. & MILLER, T.W., 1970. Temperature loss in liquid steel-refractory systems. *Iron and Steel Engineer,* 123-133.

HO, B., 1992. Characterization of interfacial heat transfer in the continuous slab casting process. Urbana-Champaign: University of Illinois. (Thesis-M.Sc.) 263p.

JIMBO, I., OZTURK, B. & FELBAUER, S. 1991. Some aspects of chemical phenomena in the mold of a continuous slab caster. (*In* Mold operation for quality and productivity. Warrendale, USA: Iron and steel society, Inc., 117-126.)

KUBOTA, J., OKIMOTO, K. SUZUKI, M, SHIRAYAMA, A. & MASAOKA, T. 1990. Mechanism of level fluctuation and mould powder catching in slab continuous casting at high speeds. (*In* ISIJ. Proceedings of the sixth international Iron and steel congress. Nagoya: ISIJ. p.356-363.)

LAING, W., CHEEK, D.L. & MUSTOE, T.N. 1997. Improving productivity and quality through control of the steel temperature in the tundish. (*In* Iron and Steel Society, Inc. Steelmaking Conference Proceedings Vol. 80. Warrendale, USA: Iron and Steel Society, Inc. p. 521-527.)

LU, M., CHEN, J. & TSAI, D. 1995. Solidification of a slab in a continuous casting mold from the viewpoint of heat transfer. (*In* Iron and Steel Society, Inc. Steelmaking Conference Proceedings Vol. 78. Warrendale, USA: Iron and Steel Society, Inc. p. 359-366.)

MAHAPATRA, R.B., BRIMACOMBE, J.K. & SAMARASEKERA, I.V. 1991. Mold behavior and its influence on quality in the continuous casting of steel slabs: Part II. Mold heat transfer, mold. *Metallurgical Transactions B*, 22 (B): 875 - 888.

MC DRAVID, R.M. & THOMAS, B.G., 1996. Flow and thermal behaviour of the top surface flux / powder layers in continuous casting molds. *Metallurgical and Materials Transactions*, 27 (B): 672-685.

MIKI, Y., SHIMADA, Y. & THOMAS, B.G. 1997. Model of inclusion removal during RH degassing of steel. *Iron and Steelmaker*, 24 (8): 31-38.

MILLS, K., 1991. The performance of casting powders and their effect on surface quality. (*In* Iron and Steel Society, Inc. Steelmaking Conference Proceedings Vol. 74. Warrendale, USA: Iron and Steel Society, Inc. p.121-129.)

MILLS, K., 1997. Mould powders for continuous casting. London: Imperial College. 125p.

MORALES, R.D., LOPEZ-RAMIREZ, S., BARRETO, J. & Palofox-Ramos, J. 2000. Modeling of thermal effects of the steel flow in tundishes. (*In* Iron and Steel Society, Inc. Process Technology Conference Proceedings. Warrendale, USA: Iron and Steel Society, Inc. p.955-989.)

NORSTROM, L. & MALM, S. 1979. A new aspect of hot ductility in nitrogen alloyed austenitic stainless steels. *Scandinavian Journal of Metallurgy*, (8): 31-33.

NUNNINGTON, R.C. & SUTCLIFFE, N., 2001. The Steelmaking and casting of Ti stabilised stainless steels. (*In* Iron and Steel Society, Inc. ISS Electric Furnace Conference. Warrendale, USA: Iron and Steel Society, Inc. p.1-34.)

OGIBAYASHI, S., MUKAI, T. & MIMURA, Y. 1987. Mold powder technology for continuous casting of low-carbon aluminium-killed steel. *Nippon Steel Technical Report*, 34: 1-10.

OLIKA, B., PAN, Y. & BJORKMAN, B. 1996. Numerical simulation and industrial investigation of the melt stratification phenomena in ladles holding molten steel. *Scaninject*, 7: 361-384.

POTTS, W.J.E. 1998. Data mining primer: overview of applications and methods. Cary, USA: SAS Institute Inc. 70p.

POTTS, W.J.E. 2000. Neural network modelling: Course notes. Cary, USA: SAS Institute Inc. 265p.

POTTS, W.J.E. 2001. Decision tree modelling: Course notes. Cary, USA: SAS Institute Inc. 133p.

SAARELAINEN, E., SAARINEN, P., VEISTARO, M. & VISA, A. 1999. Prediction of nozzle clogging – A neural computing approach. (*In* Iron and Steel Society, Inc. Steelmaking Conference Proceedings Vol. 82. Warrendale, USA: Iron and Steel Society, Inc. p. 89-95.)

SINGH, S.B., BHADESHIA, H.K.D.H. & MACKAY, D.J.C. 1998. Neural network analysis of steel plate processing. *Ironmaking and Steelmaking*, 25 (5): 355-365.

SUUTALA, N. 1983. Solidification of austenitic stainless steels. (*In* Third Scandinavian Symposium in materials science. June. p. 53-60.)

THOMAS, B.G., NAJJAR, F.M. & MIKA, L.J. 1990. The removal of superheat from continuous casting moulds. (*In* F. Weinberg International Symposium on Solidification Processing. p.131-145.)

VAN DEVENTER, J.S.J., REUTER, M.A. & ALDRICH, C. 1994. The application of neural nets in the metallurgical industry. *Minerals Engineering*, 7 (5/6): 793-809.

WANG, B., WALKER, B.N. & SAMARASEKERA, I.V. 1999. Shell growth, surface quality and mould taper design for high-speed casting of stainless steel billets. *Canadian Metallurgical Quarterly*, 39 (4):

441-454.

WESTPHAL, C. & BLAXTON, T. 1998. Data Mining Solutions. John Wiley & Sons, Inc. USA. 617p.

WIELANGA, D.,LUCAS, B. & GEORGES, J. 1999. Enterprise Miner: Applying data mining techniques: Course Notes. Cary, USA: SAS Institute Inc. 157p.

WOLF, M.M., 1980. Investigation into the relationship between heat flux and shell growth in continuous casting moulds. *Transactions ISIJ*, 20: 710-717.

WOLF, M.M. 1997. Mould powder consumption – a useful criterion? (*In* Wolf, M. Continuous Casting, Volume. 9: Initial solidification and strand surface quality of peritectic steels. Warrendale, USA: Iron and Steel Society, Inc. p.251-258.)

ZHU, M., INOMOTO, T. & SAWADA, I. 1995. Fluid flow and mixing phenomena in the ladle stirred by argon through multi-tuyere. *ISIJ International*, 35 (5): 472-479.

ZUO, G. & BJÖRKMAN,B. 2001. Monitoring the blast furnace process using neural networks and knowledge-based system. *Steel Research*, 72(4): 115-124.

APPENDIX A SAS PROGRAM FOR DATA FILTERING

```
data SQI0503.SQI052003
set all4;
format Unshrouded $3.
         Burn_Open $3.;
if Grnd_Loss < 8 and Grnd_Loss ne '';
if Ret >2;
if CCShroud ne '';
if CCB_Open ne '';
if BURNOPEN ne '';
if POKEOPEN ne '';
if NCRShroud ne '';
if MouldH2OT ne '';
if SecH2OT ne '';
if WaterFW ne '';
if WaterLW ne '';
if WaterLN ne '';
if WaterRN ne '';

if CCShroud = 'Yes' or NCRShroud = 'Yes' then
Unshrouded = 'Yes';
else if CCShroud = 'No' and NCRShroud = 'No' then
Unshrouded = 'No';

if CCB_Open = 'Yes' or Burnopen = 'Yes' then
Burn_Open = 'Yes';
else if CCB_Open = 'No' and Burnopen = 'No' then
Burn_Open = 'No';

if UNSHROUDED = 'Yes' then UNSHROUDED = 'Ye';
if BURN_OPEN = 'Yes' then BURN_OPEN = 'Ye';

if NCRlumps ne '';
if BOIL ne '';
if POWDE ne '';
if WIDTH ne '';
if LEVEL ne '';
if SPEED ne '';
if OSILA ne '';
if LEVELDIF ne '';
if SPEEDIF ne '';
if STOPDIF ne '';

if Tund_Mass > 2;* and Tund_Mass < 16;
if SuperHeat2 > 10 and SuperHeat2 < 80;
if TLN_TRN >0.1 and TLN_TRN < 2;
if TFW_TLW ne '';
if Wide_Narrow ne '';

if Type < 30900 and Type >30999 then do;
  if v1Si >=0.1 and V1Si <=0.9;
end;
else if Type >= 30900 and Type <=30999 then do;
  if V1Si <=1.7;
```

```
end;
if v1Si ne '';

if REBLOW = "No" or REBLOW = "Yes";
if CLU_RETURN = "No" or CLU_RETURN = "Yes";
if LASI_Mass = '' then LASI_Mass = 0;
if FLUO_Mass = '' then FLUO_Mass = 0;
if META_MASS = '' then Meta_Mass = 0;
if B211 = '' then B211 = 0;

if RIN_TEMP >0.1 and (RIN_TEMP-Tund_Temp) >= 6;

if (CONTACT_TIME * 60) > 5;
if Bubb_time > 4;
if Tot_ar > 0.5;
if EYE ne '';
if nm3_min_C > 0; if Al2O3 ne '';
if Ti3O5 ne '';
if TiN ne '';

if WS ne 'CON';
if GAUGE <= 3.1;
if HeatInCast ne '';

*Add column with steeltype groups;
if Type = 31663 then
Steeltype = 'T'||substr(TYPE,8,3)||'Ti';
else if.Type = 31619 or Type = 31618 or Type = 31628 then
Steeltype = 'T'||substr(TYPE,8,5);
else
Steeltype = 'T'||substr(TYPE,8,3);

if REBLOW = 'Yes' then REBLOW = 'Ye';
if CLU_RETURN = 'Yes' then CLU_RETURN = 'Ye';
 if Width >= 1450 then Width = 1565;
 else if Width >= 1300 and Width <1450 then Width = 1335;
 else if Width >= 1286 and Width <1300 then Width = 1290;
 else if Width >= 1150 and Width <1286 then Width = 1280;
 else if Width >= 1070 and Width <1150 then Width = 1080;
 else if Width >= 1030 and Width <1070 then Width = 1060;
 else Width = 990;

if SuperHeat2 > 0;

Cr_Ni2 = (Cr + 1.37*Mo+1.5*Si+2*Nb+3*Ti)/(Ni+0.31*Mn+22*C+14.2*N+Cu);
FF=2.22*(1.4*(Cr+Mo+1.5*Si+2.4*Ti+0.5*Nb)-(Ni+0.5*Mn+30*(C+N)+12.5));

if Date ne '';

run;
```

APPENDIX B VISUALISATION OF DATA

The SAS multi-plot facility was used to construct histograms for the steel plant operating variables for the skin-lamination and line-inclusion defects. Some of these figures are included in this appendix. Data were separated into bins for each independent variable and the frequency of the defect and no defect records for each bin is shown. When some of the bins contain a higher percentage of defects this indicates operating ranges to avoid.

Figure 1 shows the incidence of skin laminations as a function of mould narrow side heat removal. When the mould water temperature increase was less than 4 °C, the probability of a defect was almost a hundred percent. Figure 2 shows the incidence of line inclusions for the same variable. Levels of high defect incidence are not as clear as for the skin-lamination defect.

Figures 3 and 4 are the histograms for the heat in the cast variable (a cast consists of a number of heats cast without stopping the caster). For both defects the incidence of defects is higher on the first heat. The defect levels at heat in the cast levels of greater than six appear to have fewer defects present.

The mould level fluctuation was identified as one of the critical variables for quality. Figure 5 and 6 shows the histograms for this variable. At fluctuation levels greater than 0.035, there is a high incidence of skin laminations. It is not possible to derive any limits for the line-inclusion defect. This is because high mould level fluctuation will lead to slag entrapment (and therefore skin laminations) and will not necessarily impede inclusion removal.

The effect of superheat is not evident from Figures 7 and 8. For skin laminations the effect may be small since the steel temperature will not necessarily influence slag entrapment. Based on the literature study it is expected that more line inclusions are associated with low superheats (smaller than twenty degree Celsius). It may once again be due to interaction between variables (superheat and titanium oxide levels for example) that the individual cases do not show any correlation.

Figure 9 shows the skin-lamination defect as a function of the steel titanium nitride level. The zero-bin should be ignored since the defects grouped in this bin are related to other variables. This figure indicates that titanium nitride levels in excess of 17% should be avoided.

As the volume of argon consumed at the rinse station increases, the steel is expected to be cleaner. From Figure 10 the incidence of line inclusions for argon volume in excess of 18 m^3 is very low.

The transition slab is the first slab of a new heat (that is a new ladle) and since there is open pouring, more defects are expected on these slabs. This is not the case from the evidence of Figures 11 and 12. For both the skin-lamination and line-inclusion defect more defects were associated with non-transition slabs. This may be due to interaction between variables.

Figure 13 shows the line-inclusion defect for tundish mass. The percentage at the 24.5 ton mass is slightly higher compared to the other tundish masses. This may however also be due to other factors since fewer defects are expected at high tundish volumes since the steel residence time is longer. The longer residence time should give improved inclusion removal.

The histograms for slab width are shown in figure 14 and 15. The histograms indicate that width has no influence on the incidence of either skin laminations or line inclusions.

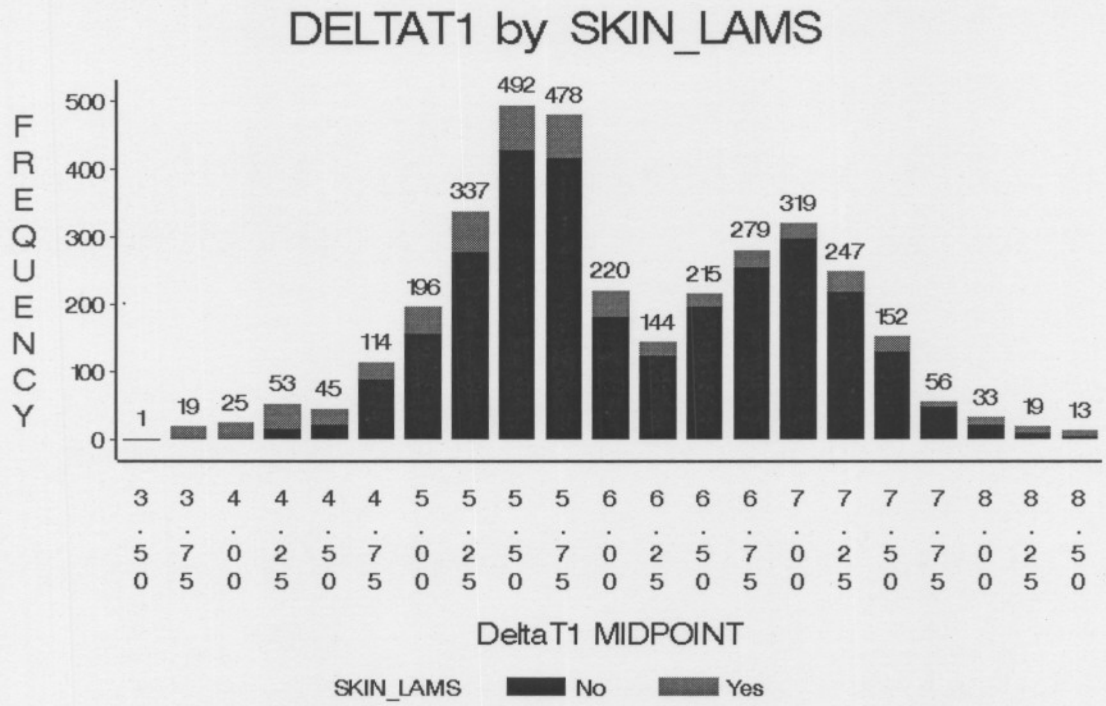


Figure 1 Mould narrow side heat removal and skin laminations

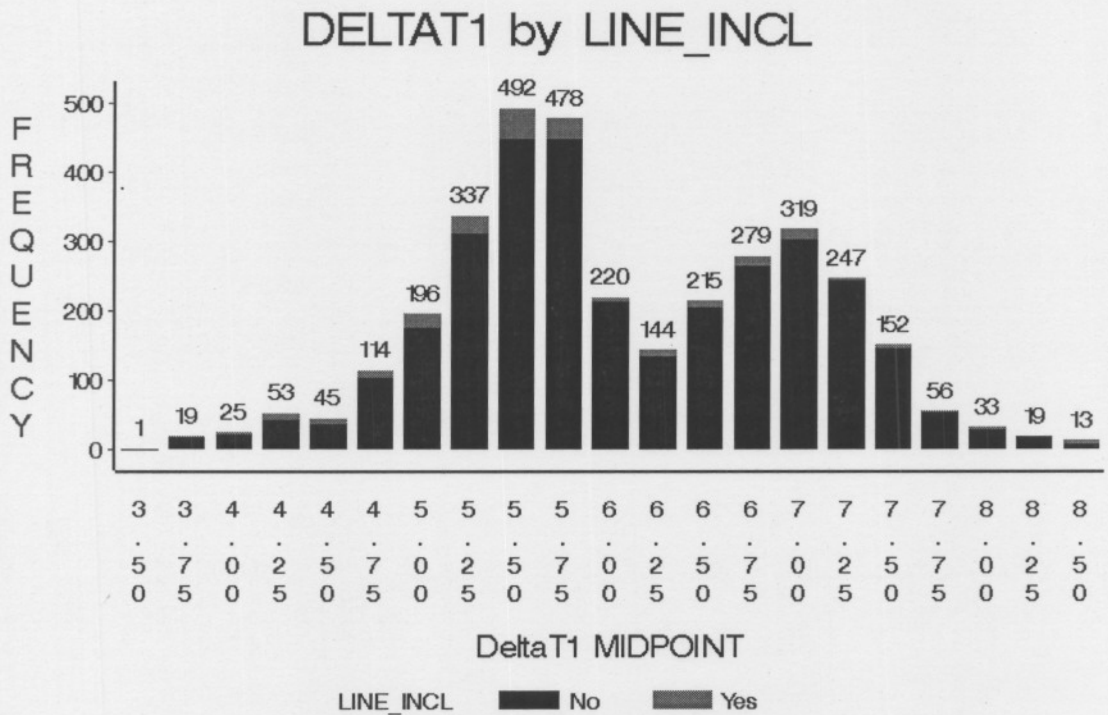


Figure 2 Mould narrow side heat removal and line inclusions

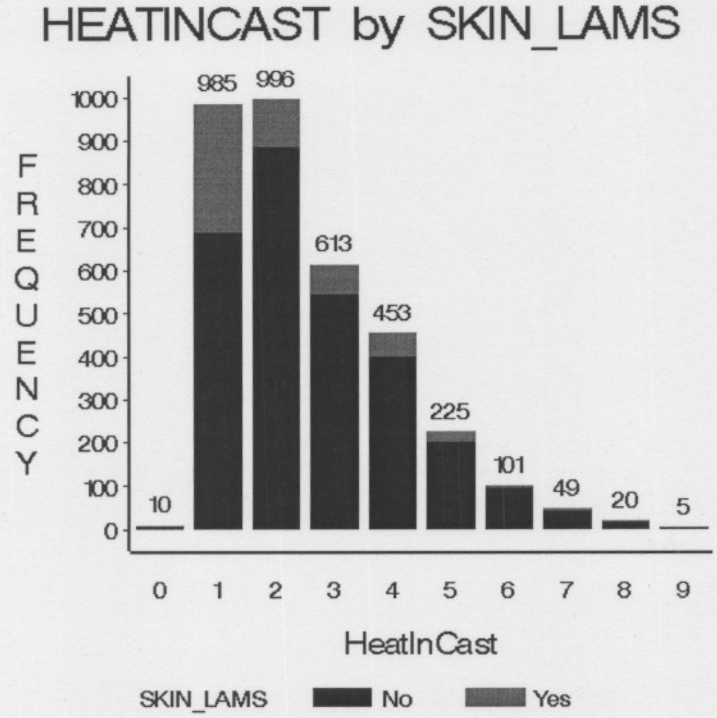


Figure 3 Heat in cast and skin lamination occurrence

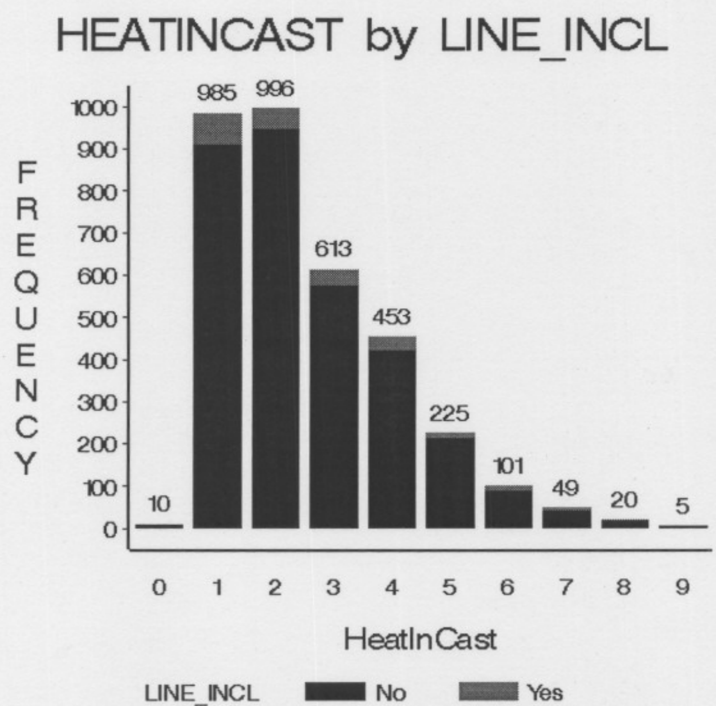


Figure 4 Heat in cast and line inclusion occurrence

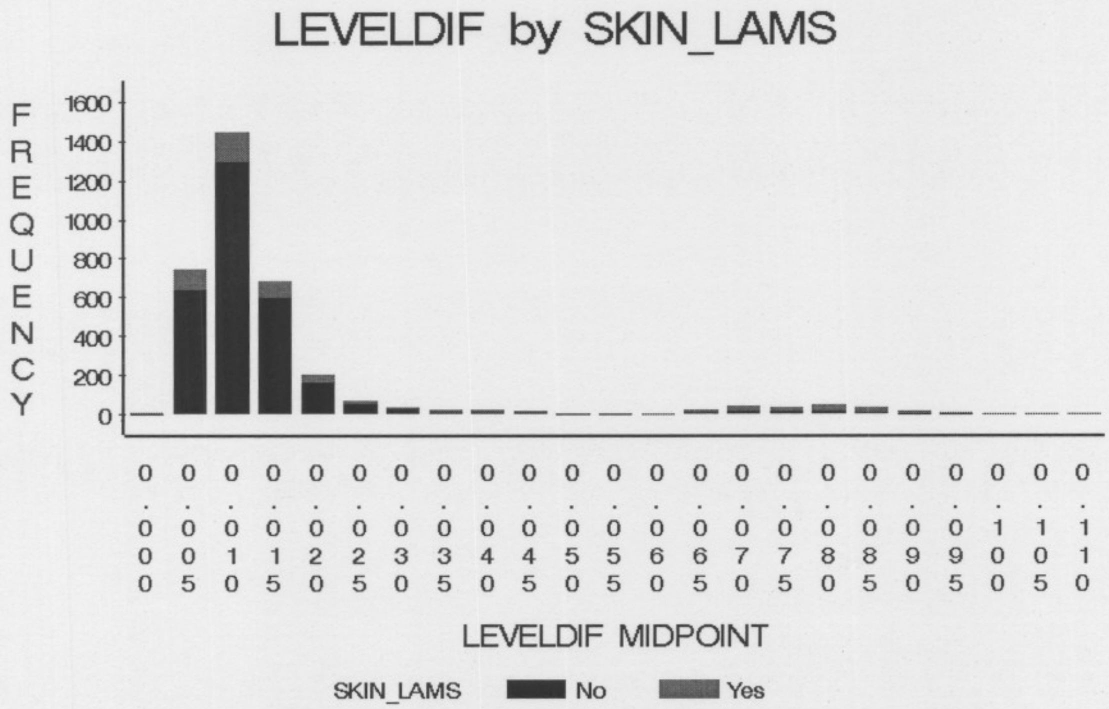


Figure 5 Mould level variation and skin lamination occurrence

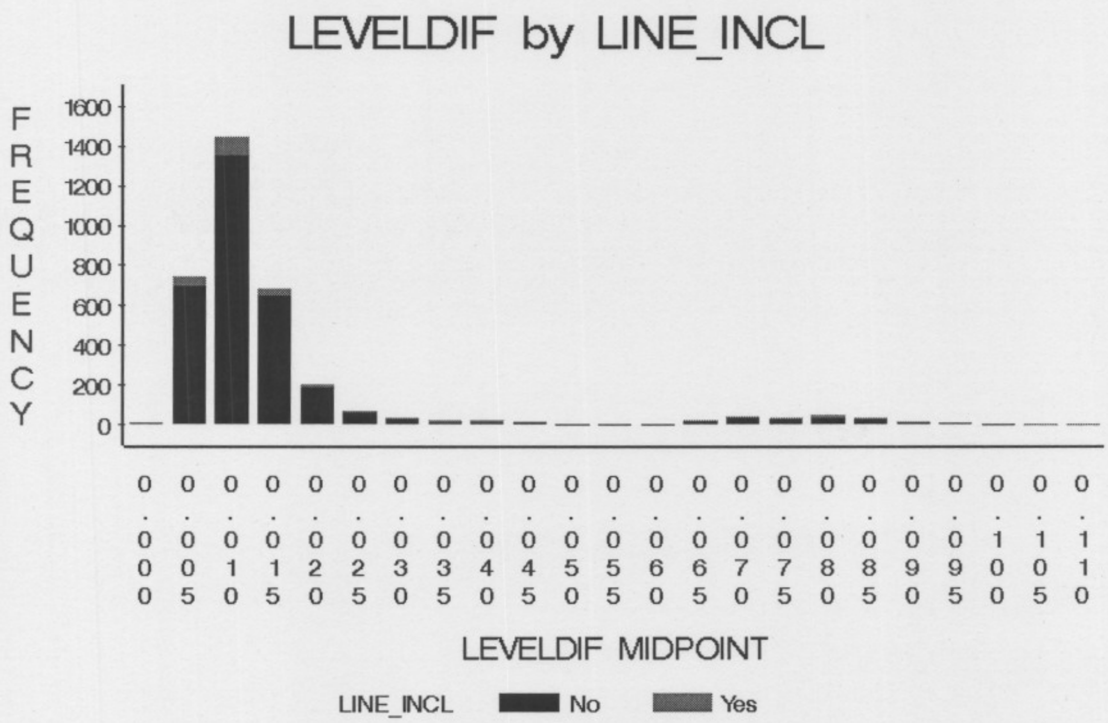


Figure 6 Mould level variation and line inclusion occurrence

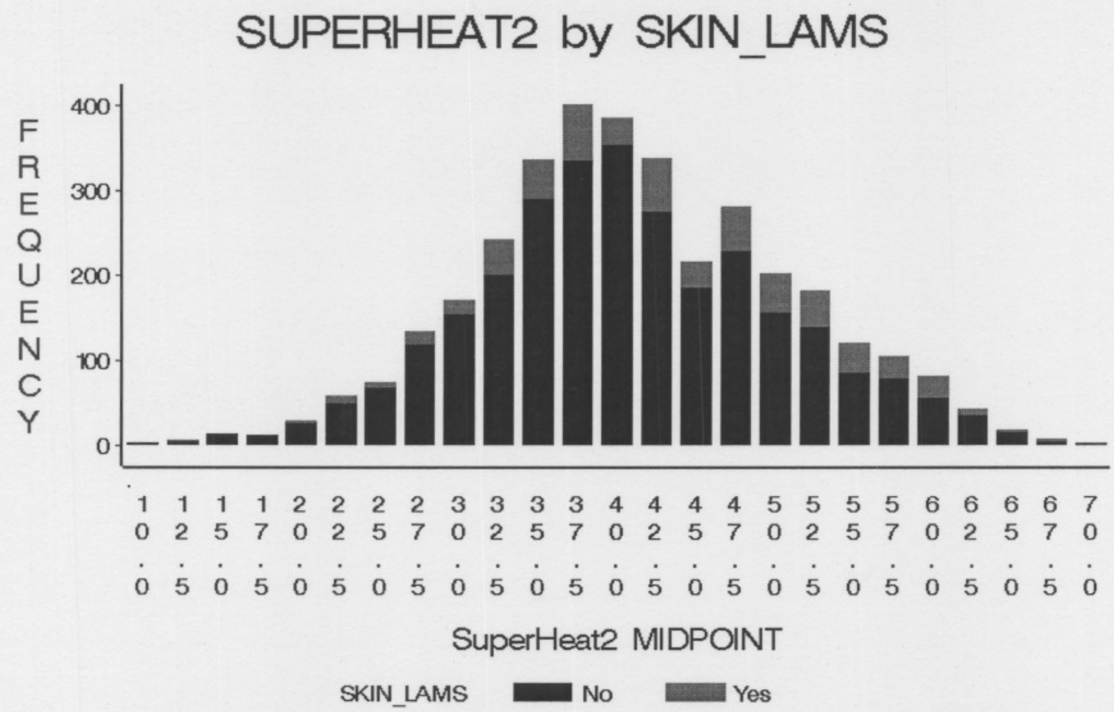


Figure 7 Superheat and skin lamination occurrence

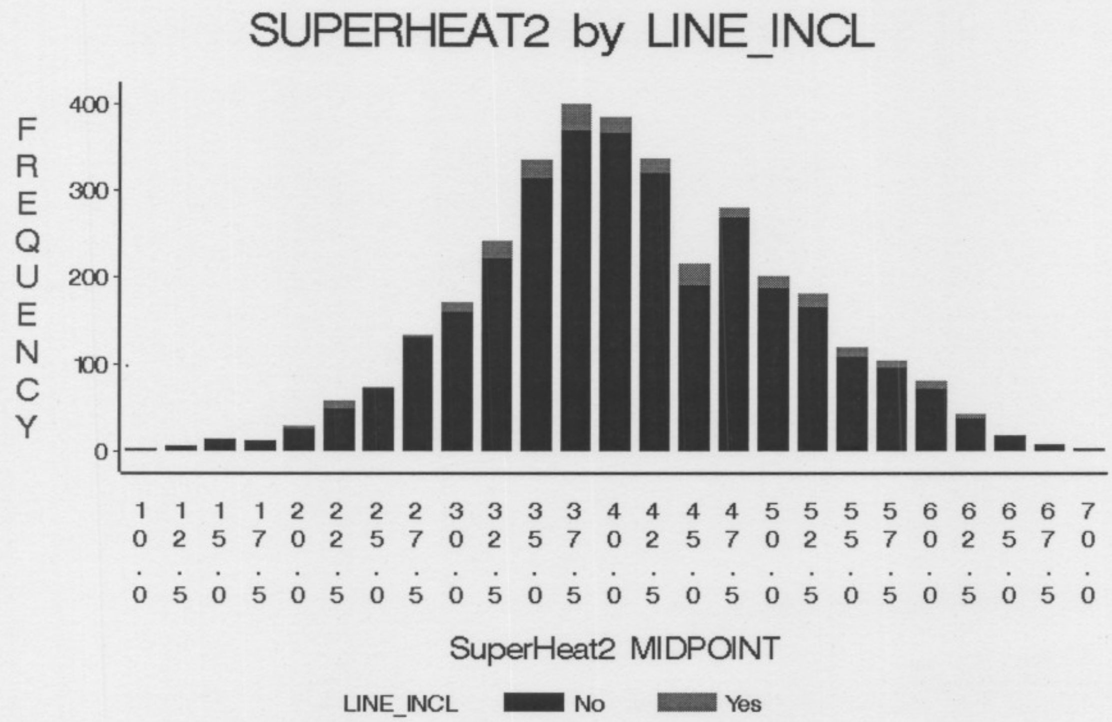


Figure 8 Superheat and line inclusion occurrence

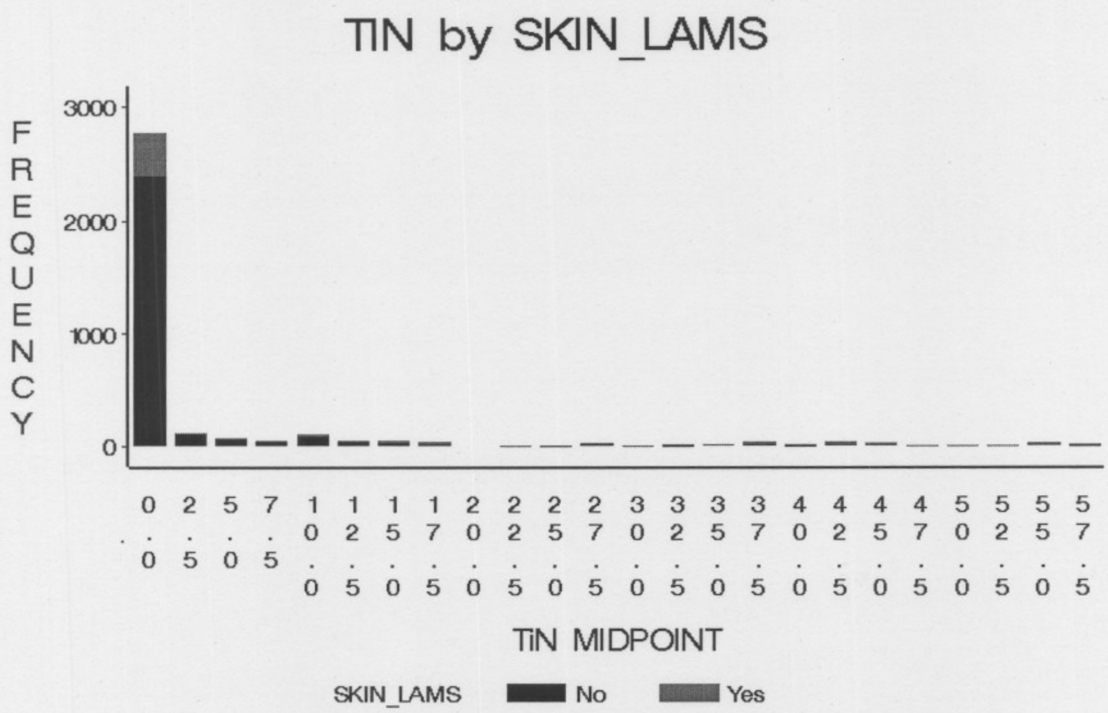


Figure 9 Titanium Nitride and skin lamination occurrence

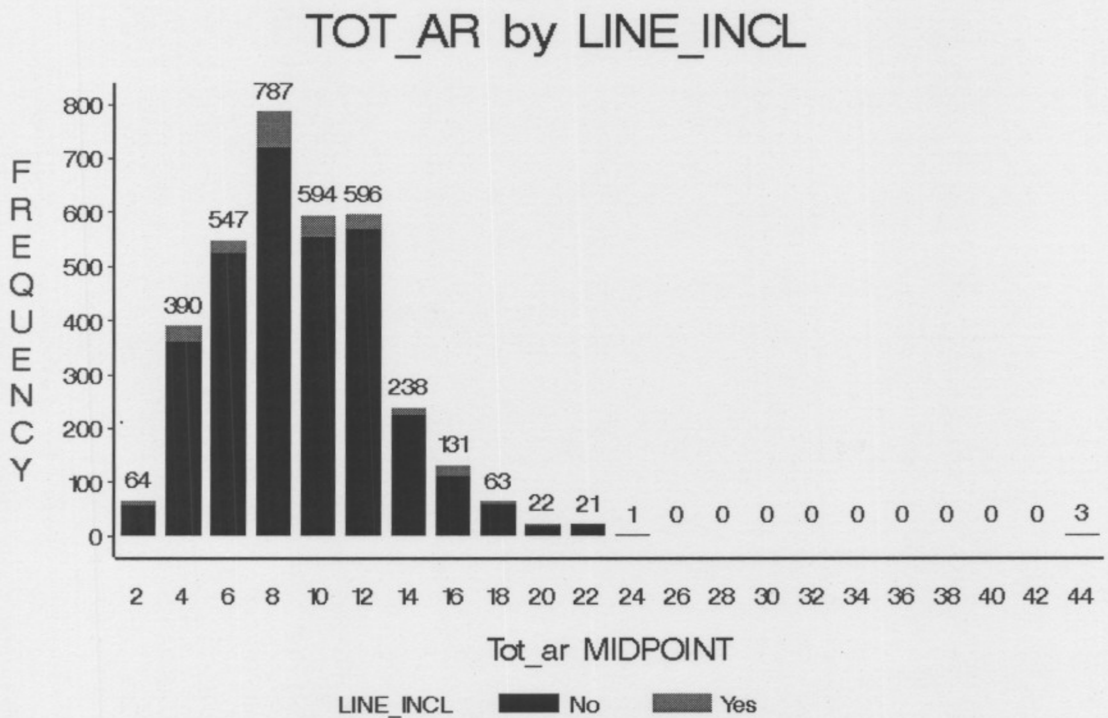


Figure 10 Total argon consumed at the rinse station and line inclusion occurrence

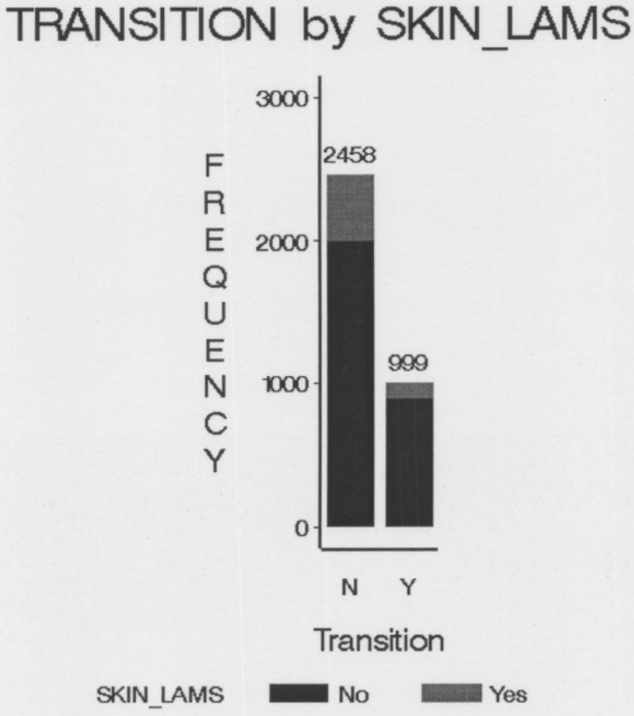


Figure 11 Transition slab and skin lamination occurrence

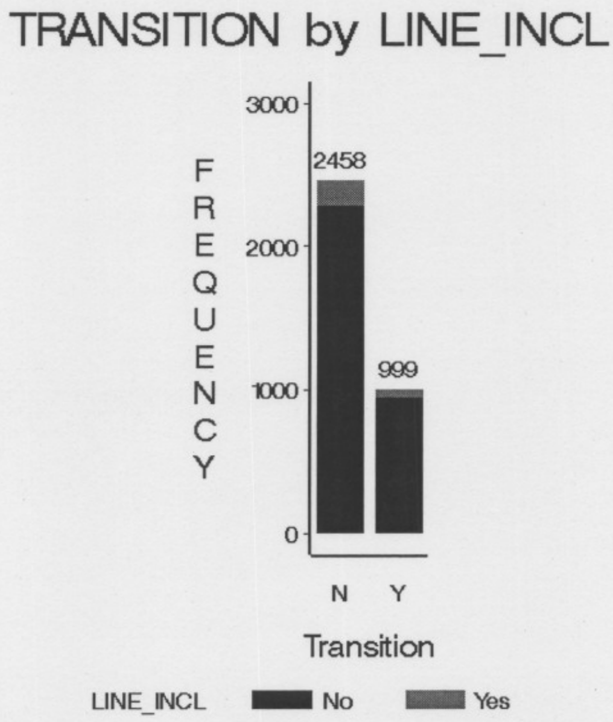


Figure 12 Transition slab and line inclusion occurrence

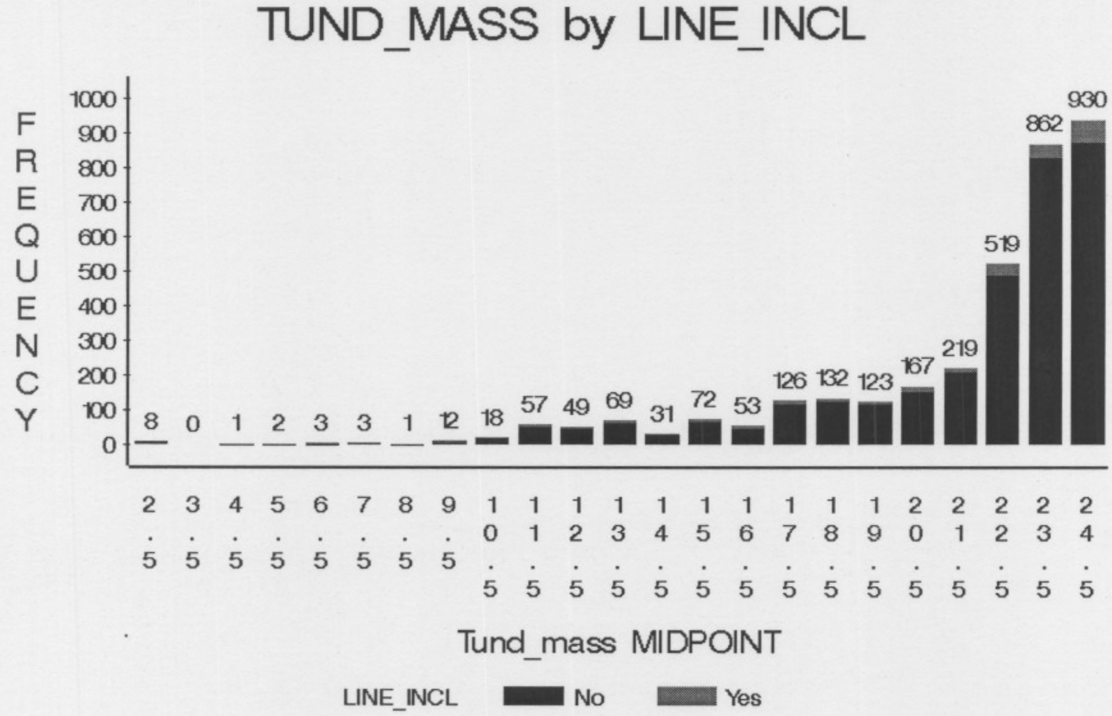


Figure 13 Tundish mass and line inclusion occurrence

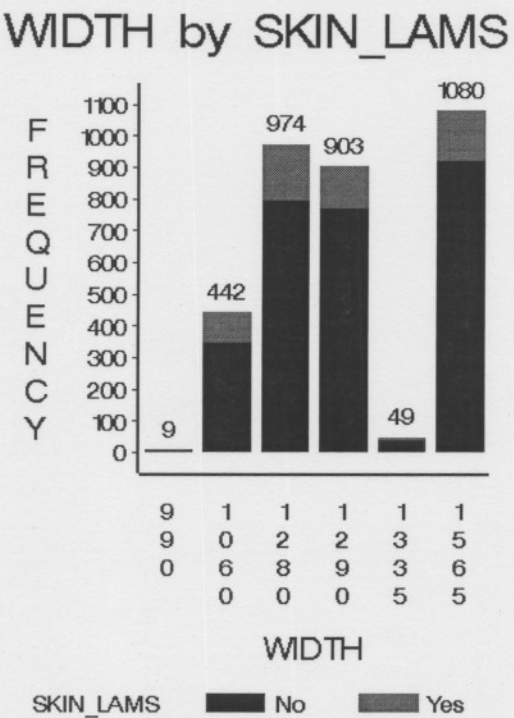


Figure 14 Slab width and skin lamination occurrence

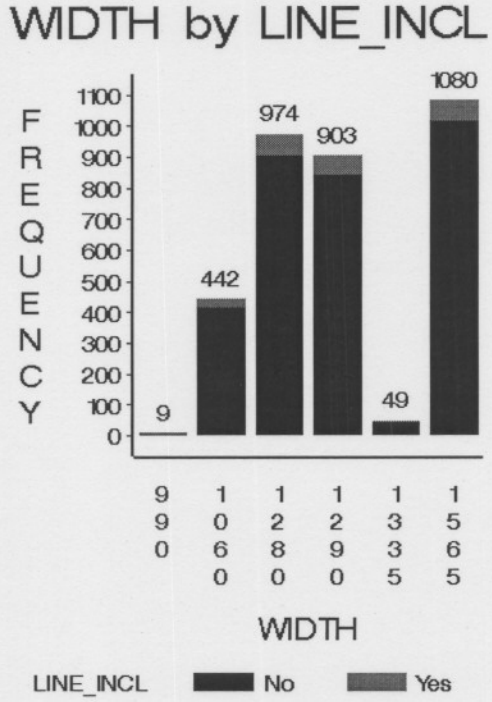


Figure 15 Slab width and line inclusion occurrence

APPENDIX C MODEL RESULTS

The average errors generated during the training of the skin-lamination neural network models are shown in Figures 1 to 4. This error is calculated between iterations during the training of the network for both the training and validation data set. These figures are for the networks with five, ten, fifteen and twenty neurons in the hidden layer respectively. On each figure the vertical line indicates the model with the minimum error for the validation data set. These models were used as the final model in each case.

In Figure 1 the error for the training set was 0.345 while the error for the validation set was 0.40 for the selected model. The error for the validation set is generally greater than that of the training data set since the training data is used to develop the model while the validation data is used to verify how good the model generalise.

The training graphs for the line-inclusion defect for the neural networks with five, ten, fifteen and twenty neurons in the hidden layer respectively, are shown in Figures 5 to 8.

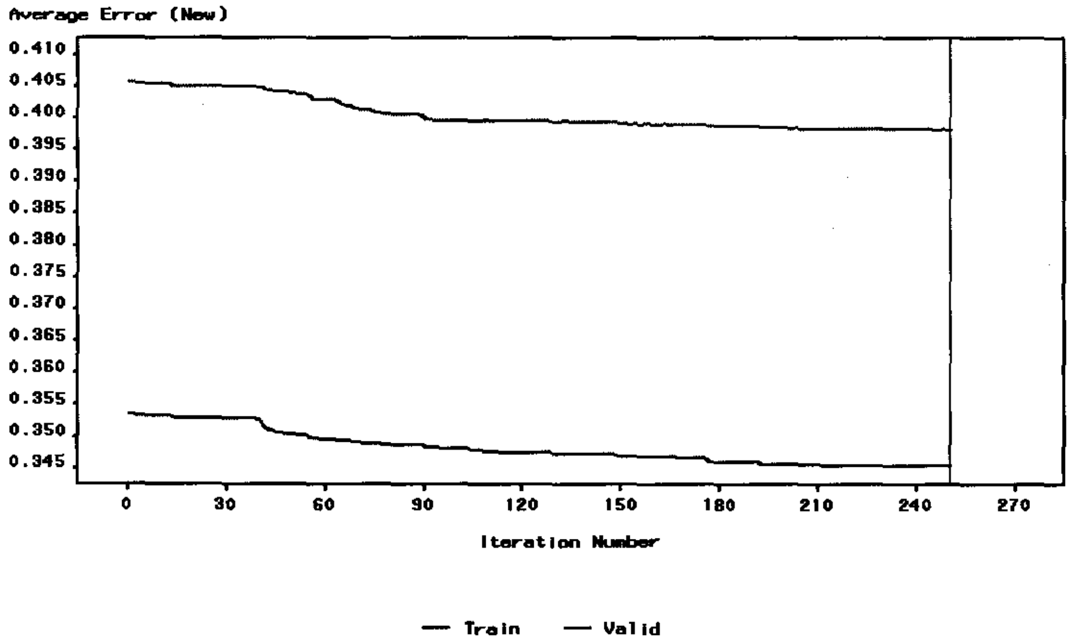


Figure 1 Skin-lamination training chart for neural network with five hidden neurons

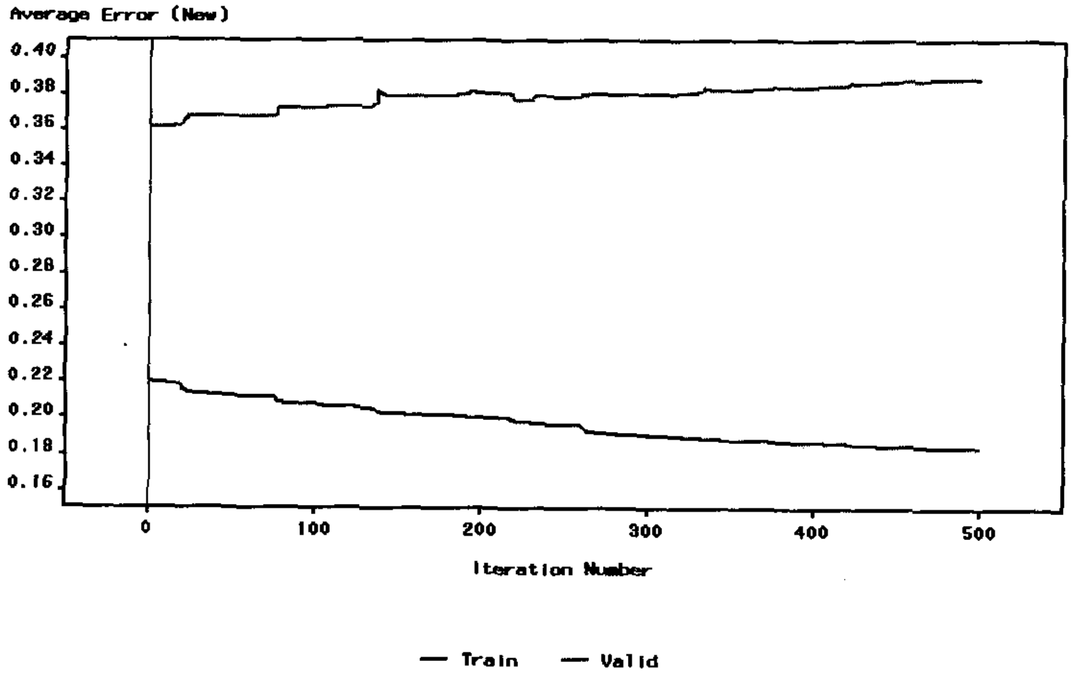


Figure 2 Skin-lamination training chart for neural network with ten hidden neurons

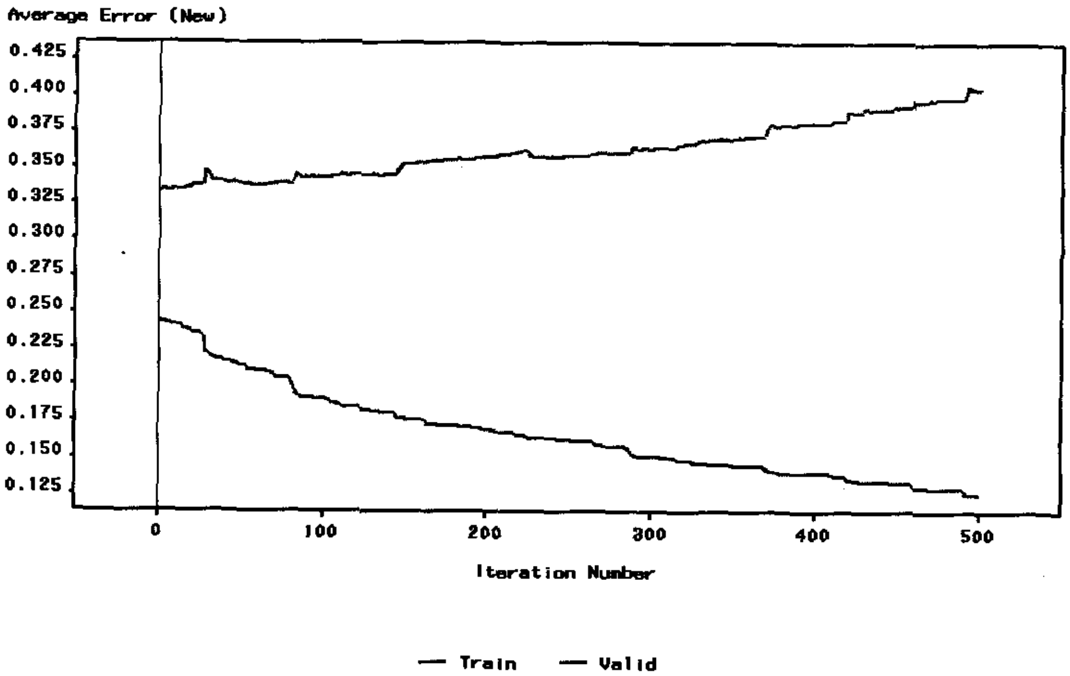


Figure 3 Skin-lamination training chart for neural network with fifteen hidden neurons

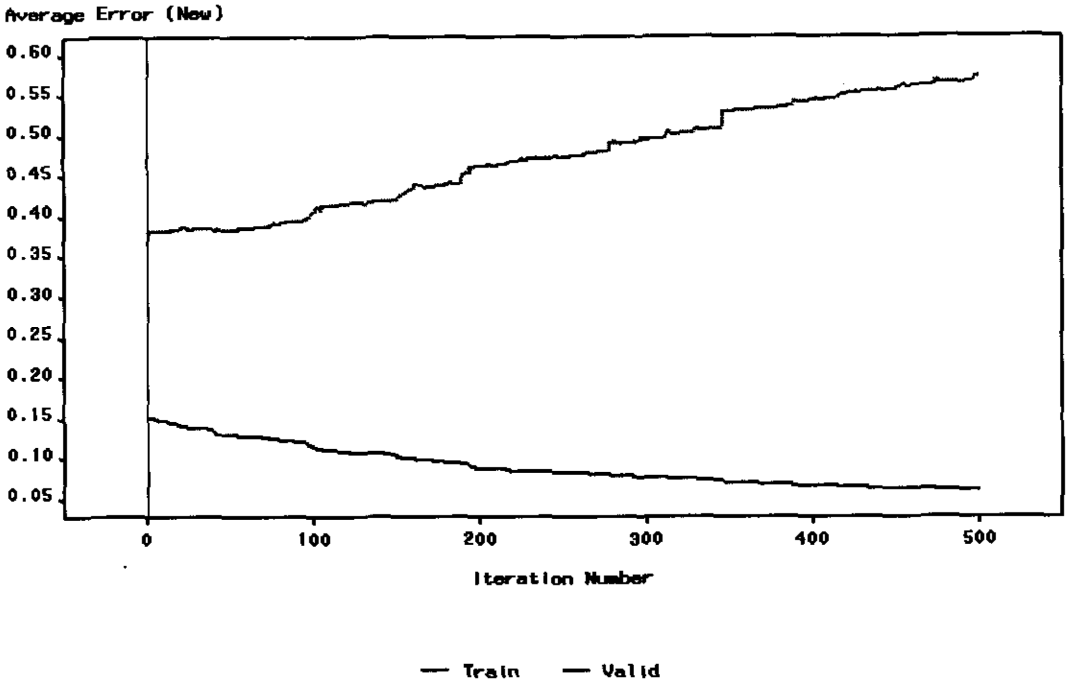


Figure 4 Skin-lamination training chart for neural network with twenty hidden neurons

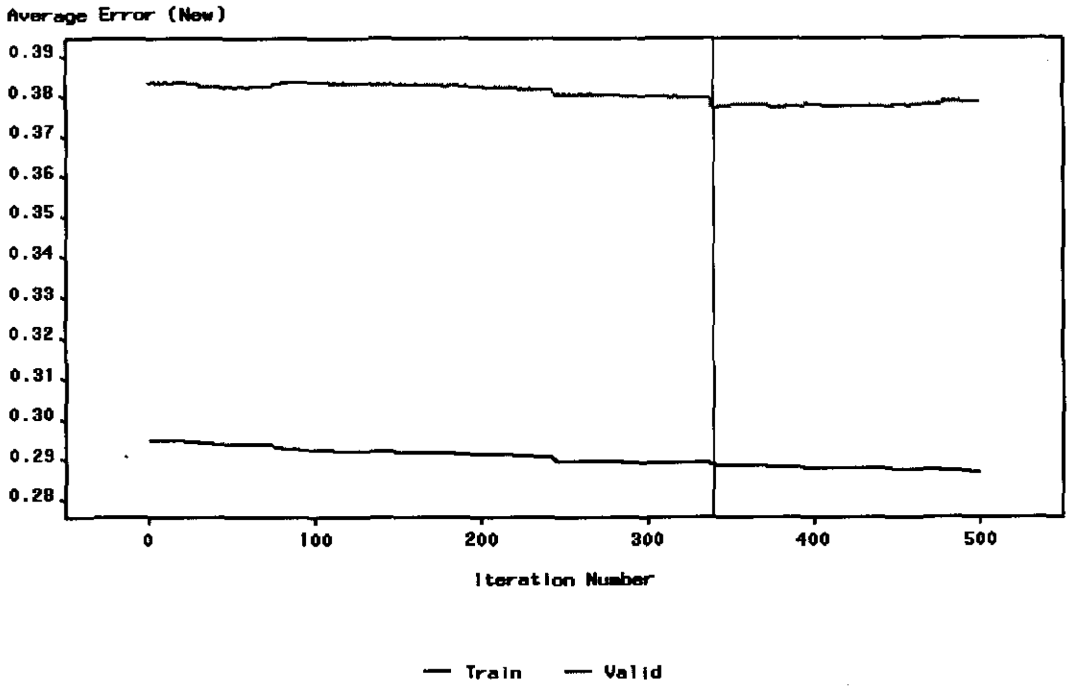


Figure 5 Line-inclusion training chart for neural network with five hidden neurons

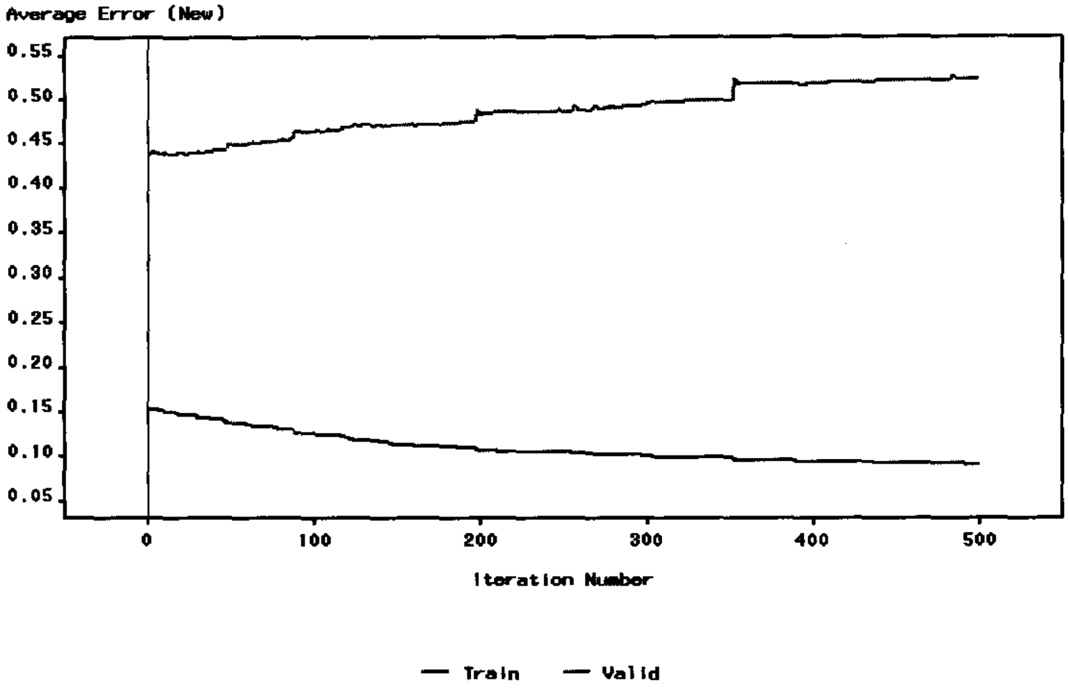


Figure 6 Line-inclusion training chart for neural network with ten hidden neurons

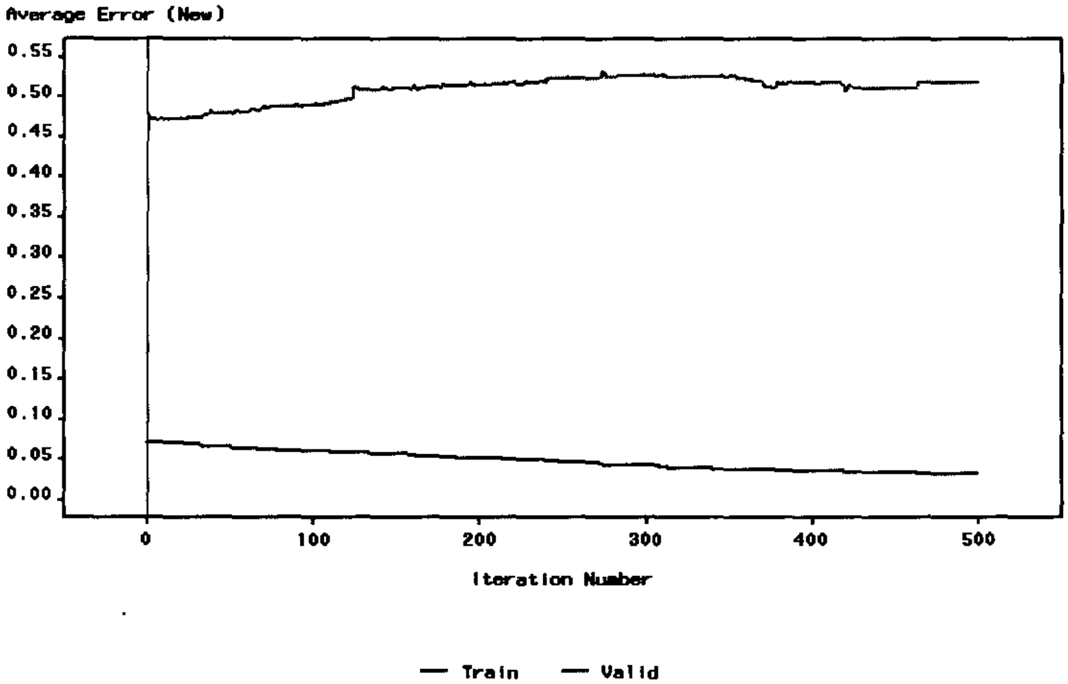


Figure 7 Line-inclusion training chart for neural network with fifteen hidden neurons

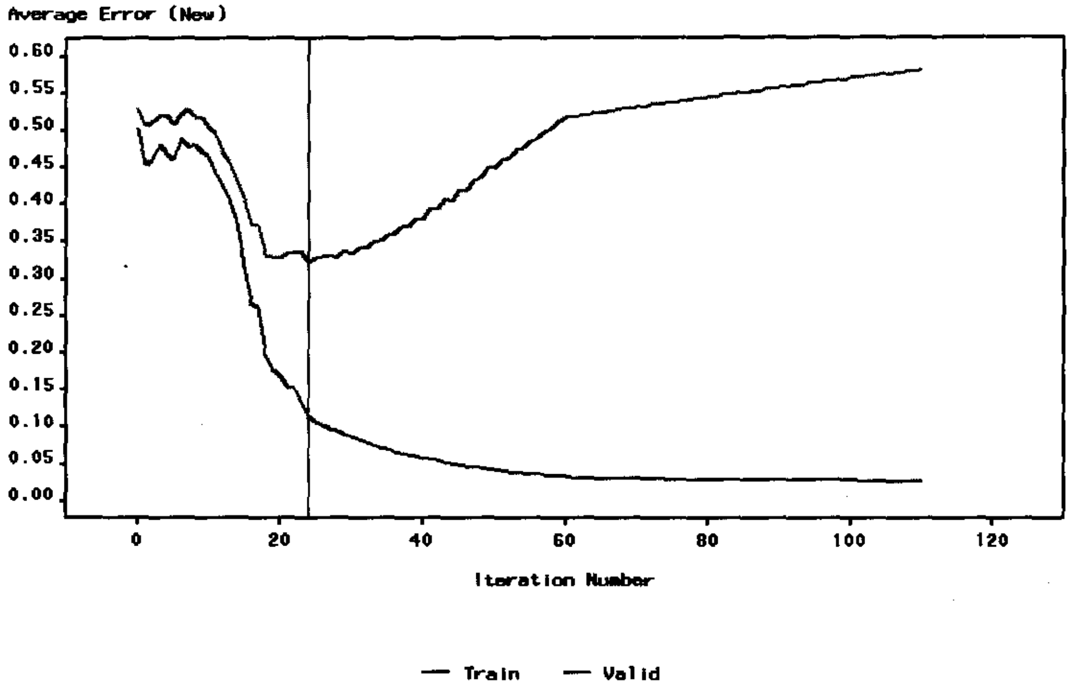


Figure 8 Line-inclusion training chart for neural network with twenty hidden neurons