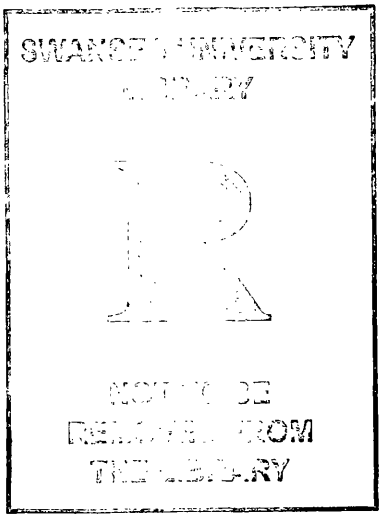
Hot mill process parameters impacting on hot mill tertiary scale formation.

Kennedy, Jonathan Ian

EPSRC Engineering Doctorate Centre on Steel Technology

Hot Mill Process Parameters Impacting on Hot Mill Tertiary Scale Formation



By

Jonathan Ian Kennedy

A thesis submitted to Swansea University
For the Degree of Doctor of Engineering

2012

Department of Materials Engineering
Swansea University

DECLARATION

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed ..... (candidate)

Date 17/07/13

STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated. Where correction services have been used, the extent and nature of the correction is clearly marked in a footnote(s).

Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed ..... (candidate)

Date 17/07/13

STATEMENT 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed ..... (candidate)

Date 17/07/13

THESIS SUMMARY

Candidate's Surname: Kennedy

Candidate's Forenames: Jonathan Ian

Candidate for the Degree of: Engineering Doctorate in Steel Technology

Full title of thesis: Hot Mill Process Parameters Impacting on Hot Mill Tertiary Scale Formation

---

Summary:

For high end steel applications surface quality is paramount to deliver a suitable product. A major cause of surface quality issues is from the formation of tertiary scale. The scale formation depends on numerous factors such as thermo-mechanical processing routes, chemical composition, thickness and rolls used. This thesis utilises a collection of data mining techniques to better understand the influence of Hot Mill process parameters on scale formation at Port Talbot Hot Strip Mill in South Wales.

The dataset to which these data mining techniques were applied was carefully chosen to reduce process variation. There are several main factors that were considered to minimise this variability including time period, grade and gauge investigated. The following data mining techniques were chosen to investigate this dataset:

- Partial Least Squares (PLS)
- Logit Analysis
- Principle Component Analysis (PCA)
- Multinomial Logistical Regression (MLR)
- Adaptive Neuro Inference Fuzzy Systems (ANFIS)

The analysis indicated that the most significant variable for scale formation is the temperature entering the finishing mill. If the temperature is controlled on entering the finishing mill scale will not be formed. Values greater than 1070 °C for the average Roughing Mill and above 1050 °C for the average Crop Shear temperature are considered high, with values greater than this increasing the chance of scale formation. As the temperature increases more scale suppression measures are required to limit scale formation, with high temperatures more

likely to generate a greater amount of scale even with fully functional scale suppression systems in place.

Chemistry is also a significant factor in scale formation, with Phosphorus being the most significant of the chemistry variables. It is recommended that the chemistry specification for Phosphorus be limited to a maximum value of 0.015 % rather than 0.020 % to limit scale formation. Slabs with higher values should be treated with particular care when being processed through the Hot Mill to limit scale formation.

Acknowledgments

This work would not have been possible if not for the assistance of a number of key people and I would like to take a moment to acknowledge their contribution to this study.

First I would like to acknowledge the invaluable assistance of Dr Fiona Robinson. Her depth of knowledge and contacts proved invaluable in running a successful project.

Dr Mark Evans for his advice and guidance on writing papers and interpreting data.

Mr Sid Petitte, whose decades of experience in the steel industry were invaluable to understanding which process parameters were important in influencing scale formation in Port Talbot Hot Mill.

Dr John Albiston for giving me advice on suitable software available within Tata steel for the project.

Dr Paul Thomas, for taking over the industrial supervisor position in the final year and aiding me in structuring my thesis.

Dr James Sullivan for his support in the early stages of the project.

I would also like to thank from the bottom of my heart Natasha Marshall, for being so supportive through the closing months of the write up. Also, for being the best person I know.

This research was funded by a grant from the Physical and Sciences Research Council (EPSRC) and TATA Steel Strip Products, UK (TSSP-UK).

And last (but not least) my family who have offered moral support and advice throughout the doctorate.

List of Publications

The following is a list of publications I have written or been involved in during my doctorate.

Kennedy, J., Evans, M., Robinson, F., Identification for control of the process parameters influencing tertiary scale formation at the hot strip mill using a binary choice model, *Journal of Materials Processing Technology, Volume 212, Issue 7, July 2012, Pages 1622-1630*

Evans, M., Kennedy, J., Thomas, P., Process parameters influencing tertiary scale formation at a hot strip mill using a multinomial logit model, *Journal of Manufacturing Science and Engineering*

Evans, M., Kennedy, J., Thomas, P., Non-linear partial least squares and its application to minimising scale formation at a steel hot mill, *Materials Science and Technology,* Volume 28, Number 12, December 2012, pp. 1513-1522(10)

Bright, G., Kennedy, J., Robinson, F., Evans, M., Whittaker, M.T., Sullivan, J., Gao Y.,Variability in the mechanical properties and processing conditions of a High Strength Low Alloy steel, *Procedia Engineering, Volume 10, 2011, Pages 106-111*

Sullivan, J., Weirman, C., Kennedy, J., Penney, D., Influence of steel gauge on the microstructure and corrosion performance of zinc alloy coated steels, Corrosion Science, Volume 52, Issue 5, May 2010, Pages 1853-1862

Contents

List of Tables

List of Figures

Chapter 1. Literature Review

1.1 Processes

This section will introduce the different processes that occur in the Hot Mill and will concentrate on the areas that have the greatest influence over scale formation.

1.1.1 Hot Mill Overview

The steel works in Port Talbot is a fully integrated steel works, producing steel from raw materials with limited scrap additions. Steel is continuously cast at Port Talbot using a continuous caster.

After being processed through the continuous caster the material is transported to the Hot Mill (see figure 1) for further processing. The purpose of the Hot Mill is to reduce the gauge of the slab to form a hot rolled coil that can either be dispatched directly to a customer or proceed for further processing. The aspects of the processing that have an effect on upstream surface quality are investigated. The Hot Mill is split into several sections including;

- Slab Yard
- Reheat furnace
- Scale breaker
- Reversing roughing mill
- Coil box
- Crop shear
- Finishing mill
- Run Out Table
- Coiler

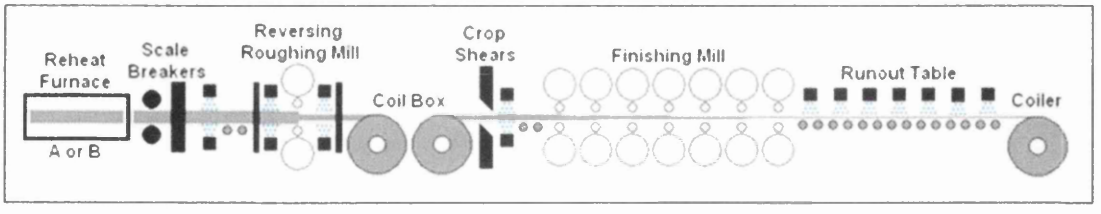


Figure 1. Overview of the Hot Mill.

1.1.2 Slab Yard

There are several purposes of the slab yard, including the following;

- To receive slabs from the continuous caster
- To store slabs before they are processed
- To provide the slabs in the correct order for the Reheat Furnace
- To identify and rectify quality issues

When a slab arrives in the slab yard one of two things will happen;

1. If the slab is part of a hot connect round it will be sent straight to the Hot Mill for processing. Hot connect slabs have several advantages including;
    a. Saving money on reheating because they start at a higher temperature when entering the Furnace.
    b. Reduce time between placed order and delivery.
    c. Saving space. Under normal conditions a slab would sit in the slab yard for several days to cool down before being processed for safety reasons.
2. Most slabs will be left to cool for several days so they can be checked and put into a round for processing in the Hot Mill.

1.1.2.1 Slab Defects

It is vital that the slabs dimensions are tightly controlled before entering the Hot Mill (see table 1). The Port Talbot reheat furnace is a walking beam furnace. This means that there are a number of stationary beams to support the slab and a number of powered beams to move the slab through the furnace. If the processed slabs are too long for the Hot Mill it can damage the Furnace and if it is too short it could get stuck in the Furnace. The dimensions are dependent on which mill the slab is going to be rolled at. Slabs that are too short (< 7 m) to be processed at Port Talbot can be processed via Llanwern Hot Mill rather than be scrapped.

Table 1. Slab Specifications (Dimensions).

	Length (m)	Width (mm)	Thickness (mm)
Port Talbot	7.0 – 10.0	640 – 1875	190 – 254
Llanwern	5.5 – 8.75	700 – 1550	190 – 254

The shape of the bar is also very important, because poor shape could lead to difficulties in charging and reheating in the Furnaces. Or in the worst case the slabs are scrapped to protect the Furnace and Hot Mill.

Table 2. Slab Specifications (Bending).

	Max. bend off horizontal (mm)	Max. bend off straight (single width) (mm)	Max. bend off straight (slit width) (mm)
Port Talbot	70	25	50
Llanwern	70	25	50

To stop slabs from bending during cooling, stacks should be arranged so that the longest slabs are on the bottom or give a maximum overhang of 1 metre. If the slabs deviate by more than the values in table 2 then the slab will be placed between 2 hot slabs. This is undertaken to bend the slab back to the correct shape. If this is not possible then the slab can be cut back until the bend is below 70 mm out of flatness. However, care should be taken to make sure that slabs do not drop below their minimum specification for processing (see table 1).

1.1.2.2 Round

The round is the order to which the slab will be processed through the Hot Mill between planned roll changes. Rounds attempt to maximize the surface and dimensional quality whilst reducing the number of work roll changes in the rolling mills to maximize profit.

The slabs in the round have rules governing how they can be positioned in the round depending on width, gauge and grade considerations. The slabs planned for the round will be checked for quality issues prior to being ordered and sent to the Furnace.

1.1.3 Furnace

After leaving the slab yard, slabs are placed on the slab charging roller tables and will be given final visual checks of their dimensions prior to entering the Furnace. The slabs are

moved to the charge end of the Furnace and enter the Furnace when appropriate space is available in Furnace A or B.

The Furnaces are long insulated rectangular metal boxes, with burners to control temperature. The slabs are moved through the Furnace using the walking beam. The Furnace contains a mixture of fixed beams (to hold the slab) and walking beams (to move the slab). The walking beams are lifted using hydraulic cylinders and rollers. The most important aspect of slab Furnace movement is that the slabs are moved in a controlled way. If control over the slab is lost then slabs could interfere with each other, damage the Furnace wall, or in extreme cases get stuck in the Furnace.

Slabs are heated to the required temperature (1180 – 1240 °C) depending on the dimensions and chemistry. The temperature of the slab is not measured in the Furnace due to the thick layer of scale preventing accurate measurements from being taken. Therefore the Furnace temperature is back calculated from the Roughing Mill after the Furnace scale has been removed and an accurate reading can be taken.

1.1.4 Scale Breaker and Reversing Rougher Mill

After leaving the Reheat Furnace the slab proceeds into the scale breaker which will remove the scale created in the Furnace. Without a scale breaker there would be a dramatic increase in the number of primary scale defects on the strip, and would also impact on surface quality by damaging rolls.

The Roughing Mill reduces the gauge of the incoming slab to between 35 mm and 45 mm, which is known as a transfer bar. The transfer bar is spread or squeezed by the edger to the required width of the coil and some of the scale formed in the Reheat Furnaces is removed. The Hot Mill in Port Talbot has a Reversing Rougher hence several passes are necessary to achieve the required transfer bar thickness.

1.1.5 Coil Box

After the Reversing Rougher Mill the slab is then held at the coil box, the purpose of which is to reduce the size of the mill and to increase the temperature of the transfer bar before it

enters the Finishing Mill. When the coiling is complete the outer lap of the coil is fed into the Crop Shears. After the coil box the top/bottom and head/tail ends are switched.

1.1.6 Finishing Mill

The purpose of the Finishing Mill (FM) is to reduce the dimensions of the transfer bar to the thickness that is required by the customer. This is achieved by passing the transfer bar through a series of close-coupled mill stands that have a reducing roll gap.

1.1.7 Finishing Mill Entry

The Finishing Mill is a vital area of the Hot Mill when considering tertiary scale formation (also called D-type scale). An overview of the Finishing Mill entry is in figure 2. The purpose of the entire scale suppression system in the front section of the Finishing Mill is to condition the transfer bar so that scale is not formed.

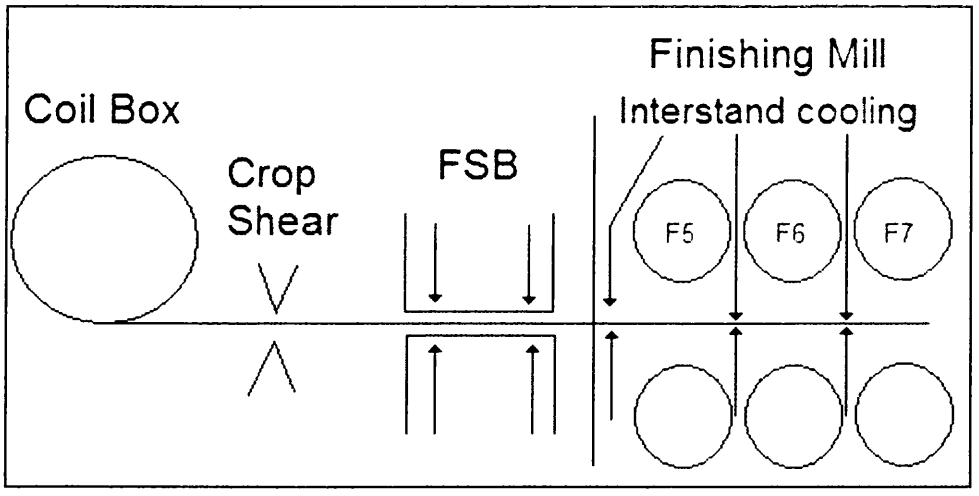


Figure 2. Finishing Mill Entry.

1.1.7.1 Crop Shear

After leaving the coil box the transfer bar enters the Crop Shear. The Crop Shear is a large shear spanning the width of the transfer bar, which will crop the end of the bar to leave a straight edge. The purpose of the Crop Shear is to ensure that the transfer bar is fed correctly into the Finishing Mill by cutting off the end of the bar so a flat edge is left.

1.1.7.2 Roll Coolant

Roll coolant is added to the rolls in the Finishing Mill, for several important reasons including;

1. To protect the work rolls from thermal failure due to the high temperatures and loads associated with rolling.
2. To control the thermal expansion (thermal crown) of the work rolls in a parabolic manner so that a parabolic strip crown and flatness are obtained.
3. To minimise the surface degradation of the work rolls. This is achieved by maintaining a stable protective oxide layer on the surface of the rolls in the early stands to minimise roll wear.

The incorrect use and control of the roll coolant can result in a number of failure modes including;

- Too much or too little water affecting roll gap profile - poor interstand shape leading to manual bending interventions - target exit shape and/or crown/profile not obtained.
- Uneven application of cooling water across the roll barrel length – asymmetric shape or a ridge on the profile.
- Inadequate volume of roll coolant - surface temperature of the roll gets too hot during rolling generating scale on the roll which can imprint on the strip.

1.1.7.3 Finishing Scale Breaker (FSB)

After the Crop Shears the transfer bar passes through the conditioning systems of the Finishing Scale Breakers (FSB). The FSB are used to remove secondary scale formed after the Roughing Mill and condition the transfer bar to prevent tertiary scale formation. This is achieved by using high pressure water from the descaling system. The FSB consists of the following:

- Two top and two bottom descaling headers, referred to as A and B legs.
- A descaling hood
    - Collects and channels rebound water away from the bar.
- FSB top and bottom pinch roll - situated at the entry to the FSB

    - These provide traction through the FSB. The top and bottom rolls need to make good contact with the transfer bar to prevent water from running back to the Crop Shear.
        - Reducing temperature at the Crop Shear will increase the resistance to shearing and increase wear on the Crop Shear.
- Damming roll - situated at the exit of the FSB
    - The top roll is fitted with a compressed air supply thus provide a downward force to ensure good contact with the strip. This prevents excess water from pooling on the top surface in the F5 guide area.
- A centrally located under bar driven roller – maintains the pass-line through the FSB.
- External deflector
    - Situated between the pinch roll and the first descaling header. This is a series of narrow steel fingers attached to a horizontal pole again preventing excess water escaping.
- A set of carry-over fingers.

The two pairs of headers (top and bottom), are all equipped with nozzles at approximately 100 mm apart. The second pair is offset by approximately 50 mm relative to the first, this evens out the cooling effects and to give full width descaling coverage should a nozzle in one of the headers become blocked. Each header is split into a central zone (approximately 1400 mm wide); this is used for the majority of material processed. For material with a width greater than 1400 mm a smaller outer section of the headers is used (approximately 300 mm at each end). Water reserves need to be carefully managed for high width material to avoid significant loss of pressure and flow.

For full width coverage, it is important that the nozzles are aligned correctly to each other. The correct descaling coverage is only achieved when the nozzles are aligned correctly allowing full overlap of sprays. This type of installation is in danger of not providing full descaling and therefore increasing the risk of defects such as scale.

To ensure that conditioning is achieved each nozzle is set at an angle to the rolling direction. This is done to ensure that the spray pattern from one nozzle does not interfere with the

adjacent nozzle. The nozzle angle is also important because some angles are more likely to suffer from clogging of the nozzle.

It should be noted that if there is excessive undulation of the transfer bar as it passes through the FSB, then the change in relative heights/distances will give a different descaling footprint. Incomplete descaling coverage is possible on the top surface as the standoff distance decreases. On the bottom surface, the standoff distance increases and more nozzle overlap results. This is another reason why the headers are offset by 50 mm.

As the nozzles are worn down through use the flow rate will increase, this will produce the following effects;

- The descaling impact pressure will drop.
    - This will reduce the amount of scale removal.
- The increased flow rates will cause additional cooling.
    - Additional cooling could aid in conditioning of the transfer bar but can cause other defects such as transverse cuts plus possibly low finishing temperatures.
- The increased overlap of the water jets will result in additional "double" cooling in the overlap area which can be seen as dark stripes on the surface exiting the FSB. This is detrimental for transverse temperature distribution.

In order to control and minimize the effects of nozzle wear, maximum flow limits are set for initiating nozzle changes.

During maintenance, a lead test (see figure 3) is undertaken as a check for blocked or misaligned nozzles and also to ensure that the carryover fingers do not interfere with the descaling spray pattern. The lead test is a board covered on both sides with lead at the same thickness as the transfer bar (45 mm) spanning the full width of the FSB. The FSB is turned on and then the lead blank is examined to identify misaligned or blocked nozzles. The test should show that the sprays overlap with each other so that the entire width is covered.

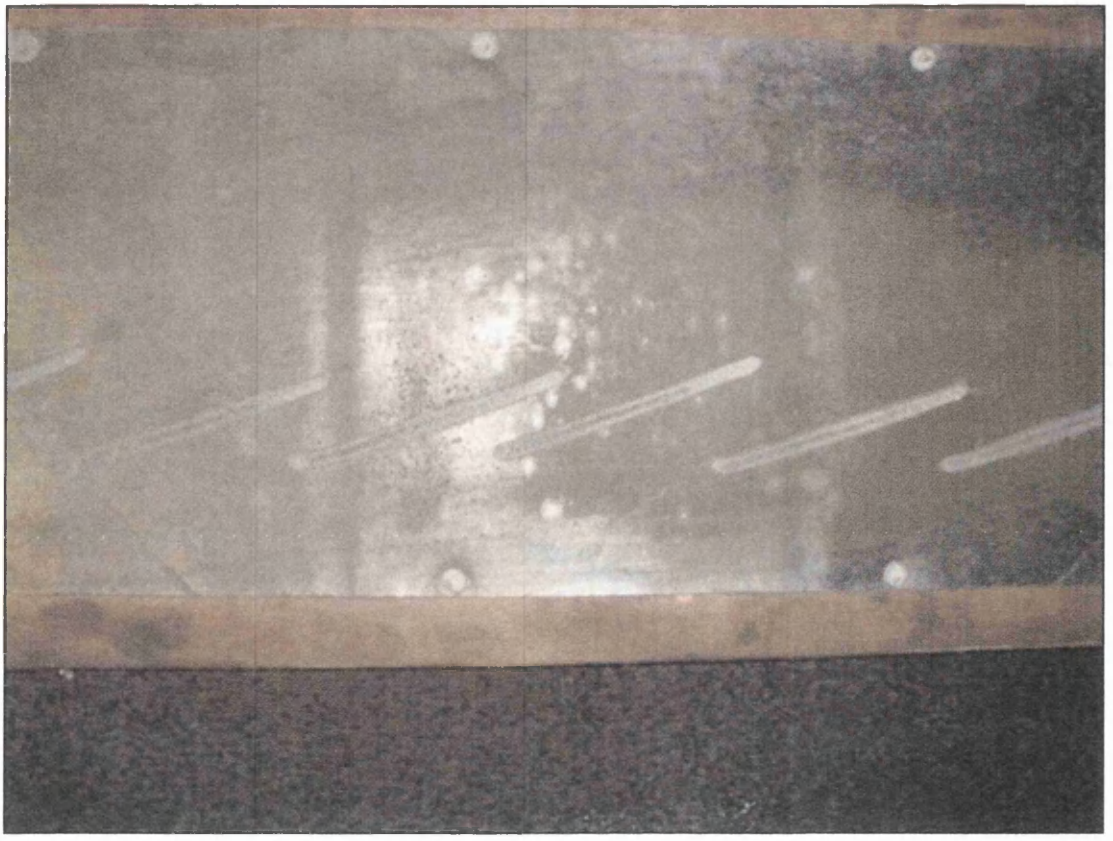


Figure 3. Lead Test. To test the effectiveness of the sprays a wooden plank covered in lead is inserted and the sprays are activated. The lines imprinted on the lead are the impact of the sprays. It is desirable to observe an overlapping pattern.

1.1.7.4 Water Containment

The FSB headers have various pieces of equipment on either side of them preventing the escape of excess descaling water. The presence of excess water pooling on the transfer bar in this area can cause:

- An overall loss of temperature.
- Localised temperature variation resulting in gauge spikes.
    - Different temperatures will affect how the transfer bar will react when it enters the Finishing Mill and give different gauges.
- Increases the risk of transverse cuts.
    - Transverse cuts are caused by localized areas of cooling. These areas will have different mechanical properties than the surrounding area and during further

processing will damage the surface which is unsuitable for many applications and can split the coil.

The purpose of the FSB hood is not only to cover the area but to stop descaling rebound water from falling onto the strip which will impede the impact of the sprays. To ensure descaling occurs water is contained within the hood. The hood contains channels to collect the rebound water and then allow it to flow away at the operator and drive side ends of the hood. During the rewinding of a cobble, the hood should be raised to prevent internal damage.

1.1.7.5 F5 Pre-Sprays

After the FSB there are the F5 pre-sprays (see figure 2), the purpose of these are to control the temperature of the transfer bar to minimize tertiary scale formation. The conditioning of the transfer bar by the F5 pre-sprays can influence a number of failures including the following;

- Excess water applied to the strip. This can result in the following;
    - Incorrect gauge.
        - Excess water can change the cooling pattern of the steel which will influence the steels reaction to the Finishing Mill rolls leading to an incorrect gauge.
    - Low Finishing Mill exit temperature.
        - Due to increased cooling mechanical property failure can occur.
    - T-cuts.
        - Caused by localised cooling.
    - Head or tail end pinching.
        - Caused by excess water added to the strip.
- Insufficient water applied to the strip, this can result in the following;
    - High Finishing Mill exit temperature.
        - Due to decreased cooling mechanical property failure can occur.
    - Scale.
        - Insufficient water applied to the strip, causes high temperatures which leads to scale formation.

After the F5 pre-sprays the transfer bar enters the Finishing Mill. The Finishing Mill consists of a series of high powered mill stands which will reduce the transfer bar into a flat continuous strip of the specified thickness. The Finishing Mill also controls the shape and temperature of the strip that in turn controls the metallurgical properties achieved by the strip. The phase transformation will occur when the correct temperature for the chemistry is reached, this will usually occur on the Run Out Table where the cooling is controlled with a water cooling system. However, with low temperatures phase transformation can also occur in the Finishing Mill.

Following the Finishing Mill is the Run Out Table (ROT). The ROT controls the mechanical properties of the strip via manipulation of the phase transformation. This is achieved by controlling the cooling pattern of the strip which will affect the microstructure present. After the Run Out Table the steel is coiled and dispatched to the customer.

1.1.7.6 Finishing Mill Scrubbers and Edge Sprays

The Finishing Mill scrubber system is situated within the guides at the entry to the first three Finishing Mill stands. There are two main purposes of the Scrubber system:

1. To condition the surface temperature of the bar to avoid reaching temperatures where tertiary scale could be formed which would then get rolled into the strip surface.
2. To provide a degree of scale removal before each roll bite.

Like the scrubbers the edge sprays are situated at the leading edge of the entry guides to the first 3 stands. The purpose of the edge cooling sprays is to cool the strip surface, along the strip edges, with the aim of preventing the surface temperature of the bar from reaching a temperature where tertiary scale could be formed.

The edge sprays are required due to the edge section (100 - 150mm in from the edge) having a higher temperature than the centre surface temperature. A 50 °C increase in edge temperature is observed compared to the centre line temperature. This temperature difference is due to heating in the Furnace.

There are a number of potential failures that are associated with the scrubbers and edge sprays, which include the following;

- Leak in header or missing nozzle
    - Uneven cooling across the strip width causes localised temperature variations. This can cause a ridge across the profile structure. After cold reduction this will appear as a shape defect.
    - Surface temperature too high in early stands. This can lead to tertiary scale on the strip surface.
- Incorrect use on Heavy Gauge products.
    - Surface temperature is too low due to too much water being applied. This can result in transverse cuts on the strip surface.

For the edge cooling sprays to work effectively, the surface of the bar needs to be free of water coming from the previous interstand cooling curtains or from badly fitting top exit strippers. Ineffective cooling will result. This also applies to the effectiveness of the top Scrubbers. The air blow-offs, situated on the side of the mill housing, need to work effectively to clear the water off the surface of the strip.

1.1.7.7 Interstand Cooling

The interstand cooling system is located in the top and bottom exit furniture of the first 4 Finishing stands (F5-F8). The purpose of the interstand cooling system is to cool the strip as it is being rolled. It comprises of a series of continuous full width cooling water curtains designed to cool the strip as it is being rolled. This is an automatic control system (but can be put into manual mode) done as part of the Finishing Mill exit temperature control. It also conditions the bar so that the formation of tertiary scale is suppressed, especially at the exit of the first 2 stands when the surface temperature of the product is at its highest.

1.1.8 Parsytec

Parsytec is a computer aided Surface Inspection System (SIS) that is in use in more than 25 rolling mills worldwide. The defect catalogue has been built using experience of inspection personnel together with information gathered from the Parsytec system. In the Hot Mill in

Port Talbot the top surface cameras are at the start of the ROT and the bottom surface is at the end of the Hot Mill.

1.1.8.1 Detection

All heterogeneous sections of the strip are detected by the Parsytec system. Issues with correct classification can occur due to environmental conditions. The main environmental conditions in the Hot Mill include water droplets, Parsytec utilises a suppression algorithm to remove these environmental factors. However, using these suppression algorithms inevitably has the potential to also suppress real defects which are covered by pseudo defects like water drops. Furthermore, high levels of pseudo defects can consume the system's computing at high rates. This can result in real defects going unrecognized while the system is busy processing pseudo defects. Disturbances created by the physical environment can never be completely eliminated even by the most sophisticated software that Parsytec has to offer. It is therefore highly beneficial to take measures to protect the surface inspection system and the inspected strip surface from adverse environmental influences.

It should also be noted that the system does not scan the surface of the strip in a uniform way. The system gives preferential treatment to areas that are likely to have roll type defects. This can result in a periodic pattern for scale appearing on the strip due to this more intensive scan.

1.1.8.2 Classification

Detected sections are classified into features and these features are compared to the Parsytec catalogue to determine which defect the feature best represents. Problems with classification arise when defects overlap each other or environmental conditions (water/steam) obscure the defect. These can make it harder for an accurate judgement to be made and more likely that the detected area will be classified as a non-classified defect.

Table 3. Parsytec Recognition Rates.

	Initial Performance		Commissioned System Performance	
Area Defects	Hit Rate	Alarm Rate	Hit Rate	Alarm Rate
Heavy Scale Medium Scale	80 %	5 %	90 %	3 %
Red Scale	not included			

Table 3 shows the expected accuracy of the Parsytec system from initial performance to when the system is working fully. The hit rate is the percentage of coils with a scale defect that are caught by the Parsytec software. The alarm rate is the percentage of coils that are incorrectly classified by the system.

1.1.8.3 Action

The operator uses the Parsytec results to determine an action for the coil. The data is also communicated to the Finishing Mill pulpit to allow corrections to be made to improve future runs.

1.1.9 Run Out Table (ROT) and Coiling

There are two functions of the ROT, the first is to connect the Finishing Mill to the coilers. The second function of the ROT is to provide an area for the correct degree of cooling to occur so that the required mechanical properties can be achieved.

After the ROT section is the coiling section. The purpose of the coiling is to coil the strip so that it can be sent to the customer or for further processing.

1.1.10 Conclusions for Process

The Hot Mill process at Port Talbot has a long history of suffering from scale related quality issues, mostly due to its layout and the power of the Finishing Mill stands. This lack of power means that harder grades need to be exposed to higher temperatures to allow them to be rolled, which results in a greater chance of scale formation. To be able to discern the impact of scale formation on the process, the process of scale formation needs to be understood.

1.2 Scale

Scale is an oxide layer that builds up on the surface of the steel when it is exposed to environmental conditions. Iron oxide can form at any point during the life cycle of the product; however it is particularly likely to arise at the Hot Mill. Higher temperatures increase the overall reaction rate due to greater kinetic energy, this causes a faster chemical reaction. This is why it is important to gain an understanding of the scale creation at the Hot Mill. However, this section will concentrate on understanding scale formation at the Hot Mill and what factors contribute to its formation.

1.2.1 What is iron oxide made of?

The scale that forms closest to the steel substrate is Wustite (FeO) followed by Magnetite (Fe_3O_4) and Hematite (Fe_2O_3) (see figure 4). The ratio of the phases can vary considerably depending on the temperature, time and steel chemistry; a ratio of 95:4:1 is to be expected (Wustite:Magnetite:Hematite). Bolt (2003) states that the important properties to consider for the oxide layer are thickness, composition, adhesion and structure of the scale formed.

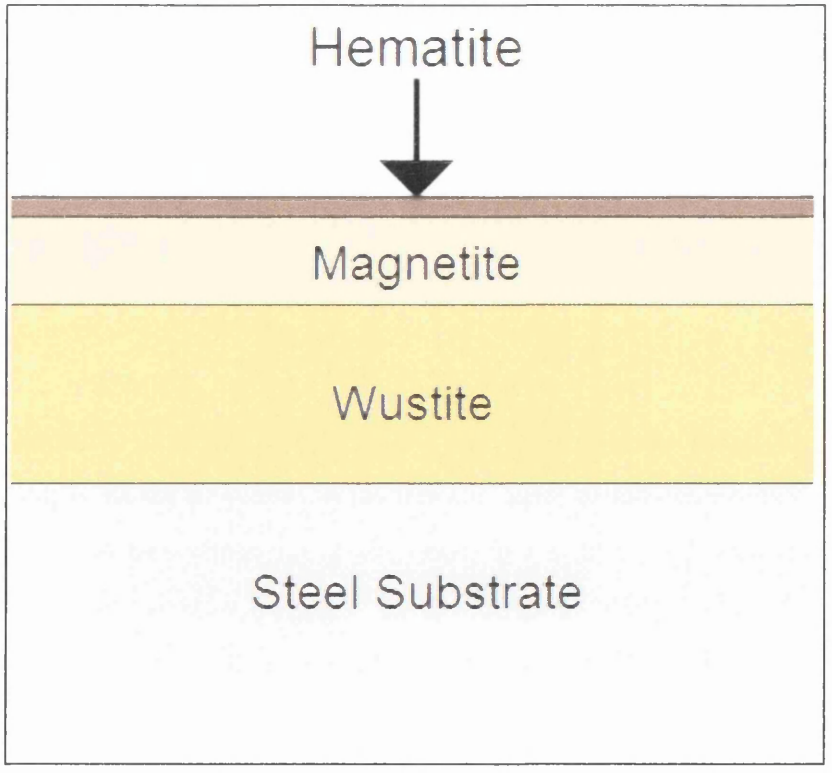


Figure 4. Composition of scale.

The thickness of the oxide layer will also be determined by its processing conditions. Bolt (2003) states that the most important parameters for thickness are the finishing temperature and gauge of the coil. Depending on the temperature of the strip, oxide could be formed at the ROT and also during coiling but that scale formation at these stages is rather modest. This is supported in the work by Yang *et al.* (2008), whereby they discuss the influence of Silicon on thickness. Yang *et al.* (2008) investigated steels with Silicon contents ranging from 0.01 - 1.91 wt. % and concluded that oxidation rates and scale thickness decreased with increasing Silicon. They also concluded that different oxides are formed at different percentages of Silicon. This work is important because it demonstrates the importance of limiting the chemistry range to avoid scales with different properties forming. Taniguchi *et al.* (2001) further demonstrated the importance of limiting the chemistry range. They found that when Silicon is at high percentages (1.14 wt. % and above), the penetration depth of the scale increases with increasing Silicon. The work of Munther and Lenard (1999) concluded that thick scale provides a degree of lubrication to the work rolls, while a thin scale has a higher coefficient of friction and is more adherent to the metal substrate, which makes it harder to remove. Therefore from the work of Yang *et al.* (2008) and Taniguchi *et al.* (2001) it can be concluded that increasing the Silicon percentage creates a thin adherent layer of scale which can a high deree of penetration into the steel substrate. However, Silicon is not likely to play an important part in this analysis due to the percentage and limited range present in the steel investigated.

The composition of these formed oxides influences the properties of the scale. The classical three-layer scale model, as observed in figure 4, reveals that when pure iron is oxidised under normal conditions, Wustite (FeO) is formed closest to the steel substrate followed by Magnetite (Fe_3O_4) and Hematite (Fe_2O_3). Each of these oxides have unique properties. Bolt (2003) states that at hot rolling temperatures, Wustite has a high plasticity, but is extremely brittle at room temperature, which means that a large amount of Wustite will result in poor scale cohesion at room temperatures. Hematite is extremely hard and brittle and is highly detrimental during all stages of processing due to the increased work roll wear it causes and the poor strip quality it creates. Magnetite is the phase that is least brittle at room temperature. This phase can be tolerated in small quantities, because it is not as hard as Hematite and shows some plasticity at hot rolling temperatures. Thinner scales are more adherent than thicker scales and the reason for their increased adherence is the higher amounts of Hematite and Magnetite, especially closer to the steel substrate. This causes

problems for further processing because Magnetite and Hematite are more adherent than Wustite, which makes it more difficult to remove.

The adherence of the scale to the steel is proportional to the composition of the scale. This is because Magnetite and Hematite are a lot harder than Wustite and are more difficult to remove by pickling/scale breakers. This increased tensile strength (see table 4) of the Magnetite and Hematite contributes to increased work roll wear and rolled in scale.

Table 4. Typical mechanical properties for scale phases (Bolt (2003)).

	Tensile Strength at room temperature
Wustite	0.4 M Pa
Magnetite	40 M Pa
Hematite	10 M Pa

1.2.2 Influences on Scale

There is no single variable that can be attributed to scale formation in the Hot Mill. To truly comprehend complex relationships between the variables affecting scale formation the relationships need to be understood. There are many potentially important variables including;

1. Mill layout
2. Temperature
3. Dimensions
4. Time between different processes
5. Chemistry
6. Water systems
7. Rolling

1.2.3 Mill layout

The mill layout is of vital importance to how the different variables will be able to interact with each other, leading to scale formation. Blazevic (1995, 1996) stated that the distance between the descaler and Finishing Mill entry has a significant impact on tertiary scale formation. The shorter the gap between the sprays and the Finishing Mill entry allows less opportunity for scale formation to occur. Even though the mill layout is of vital importance and needs to be considered when reading literature on scale, the aims of this project are not to recommend a new mill layout for Port Talbot Hot Mill rather to better understand how the variables are influencing scale formation. Due to this a more detailed discussion on the advantages of different mill layout will not be performed in this review. Instead the Port Talbot Hot Mill will be discussed.

1.2.4 Temperature

The thickness of the scale is proportional to temperature and time. Sun *et al.* (2004) observed that as temperature and oxidation time increases the scale thickness also increases. It was observed that for the first 20 seconds the scale increased in thickness according to linear growth, and then by parabolic growth after that. They explain this observation by the transport mechanism of oxidation. Initially the scale is very thin which allows rapid diffusion of oxygen with the metal at the iron-air interface. As the scale layer increases in thickness the transport rate will reduce and the diffusion will obey parabolic laws.

The composition of the scale that develops is also highly temperature dependent. At temperatures greater than 570 °C the main composition of scale is Wustite, which is generally the thickest layer. This is then followed by Magnetite and Hematite (see figure 4). The literature on what composition of scale is observed at different temperatures is highly variable, mostly due to different Hot Mill layouts, different types of steel, different experimental methods, and different temperatures. The literature review of Bolt (2000) concludes that as the temperature is increased the more likely it becomes that the scale formed will be composed of a higher percentage of Magnetite and Hematite at the expense of Wustite.

It should also be noted that the surface temperature of the transfer bar can be vastly different from the core temperature of the coil, especially when the coil interacts with interstand sprays and work rolls. The surface temperature will rise due to heat flow from the core of the coil. The amount of scale formed depends on the surface temperature at the time, but the surface temperature is also dependant on the scale thickness. The temperature recorded by the pyrometers gives a good indicator of the temperature of the steel; however it can be influenced environmental factors. If the scale formed across the coil is not consistent then the surface temperature will not be consistent.

Thermal stresses caused by thermal expansion due to cooling/heating and metallographic changes can help break up the scale. For example when austenite transforms to ferrite a 1 % volume expansion occurs, which can aid in breaking up the scale.

1.2.5 Dimensions

The main influence of the transfer bar and strip dimensions is the effect this has on temperature. Thinner gauges are more difficult to keep at the required temperature and thicker gauges have a large difference in temperature between the centre and surface of the coil. The percentage of gauge reduction will also affect the amount of temperature in the slab and how much scale will form.

1.2.6 Time

At a constant temperature the growth of scale is linear with time until the scale becomes sufficiently thick. This initial linear growth is when the Wustite is formed. After this thickness is reached the growth is determined by the diffusion of iron ions, this is when Magnetite and Hematite are formed.

1.2.7 Chemistry

The chemistry of the steel substrate has a strong influence on the type of scale that will form in the Hot Mill. The two elements that have the biggest influence on scale properties are Silicon and Phosphorus. If Silicon and Phosphorus are at levels and temperatures where they will influence scale then Silicon will have the larger impact.

The distribution of elements in the scale is not uniform. Elements that can be found throughout the scale include Manganese, Aluminium (to an extent) and Silicon (to an extent). However, the majority of elements have a high oxygen affinity and will form oxides that will accumulate in the scale at the metal/scale interface; these elements include Aluminium, Silicon, Phosphorus, Boron, Chromium, Molydbenum, Titanium, Niobium, and Zirconium. More noble elements such as Copper, Nickel and Tin will become enriched in the metal increasing the adhesion of the scale.

1.2.7.1 Silicon

The work done by Taniguchi *et al.* (2001) and by Yang *et al.* (2008) has discussed some of the influences of Silicon on scale. They found that the Silicon decreases the amount of oxidation that occurs which in turn reduces the scale thickness, but at high percentages of Silicon (> 1.14 wt. %) an oxide enriched with Silicon increases the oxides adhesion to the steel, making it harder to remove. Increased levels of Silicon have the effect of vastly increasing the oxidation rate especially above 1177 °C. This is due to the temperature of 1177 °C being the eutectic temperature of FeO-Fe2SiO4. This causes a catastrophic amount of scale to be formed due to the presence of the liquid phase at the interface and at the grain boundaries. Overall the scale will be very adherent to the metal. Silicon can also cause red scale/tiger stripes to form due to an increased metal/scale entanglement. This type of scale can be extremely difficult to remove.

Silicon has a large influence on scale, however, the grades under investigation have low levels of Silicon and a limited range. Therefore, Silicon is not likely to be observed as a significant variable.

1.2.7.2 Phosphorus

For tertiary scale formation, Bolt (2000) states that Phosphorus is of particular importance. Phosphorus reduces the overall amount of scale that is able to form on the strip. This is because Phosphorus becomes enriched at the metal/scale interface which suppresses the diffusion of Fe ions. This lack of Fe ions will not stop the oxidation from occurring, but will promote the formation of Magnetite and Hematite over Wustite.

These observations are corroborated by operator's experiences with re-phosphorised steels being very sensitive to oxide defects; increased work roll wear and increased frequency of work roll changes. This increased wear is due to increased friction caused by the harder scale (Magnetite and Hematite) being present.

1.2.8 Descaler sprays

After the coil box is the Crop Shear followed by the FSB (see figure 2). The purpose of the FSB is to break up scale on the strip before it enters the Finishing Mill. There are descaling sprays situated prior to F5 (the first Finishing Mill stand) and are used to suppress the formation of scale. There are also interstand cooling sprays between each pair of work rolls.

Silk (2001) investigates descaler effectiveness and determines that the main components of successful descaling include the impact pressures used and the angle of the descaler conditioning sprays. The slabs then proceed to the Interstand Spray System for further scale to be removed by this spray. This system is either operated manually or run in automatic mode. There are many practical problems when using the descalers. However, the most important one is the higher the flow the more descaling can be achieved but the faster the temperature is reduced. This loss of temperature is a significant concern because if the strip loses too much heat it will not be within its finishing temperature specification and will not achieve the required mechanical properties. If the temperature of the strip is lowered further then the strip might become too cold to be hot rolled and could damage the mill. This is a particular problem for thin gauges because they have less internal heat to maintain a useable working temperature.

A more accurate term for descaler sprays is conditioning sprays. As their purpose in reducing scale formation is to condition the surface of the strip to not generate scale.

1.2.9 Rolling

Munther *et al.* investigated the effect of different scale thicknesses on the friction within the Hot Mill work rolls. It was discussed that a thicker scale produces a lower degree of friction and adds a degree of lubricity to the rolls; and that a thin scale was more adherent, imparting

a higher friction and was more difficult to remove than the thicker scale. It was determined that the thickness of scale had a greater influence on friction than the amount of reduction, the rolling temperature, and the roll velocity.

There are two types of rolls that are used in the Finishing Mill; these are high chrome (HiCr) and high strength steel (HSS) rolls. The HSS rolls are cheaper and can last longer than HiCr rolls. However, due to an uneven build up of scale on the HSS rolls there will be an impact on the scale on the strip. This will result in a poor surface quality of the strip when HSS rolls are used instead of HiCr rolls. Due to high surface demands, some products (such as tinplate) will only be rolled with HiCr rolls rather than HSS rolls.

1.2.10 Downstream Scale

If scale is left to form and not removed before it reaches its end application it can have a detrimental effect on surface quality. Figure 5 indicates how scale can appear at the galvanising line. If scale is left on the surface of the coil at the coating stage then scale could cause quality issues including;

- Fleck - Worn work roll surfaces on finishing stands pick up scale from strip and impress pattern back onto strip.
- Jet - Scale formed in Reheat Furnaces or between Roughing stands - not removed by water sprays due to inadequate water pressure or blocked jets.
- Pits - Small piece(s) of scale not washed off between finishing stands that pickle out leaving deep pit(s).
- Rolled in - Single pieces of scale, rolled into the strip during hot rolling and not pickled out.

Figure 5. Scale as observed from the galvanising line.

1.2.11 Conclusions to Scale section

Scale has a detrimental effect on surface quality especially for downstream applications. The formation of scale is influenced by a number of factors including temperature, time and chemistry. These factors and those that influence them need to be considered when selecting the dataset and performing the analysis.

1.2.12 References

D.T. Blazevic, Rough surface - The major problem on thin cast hot strip mills, Spring Convention Assoc. Iron & Steel Eng., Salt Lake City, USA (1995).

D.T. Blazevic, Tertiary rolled in scale - The hot strip mill surface problem of the 1990's, in: 37th Mechanical Working and Steel Processing Conf. Proc., Vol. 33, Hamilton (Canada) 1995, Iron & Steel Society, Warrendale 1996, p. 33-38.

Bolt, H., 2000. The properties of oxide scales on hot rolled steels: A literature review,. Corus Research, Development & Technology, Ijmuiden Technology Centre.

Bolt, H., June 2003. Understanding the properties of oxide scale on hot rolled strip, in METEC Congress 03/3rd European Rolling Conference (ERC), Dusseldorf.

Munther, P. and Lenard, J., 1999. The effect of scaling on interfacial friction in hot rolling of steels. *Journal of Materials Processing Technology*, 88, 105-113.

Silk, N., October 2001. The practical aspects of hydraulic de-scaling. *Steel Times International*, 38-44.

Sun, W., Tieu, A. K., Jiang, Z., Zhu, H., Lu, C., 2004. Oxide scales growth of low-carbon steel at high tmperatures. *Journal of Materials Processing Technology*, 155-156, 1300-1306.

Taniguchi, S. Yamamoto, K., Megumi, D., Shibata, T., 2001. Characteristics of scale/substrate interface area of Si-cntaining low-carbon steels at high temperatures. *Materials Science and Engineering A*, 308, 250-257.

Yang, Y., Yang, C., Lin, S., Chen, C., and Tsai, W., 2008. Effects of Si and its content on the scale formation on hot-rolled steel strips. *Materials Chemistry and Physics*, 112, 566-571.

1.3 Data Mining

Data mining is a valuable set of techniques which are utilised for improving the understanding of a system. This is achieved by understanding the system through expert knowledge, with the aim of collecting relevant and usable data sources so that they can be interrogated to give a greater understanding of the process. Data mining techniques are used in many different sectors, including the steel industry.

Data mining is a powerful method of getting useful information from a dataset. There are however several factors which need to be considered when performing any data mining technique.

- Badly formatted data
    - This can give unexplained or inaccurate results.
- Non standard classification terms and methods
    - An example of this is the location of defects in a coil. Some systems will mark the defect at its centre while others will log the edge position. If the difference is not noted then similarities and differences in the data can be missed, and inaccurate results will be acted on.
- Lack of detailed information on the process.
    - Without the having the important variables for the system, links can be made between data that cannot be explained. Not understanding the relationships between how the data interacts with other data will result in poor results.
- Black-box techniques.
    - Some techniques are difficult to interrogate and understand how decisions have been made. If the results come with conclusions that are expensive to implement or go against past practices, then having an explainable reason will help the argument. An example of a black-box technique is an artificial neural network.
- Accuracy of the data.
    - If the accuracy of the data is not known then the accuracy of the results can be called into question.

1.3.1 Directed and Undirected Techniques

One method of classifying data mining techniques is to classify them into directed or undirected methods. Directed and undirected are also referred to as hypothesis testing and knowledge discovery. Hypothesis testing is a top-down approach, with the aim being to prove/disprove a given theory. Knowledge discovery is a bottom-up approach with the intention being to interrogate the data to find out something unknown.

There is a specific process for both directed and undirected methods. First the undirected method will be explained.

- Identify data availability and reliability, to do this attention needs to be given to the following
    - Understanding of how data is acquired (sensors, probes, manual/automatic recording systems)
    - Find out how the data is stored and the best method of extracting it
    - Investigate if errors in data alignment can occur when the data is retrieved
- Format data for analysis
    - Ensure the data can be used for the chosen data mining technique
- Create a model and use a training data set to train the model
- Use an evaluation data set to evaluate the accuracy of the model
- Use the model on live data to show similarities in the data
- Decide which similarities require further investigation
- Perform the required direct method to get further information from the data
- Investigate another hypothesis

The directed method is as follows;

- Arrange the data so that it is clustered into relevant classes
- Having the data clustered by an undirected method is standard
- An example of preclassified data is splitting process data from the galvanising line into classes for annealing code or grade
- Format data for analysis
- Ensure the data can be used for the chosen data mining technique
- Create a model and use a training data set to train the model
- Use an evaluation data set to evaluate the accuracy of the model

- Use the model on live data to show similarities in the data

A more detailed review is provided for specific data mining techniques.

1.3.2 Cluster Detection

Clustering is used to determine similarities between the data; it does this by assigning groups to the data depending on their variables. This is an example of an undirected data mining technique.

1.3.3 Binomial Regression

This type of regression can take a variety of input values and outputs them as 1 which equals success or 0 which equals failure. There are many different types of binomial regression, including logistic regression.

By fitting data to a logistic curve the probability of an event occuring can be predicted, this is called Logistic Regression. The predictor variables used can be either numerical or categorical. This model's strength comes from the fact that any value for the input can be fed into the model while the outputed value is between 0 and 1. The input value represents the factors that influence the probability of the data.

The probit model is used in binomial regression as a fast way of determining the maximum likely estimates for a model.

1.3.4 Classification and Regression Trees (CART)

CART uses a decision tree method for defining data into sets. The main advantage of decision trees is that it is easy to explain the results and work out which parameters affect the process. Due to the ease in explaining the rules, they can be expressed easily in a SQL search, which makes it easy to add new records. CART is an example of a directed data mining technique and is particularly good at classification of data.

CART partitions the inputted data using a series of rules. These rules usually consist of

interrogating the data to split it on the grounds of if it is greater than or less than a given value.

1.3.4.1 Pruning

Pruning is performed to improve the efficiency of the model. Pruning takes into account the accuracy and size of the model, with the intention of reducing the size but keeping as much of the accuracy as possible.

As the data is split generally the probability of misclassification[1] (R*(T)) will reduce. When a node[2] contains only one value, it is classified as that value and will give a misclassification result as a zero. However, in practice there are an optimum number of nodes to be used. Having a large number of nodes will result in a decrease in the R^2 value. Increasing the number of nodes beyond the optimum value will result in a detrimental impact on the prediction of any new datasets. This is due to overtraining of the training dataset, which means that the model is too specific to the dataset to give an accurate prediction of unseen datasets. This can result in the tree with a large number of nodes having a lower R^2 value than a tree with the optimum nodes. A tree with fewer nodes will not usually have enough classification power to be a useful as a predicting tool.

The optimum number of nodes can be reached by one of two ways. The first method is to grow a tree that is far too big and then prune it back to reach the optimum value. The second method is to work out a better method of predicting the R*(T) value. A good value of R*(T) can be obtained by using a test sample of a larger sample.

1.3.5 Artificial Neural Networks (ANN)

Artificial neural networks are based on the neural connections present in the brain. The neural network uses a training set to generalize relationships for the data, and then it is used to predict general data sets. Neural networks have two weaknesses. The first is their sensitivity to the given format of the data it is used on. A change in format can give vastly different

[1] When R*(T)=1=100 % misclassification, R*(T)=0=100 % classified
[2] A node is the result of a split.

results; so a vast amount of time is spent formatting the data and removing anomalous results prior to analysis. The other weakness is the lack of description in the results output by the neural network. Neural networks can be applied to undirected data mining techniques in the form of Self Organising Maps (SOM's).

SOM's output their results in the form of a 2D map. Further analysis of the map can be carried out using clustering, to attain information on the different properties.

1.3.6 Multivariate Adaptive Regression Splines (MARS)

MARS is a non-parametric regression analysis that is used to improve the accuracy of linear models. There are many benefits and disadvantages to the MARS technique including the improved flexibility over the linear regression technique; however it is more difficult to prove that the MARS model is more accurate than it is for the linear model. Explanation of results is easier for MARS models than neural networks, because they are represented graphically and the knots can be explained using the rules that created them. Another advantage over neural networks is the data does not have to be prepared before it is used, due to the hinge functions automatically partitioning the data.

1.3.6.1 Hinge Function

The addition of a hinge function allows the MARS model to partition data in a similar way to CART models. This partitioned data can be treated independently from each other which is beneficial for inputs that act differently under changing conditions. This is achieved within the hinge function with the addition of a constant that is usually referred to as a knot. The hinge function is usually set to zero for part of its range and will change when the conditions are met. A simple hinge function will often look like this:

max(0,x-c) or max(0,c-x)

c=constant/knot

When the hinge function achieves its knot value it is activated and the data is partitioned.

MARS is used for both categorical and numerical data, but the hinges often give improved accuracy for numerical data. The use of large datasets is common with MARS models, but some other techniques can obtain the required information faster, such as CART.

1.3.7 Principle Component Analysis (PCA)

The purpose of PCA is to turn variables into linararity uncorrelated variables called principle components. This is useful if the dimensionality of the dataset needs to be reduced but the variability needs to be maintained.

1.3.8 Partial Least Squares (PLS)

The PLS method has many elements in common with PCA. The main difference is that rather than trying to reduce the variance it finds a linear regression of the data by projecting the observed and predicted values in to a new space.

1.3.9 Genetic Algorithms

Genetic algorithms were developed around the idea of using the principles of natural selection to identify the optimum set of parameters to work out the best result. The model is run several times and the least predictive results are removed each time, so that the best set of parameters is converged on as the program is run. Genetic algorithms are often used to improve other techniques such as neural networks. This is an example of a directed data mining technique.

1.3.10 Multinomial Logistical Regression (MLR)

Instead of a binary model that has two possible outcomes, 1 and 0. A multinomial model can have multiple outcomes, which can (in some cases) be used to more accurately represent the dataset.

1.3.11 Adaptive Neural Fuzzy Inference System (ANFIS)

The ANFIS model combines elements from neural networks and fuzzy logic principles.

1.3.12 Conclusions to Data Mining

There are many different techniques used in data mining, all with their own strengths and weaknesses. The key factor in successful data mining is to understand the data which is available, and to be able to perform a successful analysis from that data.

Whatever statistical approach is used to model the formation of scale at the Hot Mill, it must address the issues associated with co-linearity and measurement error. One approach is to directly model the count data collected by the Parsytec system, i.e. the recorded number of defects on the top (and or bottom) surface of the coil. Traditionally, this is done using a Poisson distribution with the mean of the distribution being made a linear function of the process variables. For the data set collected for this thesis this is not straightforward because this scale count data has many "missing" values - for example none of the coils in the data set had scale counts in the range 991 to 1002 (there are no values within this range). Indeed there are many other such missing value counts in this data set (over the recorded range of 0 – 4000 counts). To use the Poisson model would therefore require a form of truncation and consequent loss of information. A nonlinear variation of this approach would model the mean of the distribution using, for example, a neural network.

The following chapters will explain the methodology behind several different data mining techniques. The methods that will be included encompass the following;

- Partial Least Squares (PLS)
- Logit
- Principle Component Analysis (PCA)
- Multinomial Logistical Regression (MLR)
- Adaptive Neuro Inference Fuzzy Systems (ANFIS)

Due to the inherent strengths and weaknesses of different data mining techniques it is likely that multiple techniques will need to be employed to achieve the best results.

Chapter 2. Aims of Research Project

This project aims to use a collection of data mining techniques to better understand the influence of Hot Mill process parameters on scale formation at Port Talbot Hot Strip Mill in South Wales. This understanding will then be used to recommend operating conditions designed to minimise the formation of scale at the Hot Mill.

The dataset to which these data mining techniques were applied was carefully chosen to reduce process variation. There are several main factors that were considered to minimise this variability including time period, grade and gauge investigated. The following data mining techniques were chosen to investigate this dataset;

- Partial Least Squares (PLS)
- Logit Analysis
- Principle Component Analysis (PCA)
- Multinomial Logistical Regression (MLR)
- Adaptive Neuro Inference Fuzzy Systems (ANFIS)

These data mining techniques were chosen because they represent different ways of investigating the dataset. Some of the differences include diverse ways to represent the dependant variable used in the analysis. Three different representations were used binary, multinomial and continuous.

Chapter 3. Experimental Techniques

3.1 Data Collection

The purpose of this section was to understand how data was collected from the available process parameters, with the intention of better understanding their influence on scale formation by using data mining techniques.

3.1.1 Considerations

To be able to achieve a successful data mining investigation, the data sampled needs to be carefully selected to be able to accurately represent the process being investigated. The main factors that need to be considered for scale formation include;

- A large dataset
    - Many of the data mining techniques require large datasets to be effective
- A comparable dataset
    - The cases in the dataset investigated need to be comparable to each other
- A clean dataset
    - Data recording anomalies need to be removed, these include;
        - Probe recording errors
- Comparable time period
    - Due to changes and improvements being implemented to the process of the Hot Mill, longer time periods may not be comparable for the duration of the time period.
- Representative independent variables
    - The variables selected to represent the process
- Representative dependent variables
    - A variable will need to be chosen to represent the effect of the process variables on scale.
- Processing
    - The material needs to be processed the same way, with particular attention to;
        - Gauge and grade
            - Different gauges and grades will react and be reacted to differently when being processed in the Hot Mill
        - Roll Type

- HSS/HiCr rolls will react differently to rolling over time and could give different results.

- Schedule
  - It has been noticed that the coils processed in the Hot Mill are less stable when the gauge or grade is changed
- Application
  - The Dataset needs to represent material for the same application so the results are meaningful

The above problems were overcome by using data associated with coils processed for Tinplate applications. The tinplate products have several advantages over other products including;

- The tinplate product is intended for surface critical application so scale levels must be low.
- All the coils in the dataset have the same chemistry requirements.
- To minimise the effect of different heating/cooling cycles, only a single gauge of 2.1 mm has been investigated.
- This is a high quantity product type so a lot of material is processed, and they are processed in their own rounds.

Coil data was collected between September 2009 and March 2010 for all the process variables described above and table 5 gives some descriptive statistics for this sample of data, together with the full name of each process variable and its short hand descriptor – x_j. The sample size was n = 1577.

Table 5. Process Variables and their Sample Means and Standard Deviations.

Process Variable	xj	Mean	Standard Deviation
Wt. % Carbon	x1	0.069	0.0042
Wt.% Manganese	x2	0.485	0.0226
Wt. % Phosphorus	x3	0.014	0.0031
Wt. % Sulphur	x4	0.014	0.003
Wt. % Silicon	x5	0.005	0.0032
Wt. % Copper	x6	0.016	0.0059
Wt. % Nickel	x7	0.011	0.003
Wt. % Chromium	x8	0.018	0.004
Wt. % Aluminium	x9	0.034	0.0047
Wt. % Tin	x10	0.003	0.0018
Wt. % Nitrogen	x11	0.0125	0.0009
Wt. % solAl	x12	0.031	0.0043
Maximum RM Exit Temperature, °C	x13	1121.75	17.87
Minimum RM Exit Temperature, °C	x14	918.34	11.97
Average RM Exit Temperature, °C	x15	1099.43	16.18
Maximum CS Temperature, °C	x16	1099.88	19.89
Minimum CS Temperature, °C	x17	951.69	48.59
Average CS Temperature, °C	x18	1069.55	20.8
Maximum FM Exit Temperature, °C	x19	907.16	7.18
Minimum FM Exit Temperature, °C	x20	829.87	28.96
Average FM Exit Temperature, °C	x21	881.05	5.75
ROT Temperature, °C	x22	731.35	25.13
Maximum Coiling Temperature, °C	x23	693.133	28.3786
Minimum Coiling Temperature, °C	x24	438.354	53.8953
Average Coiling Temperature, °C	x25	561.627	2.7173
Furnace Aim Temperature - Average Temperature, °C	x26	4.77933	14.587
Furnace Centre/Surface Difference, °C	x27	44.9575	24.2802
Furnace Surface Temperature, °C	x28	1227.13	19.5941
Furnace Centre Temperature, °C	x29	1182.48	17.7393
Scrubber Pressure (set)	x30	1.748	0.6603

ISS (set)	x31	-0.692	0.4616
Thread Speed	x32	11.370	0.1772
Slab Length (m)	x33	9.677154	0.470653
Slab Width (m)	x34	0.95601	0.047791
Slab Gauge (m)	x35	0.235172	0.003213
Slab Weight (tonnes)	x36	17.05425	1.163756
Transfer Length (m)	x37	61.01447	4.611045
Transfer Gauge (m)	x38	0.040309	0.002432
Coil Width (m)	x39	906.4052	52.28183
Top Scale Count	x40	126.200	219.9389
Bottom Scale Count	x41	250.008	402.6597
Indictor Variable	Y	2.48193	1.09386
Scale Count Indicator (Binary)	YB	0.327	0.4693

RM = Roughing Mill, CS = Crop Shear, FM = Finishing Mill, ROT = Run Out Table, ISS = Interstand Spray System (-1 = Manual and 0 = Automatic), Srubber pressure (0 = low pressure, 1 = medium pressure, 2 = high pressure). YB = 1 when the scale count on the bottom of the sample exceeds 200 and zero otherwise.

The data collected for this investigation was obtained from Port Talbot Hot Mill. The temperature data was collected from temperature sensors at various points along the Hot Mill. The temperature sensor is positioned in the centre of each section of the hot mill and records the temperature along the entire length of the coil. The temperature data is converted into maximum, average, and minimum for each coil. The chemistries of the coil are measured at the secondary steel making process stage. The scale count is the absolute bottom scale count from the Parsytec inspection system.

3.2 Statistical Methods

This chapter will lay out the methodology used in the Logit, PCA, PLS, MLR and ANFIS models used in the analysis of the dataset described in the data collection section.

3.2.1 Binary Statistical methodology

The Logit, PCA and PLS all use a binary split to describe the dependent variable of the bottom scale count. The dependent variable takes on a value of unity when the scale count on the bottom of table 5 exceeds 200 and zero otherwise. The cut of 200 was chosen for two main reasons. The first reason was that scale counts lower than this are unlikely to be a problem for further processing but as the scale count increases above 200 it becomes more difficult to determine if the scale will cause a problem for further processing without understanding how adherent the scale is. The second reason is that a split of 200 makes the population of bottom surface affected coils represent more than 30 % of the dataset which makes it easier to be confident about the output of the models, i.e. it creates a more balanced data set from which to apply non-linear Partial Least Squares.

Chapter 4. Identification, for control, of the process parameters influencing tertiary scale formation at the hot strip mill using a binary choice model

4.1 Principal component analysis

Principal component analysis (PCA), see Jolliffe (2002) for a good review of this technique, is an extensively used technique of multivariate linear data analysis. The purpose of PCA is to reduce the dimensionality of the data but still maintain its variability and to remove the correlation amongst the process variables used to control the Hot Mill process.

The principal components (PC) are calculated as follows. The p process variables (x_1 to x_p) are standardised to have a zero mean and unit variance to avoid the problems associated with units of measurement. Then p linear combinations of these variables can be formed as follows

$$\begin{aligned} PC_1 &= a_{11}z_1 + a_{12}z_2 + \cdots a_{1p}z_p \\ PC_2 &= a_{21}z_1 + a_{22}z_2 + \cdots a_{2p}z_p \\ &\vdots \\ PC_p &= a_{p1}z_1 + a_{p2}z_2 + \cdots a_{pp}z_p \end{aligned}$$

Eq. (4.1a)

where the z variables are the standardised values of the x_j (j = p) process variables shown in table 5. The so called loadings, a_{11} to a_{1p}, are then chosen so as to maximise the variance of PC_1 subject to the normalising condition (the purpose of the normalising condition is to maintain the same sum of loadings squared across all of the principle components):

$$a_{11}^2 + a_{12}^2 + \cdots + a_{1p}^2 = 1$$

Eq. (4.1b)

PC_1 is then stated to be the first principal component and is a linear function of the z's (and thus the process variables) that has the highest variance. PC_2 then has the next highest variance and so on until all the variation in the z's is picked up by the p principal components. That is

$\text{variance}(PC_1) + \text{variance}(PC_2) + ... + \text{variance}(PC_p) = \text{variance}(z_1) + \text{variance}(z_2) + ... + \text{variance}(z_p) = p$

with

$\text{variance}(PC_1) > \text{variance}(PC_2) > ... > \text{variance}(PC_p)$

Various different algorithms can be used to find values for these loadings, ranging from the simple summation method put forward by Burt (1945) (of which a good review can be found in Childs (1970)) to algorithms that calculate a spectral decomposition of the correlation matrix amongst all the x variables (see the text by Mardia *et al.* (1979)). The Eigen values from this decomposition measure the variation in all the process variables explained by each principal component. Thus the first Eigen values from this decomposition measures the variation in all the process variables explained by the first principal component. The Eigen Vector from this decomposition contains the required loadings for each principal component. By construction these principal components are also all orthogonal and hence completely uncorrelated with each other. The spectral decomposition method will be used for this Principle Component Analysis.

Further, Cattell (1952) has suggested that only those principal components having an Eigen value greater than 1 should be considered as essential and therefore retained in the analysis. This simple rule allows for a substantial reduction in the number of variables included in the final statistical model.

4.2 Binary choice models

4.2.1 A generalised framework

A binary choice model views the outcome of a discrete incident (such as the occurrence of a large amount of scale) as a reflection of an underlying regression. The cost associated with any additional processing or downgrading of a heavily scaled coil (defined above as 200+ observations of scale) is modelled using an unobserved variable y_i^* such that

$$y_i^* = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \varepsilon = x\beta + \varepsilon_i$$

Eq. (4.2a)

where x_{ij} is the i^{th} measurement on process variable j associated with the Hot Mill and the β_j are the parameters that are to be estimated from the data set. β is a vector containing the p+1 parameters of this model and x is the data matrix with the columns of this matrix containing each of the process variables (including an initial column of 1's) and each row contains the i^{th} measurement on these variables. ε_i is assumed to follow some standardised distribution with known variance - to be discussed further below. The cost of having large amounts of scale is not directly observed, just whether the coil is heavily scaled or not. Therefore observations at the Hot Mill take the form

$y_i = 1$ if $y_i^* > 0$ Eq. (4.2b)

$y_i = 0$ if $y_i^* \leq 0$

Provided that a constant (β_0) is included in Eq. (4.2a) the assumption of a zero threshold for y_i^* is an innocent one. The process variables are then taken to determine the probability of $y_i^* > 0$

$\text{Prob}(y_i^* > 0 \mid x) = \text{Prob}(y_i = 1 \mid x) = F(x, \beta)$ Eq. (4.2c)

$\text{Prob}(y_i^* \leq 0 \mid x) = \text{Prob}(y_i = 0 \mid x) = 1 - F(x, \beta)$

where $F(x, \beta)$ is a function, dependent on x and β, that determines the probability that $y_i = 1$ given the values in x, with the natural requirement that $F(x, \beta)$ should tend to 1 as $x'\beta$ tends to infinity (and $F(x, \beta)$ tends to zero as $x'\beta$ tends to minus infinity). There are many cumulative density functions that fit these criteria, but most applications of this type of model are confined to the use of the normal distribution (giving rise to the Probit model - see Maddala (1991) for a good introduction to this model) and the logistic distribution (giving rise to the Logit model - see Maddala (1991) for a god introduction to Logit model).

The following model can be seen as a generalisation of these two binary choice models that enables discrimination between these and other functions through use of standard likelihood ratio tests. From Eq. (4.2a-c) it follows that

$$Prob(y_i = 1) = P_i = Prob\left\{\varepsilon_i > -\left[\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}\right]\right\} = 1 - F\left\{-\left[\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}\right]\right\}$$

Eq. (4.3a)

Eq. (4.3a) makes it clear that F is the cumulative distribution function for ε. One very general specification can be derived from the generalised F distribution. Prentice (1975) has suggested a non-degenerate and standardised version of this distribution whereby the probability density function is given by

$$f(w) = \frac{(m_1/m_2)^{m_1}}{\{T(m_1)T(m_2)\}/\Gamma(m_1+m_2)} \frac{\exp(m_1 w)}{\left(1+\{m_1/m_2\}e^{w}\right)^{(m_1+m_2)}} \qquad \text{Eq. (4.3b)}$$

where

$$w_i = -\left[\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}\right]$$

Eq. (4.3c)

and where f() is the derivative of F() with respect to w and Γ() is the gamma function. In Eq. (4.3a-c), the parameters m_1 and m_2 determine the cumulative distribution function for w_i and thus ε_i. That is, m_1 and m_2 determine the form of the binary choice model such as the Logit and Probit model. Unfortunately, there is no closed form expression for F() within this setting except for some special cases. To take some well-known examples, first consider the case where $m_1 = m_2 = 1$. Then

$$1 - F[w_i] = \frac{exp(w_i)}{1 + exp(w_i)}$$

Eq. (4.4a)

which is the cumulative (standardised) logistic function giving rise to the Logit model which can also be expressed as the log odds model

$$Ln\left[\frac{P_i}{1-P_i}\right] = -\left[\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}\right]$$

Eq. (4.4b)

For finite m_1 and m_2 but with $m_1 = m_2$ this logistic function becomes the generalised logistic. Then when $m_1 = 1$ and $m_2 = 0$

$$1 - F(w_i) = 1 - exp[-exp(w_i)]$$

Eq. (4.5a)

which is the cumulative (standardised) extreme value function giving rise to the proportional hazards model which can also be expressed as

$$Ln[-Ln(1-P_i)] = -\left[\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}\right]$$

Eq. (4.5b)

Other special cases include the probit model when $m_1 = \infty$ and $m_2 = \infty$ and the generalised gamma cumulative function when just $m_2 = \infty$ (see Lawless (2003) for a descriptions of this function).

Whilst there are no closed form expression for the integral of Eq. (4.3b) when m_1 and m_2 are not restricted to the special cases outlined above, values for this integral at a given value for w_i can be obtained numerically. If m_1 and m_2 are both finite, values for this integral at a given value for w_i are found by calculating areas under the Beta distribution with m_1 and m_2 degrees of freedom. More precisely,

$$P_i = 1 - F\left\{-\left[\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}\right]\right\} = B(s_i, m_1, m_2)$$

Eq. (4.6a)

where $B(s_i,m_1,m_2)$ is the area under the Beta distribution with m_1 and m_2 degrees of freedom to the left of s_i, where

$$s_i = \left\{1 + \frac{m_1 exp[w_i]}{m_2}\right\}^{-1}$$

Eq. (4.6b)

For finite m_1 and $m_2 = \infty$, values for this integral at a given value for w_i are found by calculating areas under the Chi distribution with $2m_1$ degrees of freedom. More precisely,

$$P_i = 1 - F\left\{-\left[\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}\right]\right\} = x(2v_i, 2m_1)$$

Eq. (4.6c)

where $x(2v_i,2m_1)$ is the area under the Chi distribution with $(2m_1)$ degrees of freedom to the left of $2v_i$, where

$$v_i = m_1 exp\left[\frac{w_i}{\sqrt{m_1}}\right]$$

Eq. (4.6d)

The integrals calculated using Eq. (4.6c, 4.6d) tend to those obtained from the standard normal distribution (and hence yields the Probit model) as m_1 tends to ∞. For finite m_2 and $m_1 = \infty$, the procedure is to replace v_i with $-v_i$ and to replace m_1 with m_2 in Eq. (4.6c, 4.6d). Most statistical packages have routines capable of calculating such areas and as demonstrated by Prentice (1975) this is best achieved by reparametrising m_1 and m_2 as

$p = (m_1 + m_2)^{-1}$ and $q = (m_1^{-1} + m_2^{-1})(m_1^{-1} + m_2^{-1})^{-1}$ Eq. (4.6e)

4.3 Parameter estimation

Since the y_i are realisations of a binomial process with probabilities given by Eq. (4.6a or 4.6c) we can write the likelihood function as

$$L = \prod_{y_i=1} P_i \prod_{y_i=0} (1 - P_i)$$

Eq. (4.7a)

so that the log likelihood is given by

$$log[L] = \sum_{i=1}^{n} y_i log(P_i) + \sum_{i=1}^{n} (1 - y_i) log(1 - P_i)$$

Eq. (4.7b)

where n is the sample size of the data set and P_i is calculated from Eq. (4.6a or 4.6c) for given values of m_1 and m_2. Given these values for m_1 and m_2, the β parameters in Eq. (4.6a, 4.6c) are chosen so as to maximise the log likelihood in Eq. (4.7b).

4.4 Model identification

Taking a general to specific approach to model identification, all the process variables shown in table 5 are initially included in the model (either individually or as part of a principal component). Then a grid search is carried out at different values for m_1 and m_2 using this general model. The log likelihood defined by Eq. (4.7b) at each (m_1, m_2) combination is stored and the largest of them recorded. Call this statistic $log[L]_{max}$. Clearly the (m_1, m_2) value combination associated with this value determines the distribution most supported by the data. However, any (m_1, m_2) combination that produces

$$\Lambda_r = -2[log(L)_{max} - log(L)_r]$$

Eq. (4.8a)

with a value less than 3.84 is also supported by the data at the 5 % significance level. $\log[L]_r$ is the log likelihood associated with the r^{th} (m_1, m_2) combination. This follows from the fact that Λ_r, under the null hypothesis that the r^{th} (m_1, m_2) combination is the true combination, is asymptotically Chi square distributed with 1 degree of freedom.

Once the correct values for m_1 and m_2 are determined in the way described above, a data based simplification of the model then takes place. The model associated with $\log(L)_{max}$ is called the unrestricted model. If the j^{th} explanatory variable is then removed from this model, a restricted model is obtained with a log likelihood given by $\log(L)_R$. This is achieved by removing the x_j with the smallest student t statistic. The final simplified model is obtained by removing all those variables for which

$$\Lambda_R = -2[log(L)_{max} - log(L)_r]$$

Eq. (4.8b)

has a value in below of 3.14. Finally, the adequacy of this parsimonious model can be assessed in one of two ways. First the proportion of correct predictions can be calculated, where a prediction is considered correct if Eq. (4.6a or 4.6c) yields a value above 0.5 when $y_i = 1$ or alternatively when Eq. (4.6a or 4.6c) yields a value equal to or below 0.5 when $y_i = 0$. The count R^2 defined in Eq. (4.8c) below then lies between 0 and 1 with better performing models having a value closer to 1.

$$Count\ R^2 = \frac{number\ of\ correct\ predictions}{total\ number\ of\ observations}$$

Eq. (4.8c)

4.5 Results and evaluation

4.5.1 Principal component testing procedure

With the use of principal component analysis the correlations existing between the process variables were investigated. The analysis was able to reduce both the number of process variables and remove the correlations among the process variables used in the model. This allowed the model to predict the scale count and from this to gain an understanding of how to

run the Hot Mill so as to minimise scale formation. The maximum correlation coefficient between all the chemistry variables was 0.47 and this was between Copper and Nickel. Further, most of the correlation coefficients were less than 0.1 in absolute value. These relatively low correlations mean that there is little benefit to be gained from forming the principal components of these chemistries. However, table 6 shows that all the temperature variables are highly correlated – especially for example between the maximum and average Rougher Mill temperatures (x_{13} and x_{15}). In general, it can be seen from table 6 that the most highly correlated relationships occurred between the maximum and average temperatures at the Rougher Mill and at the Crop Shear.

Table 6. Correlation matrix for the temperature process variables. A value of 1 represents a perfect correlation and a value of 0 represents zero correlation. A negative value shows an interaction in the opposite direction.

	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	x_{20}	x_{21}	x_{22}
x_{13}	1									
x_{14}	0.04	1								
x_{15}	0.92	0.13	1							
x_{16}	0.76	0.73	-0.01	1						
x_{17}	0.3	0.26	0.04	0.38	1					
x_{18}	0.53	0.57	0.01	0.69	0.3	1				
x_{19}	0.46	0.42	-0.03	0.61	0.46	0.23	1			
x_{20}	0.16	0.12	0.03	0.18	0.12	0.13	0.16	1		
x_{21}	0.48	0.44	-0.04	0.63	0.51	0.26	0.7	0.25	1	
x_{22}	0.06	-0.03	0.02	0.02	0.06	-0.01	-0.01	0.08	0.02	1

Given these high levels of dependency between the temperature process variables, principal components of these temperatures need to be constructed to remove these correlations.

Table 7 shows how much of the variability in all the temperature variables shown in table 5 are accounted for by the principal components.

Table 7. Principle component results for the temperature process variables.

Components	Eigen values	Variation, %	Cumulative Variation, %
PC_1	4.201	42.01	42.01
PC_2	1.106	11.06	53.07
PC_3	1.043	10.43	63.51
PC_4	0.9741	9.74	73.25
PC_5	0.8418	8.42	81.67
PC_6	0.7769	7.77	89.44
PC_7	0.4999	5.00	94.44
PC_8	0.2805	2.81	97.24
PC_9	0.2040	2.04	99.28
PC_{10}	0.0718	0.72	100.00

The first three principal components accounted (see table 7) for nearly 64 % of the variability present in all ten temperature process variables, whilst the first four principal components accounted for nearly three quarters of the variability present in all ten temperature process variables. Using Cattell's rule (1952) of only needing to use those principal components having an Eigen value of around unity or more, the ten temperature process variables can be replaced by the first three or four principal components. These principal components had the form

$$PC_1 = -0.0233(x_{13} - 1121.8) - 0.0016(x_{14} - 918.3) - 0.0251(x_{15} - 1099.4) - 0.0223(x_{16} - 1099.9) - 0.0045(x_{17} - 951.7) - 0.0179(x_{18} - 1069.6) - 0.0493(x_{19} - 907.2) - 0.0045(x_{20} - 829.9) - 0.0644(x_{21} - 881.1) - 0.0003(x_{22} - 731.4)$$

Eq. (4.9a)

$$PC_2 = 0.0156(x_{13} - 1121.8) + 0.0431(x_{14} - 918.3) + 0.0214(x_{15} - 1099.4) + 0.0008(x_{16} - 1099.9) + 0.0008(x_{17} - 951.7) - 0.0007(x_{18} - 1069.6) - 0.0384(x_{19} - 907.2) - 0.0116(x_{20} - 829.9) - 0.0622(x_{21} - 881.1) - 0.0185(x_{22} - 731.4)$$

Eq. (4.9b)

$$PC_3 = -0.0037(x_{13} - 1121.8) + 0.0597(x_{14} - 918.3) + 0.0005(x_{15} - 1099.4) - 0.0027(x_{16} - 1099.9) + 0.0022(x_{17} - 951.7) + 0.0003(x_{18} - 1069.6) - 0.0149(x_{19} - 907.2) + 0.0116(x_{20} - 829.9) - 0.0007(x_{21} - 881.1) + 0.0235(x_{22} - 731.4)$$

Eq. (4.9c)

$$PC_4 = 0.0043(x_{13} - 1121.8) + 0.0085(x_{14} - 918.3) + 0.0080(x_{15} - 1099.4) + 0.0040(x_{16} - 1099.9) - 0.0054(x_{17} - 951.7) + 0.0078(x_{18} - 1069.6) - 0.0047(x_{19} - 907.2) - 0.0249(x_{20} - 829.9) - 0.0086(x_{21} - 881.1) + 0.0233(x_{22} - 731.4)$$

Eq. (4.9d)

The loading coefficients in front of each variable show how heavily weighted each variable is within each principle component. If the loading coefficients are similar , then instead of creating principle components a simpler alternative would be to take their sums (if the loadings are positive) or differences (if the loadings are negative). Thus in the first principal component, the loadings in front of the maximum Rougher Mill (x_{13}), the average Rougher Mill (x_{15}) and the maximum Crop Shear (x_{16}) temperatures are very similar, suggesting that an appropriate dimensionality reduction would be to look at the differences between these three temperatures. The loadings in front of the minimum Rougher Mill (x_{14}) and the average Crop Shear (x_{18}) temperatures are very similar, suggesting that an appropriate dimensionality reduction would be to look at the differences between these two temperatures. The loadings in front of the minimum Crop Shear (x_{17}) and the minimum Finishing Mill (x_{20}) temperatures are very similar, suggesting that an appropriate dimensionality reduction would be to look at the differences between these two temperatures. The loadings in front of the maximum (x_{19}) and average Finishing Mill (x_{21}) temperatures are also very similar, suggesting that an appropriate dimensionality reduction would be to look at the differences between these two temperatures.

4.5.2 Distribution analysis

The β parameters in Eq. (4.6a and 4.6c) were estimated by maximising the log likelihood in Eq. (4.7b) at given values for p and q (with p and q defined by Eq. (4.6e)). All the chemistry

variables in the table 5 were included in Eq. (4.6a and 4.6c), and all the temperature variables were replaced by the first three principal components specified in the previous section. Figure 6 shows the maximised log likelihoods associated with different values for p and q. It can be seen from figure 6 that the largest maximised log likelihood occurs where p = 1 and q = 0.5. The area of the graph below the dashed line in figure 6 shows which maximised log likelihoods are significantly different from this one at the 5 % significance level using the test statistic in Eq. (4.8a). It is clear from figure 6 that both the Logit and Probit models are therefore supported by the data and have log likelihoods very close to the largest log likelihood. Consequently, and to keep the computational aspects of the work simpler, all further results shown below are based on the Logit model.

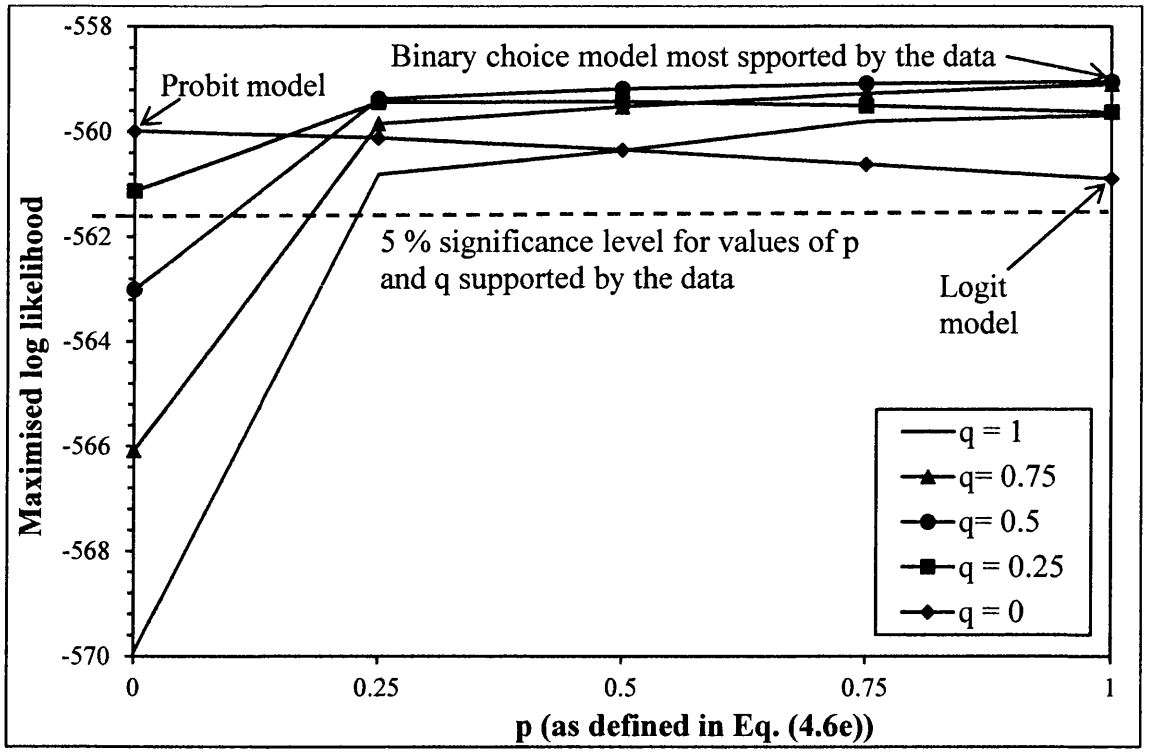


Figure 6. The maximised log likelihood obtained at different values for p and q.

4.6 Logit model results

Using the data based simplification procedure described in Section 4.4 above, a simplified Logit model was derived and this model is summarised in table 8. Only a few chemistries were statistically significant at the 1 % significance level, with these being Phosphorous, Silicon, Copper, Nickel and Chromium. Further, only 2 of the above derived principal

components were statistically significant at this level. Further, of the n = 1577 observation of y, 1140 of these were predicted correctly by the model. This is an overall correct prediction rate of 73 %. However, the model was significantly better at predicting the Hot Mill conditions leading to no significant scale formation (i.e. $y_i = 0$), as the model predicts when $y_i = 0$, some 98 % of the time.

Table 8. Logit model results.

Process Variable	Coefficient	Std.Error	t-value
Constant	-2.49	0.46	-5.38
Percentage of Phosphorus (x_3)	228.28	25.13	9.08
Percentage of Silicon (x_5)	-88.87	22.25	-3.99
Percentage of Copper (x_6)	72.66	13.33	5.45
Percentage of Nickel (x_7)	-184.66	28.79	-6.41
Percentage of Chromium (x_8)	-64.12	19.74	-3.25
PC_1	-0.391	0.04	-10.3
PC_2	-0.232	0.07	-3.32

From table 8 it can be observed that each principal component has a negative impact on the log odds ratio as does the amount of Silicon, Nickel and Chromium. Phosphorous and Copper on the other hand have a positive impact on this ratio. For an indication of which variables have the biggest effect on the probability of scale forming at the Hot Mill it is better to convert the coefficient values in table 8 into quasi - elasticities. These elasticities, contained in table 9, show the percentage change in the probability of scale forming following a percentage change in each of the process variables. These elasticities are calculated around the mean values – shown in table 5 for each process variable as the elasticity will vary with the magnitude of each process variable.

Table 9. Quasi-elasticities.

Process Variable	Quasi – Elasticity
Percentage of Phosphorus (x_3)	0.62
Percentage of Silicon (x_5)	-0.09
Percentage of Copper (x_6)	0.24
Percentage of Nickel (x_7)	-0.39
Percentage of Chromium (x_8)	-0.22
PC_1	-0.08
PC_2	-0.05

From table 8 it can be seen that when the Phosphorous content is at 0.014 % wt (i.e. at its mean value shown in table 5) a further 1 % increase (0.00014 % wt) in the Phosphorous content will increase the probability of scale formation by just over 0.6 percentage points. It can also be observed from table 9 that when the Nickel content is at 0.01 % weight (i.e. at its mean value shown in table 5) a further 1 % increase in the Nickel content will decrease the probability of scale formation by just under 0.4 percentage points. Given the size of the other elasticities in table 9, these are therefore the most important chemistries for controlling scale formation at the Hot Mill, in that changes in these two chemistries will bring forth the biggest change in the probability of scale forming when the temperature variables are at their mean values shown in table 5.

As for the temperature variables, it can be seen from table 9 that when the first principal component is at its mean value a further 1 % increase in this component will decrease the probability of scale formation by just under 0.1 percentage points. The temperatures in the first principle component with negative loadings (see Eq. (4.9a) above for PC_1), have a positive impact of the probability of scale forming. This relationship is looked at in more detail below. Prior to this it is important to look at some probability plots as the above elasticities only apply at the mean values for each process variable. To have a full understanding, the complete range of values must be looked at and this is shown in the probability plots of figure 7.

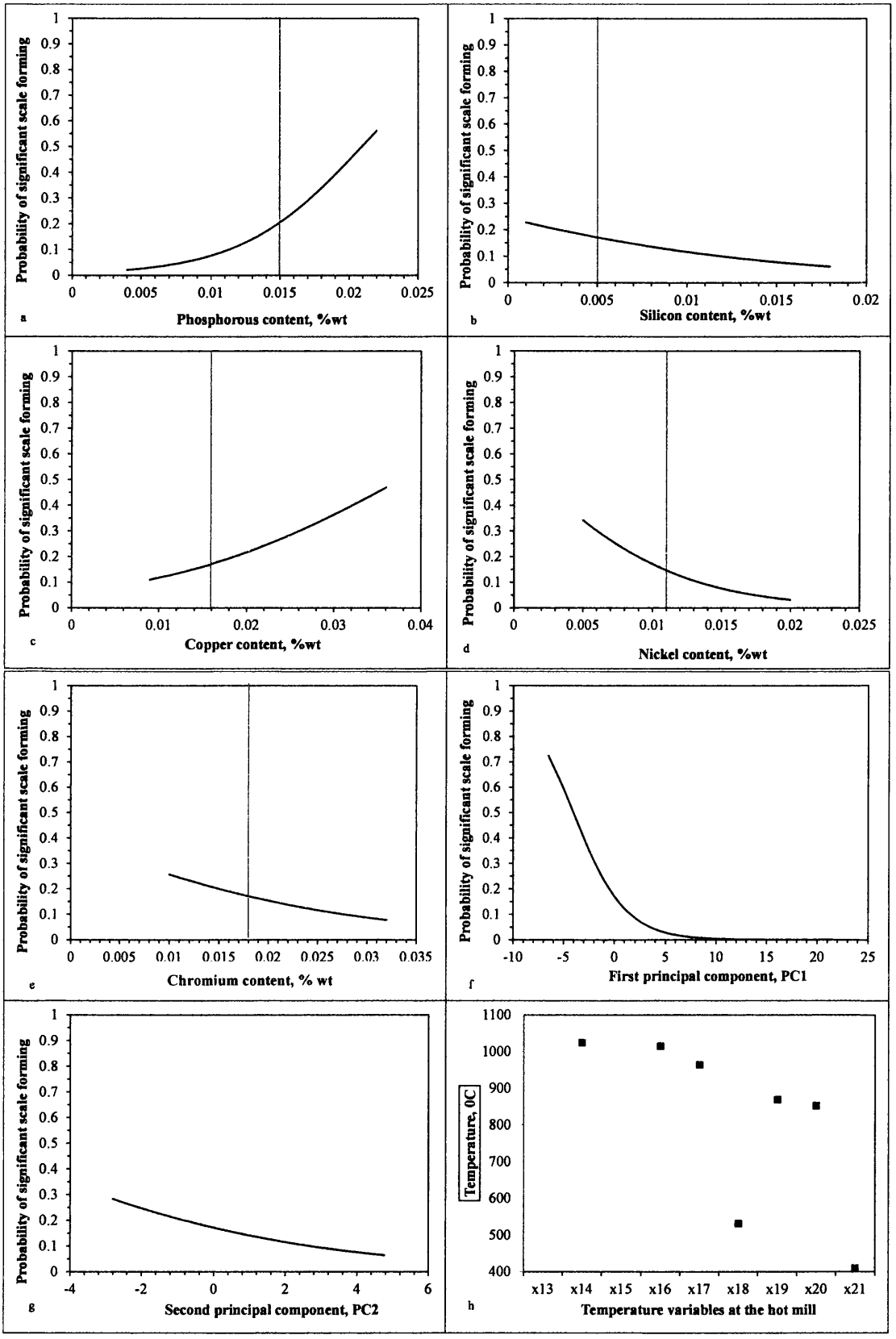


Figure 7. Probability of scale formation with variations in the significant process variables. The vertical line of a – e indicates the mean value for that variable.

In figure 7 the probabilities shown on the vertical axis are calculated from Eq. (4.6a and 4.6c) by varying one of the process variables whilst holding the other process variables at their mean values. These probabilities can also be interpreted as the proportion of coils manufactured that have a significant number of flaws - so preventing them being used for the high value added markets. Plots like those shown in figure 7 can therefore be used for optimising the process, i.e. finding how to operate the Hot Mill so that only a small proportion of manufactured coils have scale counts that prevent them from being used in high end applications. For example, to keep the percentage of defective coils below 10 %, figure 7a reveals the Phosphorous content must be below 0.011 % wt, when running the Hot Mill at the mean values for all the other process variables (see table 5 for these mean values). Again, to keep the percentage of defective coils below 10 %, figure 7d reveals the Nickel content must be higher than 0.013 % wt., when running the Hot Mill at the mean values for all the other process variables. As a final example, figure 7f shows that to keep the percentage of defective coils below 10 %, the first principal component must be below 1.9 when running the Hot Mill at the mean values for all the other process variables.

Using Eq. (4.9a) and $PC_1 = 1.9$, a particular Hot Mill temperature can be calculated assuming the process runs at the mean values for all the other temperature variables at the Hot Mill. These calculated temperatures are shown in figure 7h. It can be seen that to keep the percentage of defective coils below 10 %, the maximum Rougher Mill temperature must be kept at 1040 °C or less. When looking at the average Rougher Mill temperature the temperature should be 1025 °C or less. The corresponding temperatures for the maximum, average and minimum Crop Shear temperatures are 1020 °C, 970 °C and 540 °C respectively. The corresponding temperatures for the maximum, average and minimum Finishing Mill temperatures are 870 °C, 852 °C and 410 °C respectively. No conditions could be found for the minimum Rougher Mill temperatures or the Run Out Table temperature when all other temperatures are at the mean values shown in table 5.

This type of calculation can be done for any PC_1 value, and therefore for any frequency of significant scale formation to indicate what temperatures to run the Hot Mill at to achieve this required frequency of scale formation. The final impression to be gleamed from figure 7 is that the variables to keep under strict control, so as to remove the possibility of very high proportions of coils being produced with a significant number of defects (probabilities above

50 % for example), are Phosphorous, Copper and the various temperatures as summarised in the first principal component.

4.7 Logit Conclusions

The principal component analysis above showed that the temperatures at the Hot Mill were all highly correlated. However, nearly 75 % of the variation in the ten Hot Mill temperatures could be accounted for by just four principal components that were completely uncorrelated with each other. Both the Logit and Probit models were supported by the collected data set and when a Logit model was estimated, only the first two principal components were statistically significant at the 1 % significance level, together with just five chemistries – Phosphorous, Silicon, Copper, Nickel and Chromium. The Logit model also had a correct prediction rate of 73 %.

It appears from this analysis that the easiest way to reduce scale formation is to lower the temperature entering the Finishing Mill. It was observed that average Crop Shear temperatures greater than 1050 °C have a significantly higher percentage chance of forming scale at levels greater than a count of 200 on the bottom surface. Therefore, reducing the average temperature entering the Finishing Mill will reduce the amount of scale formed on the surface of the coil.

The most practical way of achieving this is to reduce the amount of variability of the temperature. This reduction of the standard deviation avoids the issue of failure to achieve finishing temperature requirements for mechanical properties. With increased temperature control the average temperature could be reduced with no detrimental impact.

4.8 References

Burt, C.L., 1945. How the Mind Works. Allen & Unwin, London.

Cattell, R.B., 1952. Factor Analysis. Harper, New York.

Childs, D., 1970. The Essentials of Factor Analysis. Holt, Rinehart and Winston, London.

Jolliffe, I.T., 2002. Principal Component Analysis, 2nd edition. Springer-Verlag, New York.

Lawless, J.F., 2003. Statistical Models and Methods for Lifetime Data, 2nd edition.
John Wiley & Sons, New Jersey.

Maddala, G.S., 1991. Introduction to Econometrics, 3rd edition. John Wiley & Sons Ltd., New York, p. 322.

Mardia, K.V., Kent, J.T., Bibby, M., 1979. Multivariate Analysis. Academic Press, London.

Prentice, R.L., 1975. Discrimination among some parametric models. Biometrika 62, 607.

Chapter 5. A Partial Least Squares Generalized Linear Regression Algorithm

The Partial Least Squares (PLS) regression method in its basic form applies for one single response variable y and is non-iterative. It is particularly useful when the x_j variables are closely correlated with each other. Marx (1996) proposed a generalization of the PLS algorithm to generalized linear models (see McCullagh and Nelder (1989) for an excellent account of generalized linear models). This approach is very useful because the binary logit model can be given a generalised linear regression representation. Marx (1996) used the fact that, in the context of the exponential family, maximum likelihood estimates are obtained by an iterative weighted least squares procedure. Marx's approach consisted of replacing the iterative weighted least squares step by a sequence of PLS regressions.

5.1 The Generalised Linear Model

Suppose that a sample of i = 1 to n specimens have been scanned for scale formation on the ROT of the Hot Mill. Then let y_i be the binary response that equals unity when significant scale (defined as a scale count for the Parsytec system in excess of 200) is detected on the bottom of the tested specimen and zero otherwise. Next let π_i be the probability of significant scale formation and also let $x_{1i},..,x_{pi}$ be all the explanatory or process variables described in table 5. Finally, let $g(\pi_i)$ be the link function. Then the generalized linear regression model has the following form:-

$$g(\pi_i) = \sum_{j=1}^{p} \beta_j x_{ji}$$

Eq. (5.1a)

where

$$\eta_i = g(\pi_i) = ln\left[\frac{\pi_i}{1-\pi_i}\right]$$

Eq. (5.1b)

and where there are p process variables. The model parameters, β_j, can be estimated using the iterative weighted least squares procedure put forward by McCullagh and Nelder (1989). This procedure involves the creation of an adjusted dependent variable z_i with components

$$z_i = \eta_i + (y_i - \pi_i)\left[\frac{d\eta_i}{d\pi_i}\right]$$

Eq. (5.2a)

If then, the following (nxn) diagonal matrix of weights is calculated

$$W = diag\left[\frac{\left(\frac{d\eta_i}{d\pi_i}\right)^2}{\pi_i(1-\pi_i)}\right]$$

Eq. (5.2b)

then the following standard weighted least squares formula can be used to estimate the β_j

$$\hat{\beta} = (X^T W X)^{-1} X^T W Z$$

Eq. (5.2c)

where X is a (nxp) matrix of the process variables $x_1, \ldots, x_p$ such that the first column of X contains the n values for x_1 and the last column of X contains the n values for x_p. X^T is the transpose of X and Z is an (nx1) vector containing the n values for z_i as defined by Eq. (5.2a). β is a (px1) vector containing the estimates for β_1 to β_j. This has to be an iterative solution where initial guesses for β_1 to β_j are inserted into Eq. (5.1a-b) to enable values for z_i and W to be calculated from Eq. (5.2a-b). Eq. (5.2c) then gives revised or improved estimates for β_1 to β_j. These revised values are then used to recalculate values for z_i and W from which further revised estimates for β_1 to β_j are obtained. This iterative process continues until the changes in all the β_j values are very small.

5.2 Partial Least Squares within a Generalised Linear Model

Within this setting, the following Partial Least Squares (PLS) generalized linear regression algorithm can be stated as follows. This algorithm follows closely that given by Bastien and Tenenhaus (2005). Let X_0 be the matrix containing the n values for the standardized input variables $x_{01i}, \ldots, x_{0p}$

$$x_{0ji} = \left[\frac{x_{ji} - \hat{x}_j}{s_j}\right] \quad j = 1, p \; and \; i = 1, n$$

Eq. (5.3)

where $\overline{x}_j$ is the sample average for the n values on variable x_j and s_j the corresponding sample standard deviation (the last two columns of table 5 show these two statistics for each process variable). Then, the first column of X_0 contains the n values for x_{01}, and the last column of X_0 contains the n values for x_{0p}. Within this setting the PLS components can be derived as follows:

5.2.1 Determination of the first PLS component t_{1i}:

1. For each j = 1 to p, compute the regression coefficient β_{1j} of x_{0j} in the generalized linear regression model given by Eq. (5.1a-b) of y_i on x_{0ji} using the iteratively weighted least squares procedure described above.

2. For each j = 1 to p, compute the standard error, se_{1j}, of the regression coefficient β_{1j} of x_{0j} in the generalized linear regression model given by Eq. (5.1a-b) of y_i on x_{0ji} using the iteratively weighted least squares procedure described above. Then compute the student t statistic $v_{1j} = \beta_{1j} / se_{1j}$.

3. Form the vector $\beta_1 = (\beta_{11},\ldots,\beta_{1p})^T$, and a (pxp) diagonal matrix of weights $v_1 = diag(v_{1j})$ and the (px1) vector $v_1\beta_1$.

4. Compute the vector $t_1 = X_0 v_1 \beta_1$. t_1 is a vector made up of the n values of t_{1i} and constitutes the first PLS component.

This PLS component is essentially a weighted average of the p predictors of y_i, namely a weighted average of the $\beta_{1ji}x_{0ji}$ series. A cruder but simpler version of this is to remove from β_1 those coefficients that are found to be statistically insignificant at the 1 % significance

level using the student t test, and of course the corresponding process variable from X_0, when forming the vector t_1. (The justification being that v_{1j} will be small if the student t value for $\beta_{1j} = 0$ is less than the 1 % significance level).

5.2.2 Determination of the second PLS component t_{2i}:

1. Compute the residuals from the set of linear regressions of each column of X_0 on t_1 and store these residuals in the matrix X_1. (For example, x_{1j} would be the first column in X_1 and would contain the n residuals derived from the linear regression of x_{01i} on the t_{1i}.)

2. For each j = 1 to p, compute the regression coefficient β_{2j} of x_{1j} in the generalized linear regression model of y_i on t_{1i} and x_{1ji} using the iteratively weighted least squares procedure described above.

3. For each j = 1 to p, compute the standard error, se_{2j}, of the regression coefficient β_{2j} of x_{1j} in the generalized linear regression model of Eq. (5.1a-b) of y_i on x_{1ji} using the iteratively weighted least squares procedure described above. Then compute the student t statistic $v_{2j} = \beta_{2j}/ se_{2j}$.

4. Form the vector $\beta_2 = (\beta_{21},\ldots,\beta_{2p})^T$, a (pxp) diagonal matrix of weights $\mathbf{v}_2$ = diag(v_{2j}) and the (px1) vector $v_2\beta_2$.

5. Compute the vector $t_2 = X_1 v_2 \beta_2$. t_2 is a vector made up of the n values of t_{2i}. It is essentially a weighted average of the p predictors of y_i given in step 1. Again a cruder but simpler version of this is to simply delete from β_2 those coefficients that are found to be statistically insignificant at the 1 % significance level using the standard student t test, and the corresponding transformed process variable from X_1, when forming the vector t_2.

5.2.3 Determination of the other PLS components:

This procedure is iterated for the other PLS components t_{hi}. At each step the generalized linear regression of y_i on components $t_{1i},\ldots, t_{hi}$ is carried out. The procedure is stopped, and the component t_{m+1i}, not included in the model if that component is not statistically significant at the 5 % significance level using the standard student t test. The final regression equation is obtained by expressing the generalized linear regression of y_i on $t_{1i},\ldots, t_{mi}$ and this can also be written as a function of the original variables by making use of the standardisation procedure given in Eq. (5.3) and the means and standard deviations given in table 5. In the case of

ordinary multiple regression, this algorithm gives the usual PLS regression when there is no missing data.

5.3 PLS Results and discussion

5.3.1 The PLS Components and the PLS Logit Model

Table 10 shows the estimated values for the β_{1j} weights of the first PLS components associated with all the different alloying elements. (See table 5 for a description of the shorthand used for each of the process variables listed in table 10).

Table 10. Estimated values for the β_{1j} weights for all alloying elements.

Process variables	Logit slope coefficient, β_{1j}	Elasticity (%)	Count R^2 (%)
x_1	-0.0298 [-0.59]	-0.24	49.08 (54.26, 46.56)
x_2	-0.0807 [-1.60]	-0.87	50.35 (57.56, 46.84)
x_3	0.6371 [11.15]#	1.43	65.57 (56.98, 69.75)
x_4	0.1065 [2.11]	0.25	57.64 (46.32, 63.18)
x_5	0.1713 [3.31]#	0.14	60.88 (35.47, 73.23)
x_6	-0.0585 [-1.16]	-0.08	47.69 (58.91, 42.22)
x_7	-0.2122 [-4.14]#	-0.39	53.52 (45.54, 57.40)
x_8	-0.0937 [-1.85]	-0.21	50.86 (53.88, 49.39)
x_9	-0.1892 [-3.71]#	-0.68	56.50 (54.84, 57.30)
x_{10}	-0.2360 [-4.50]#	-0.19	48.19 (75.58, 34.87)
x_{11}	-0.0507 [-1.00]	-0.35	48.19 (50.39, 47.13)
x_{12}	-0.1935 [-3.80]#	-0.7	54.85 (57.56, 53.53)

Student t value for testing the null hypothesis that $\beta_{1j} = 0$ shown in parenthesis.

Coefficient is significantly different from zero at the 1 % significance level.

Elasticity measures the percentage change in the probability of observing scale resulting from a one percentage point change in the value of process variable x_j. This elasticity is calculated using $\pi = 0.5$ and using the values for x_j at this probability. Count R^2 shows the percentage number of correct predictions which is further broken down, in brackets, into the percentage number of $y_i = 1$ and $y_i = 0$ correct predictions respectively.

Only the slope coefficients in front of the elements P, Si, Ni, Al, Sn and solAl (x_3, x_5, x_7, x_9, x_{10}, x_{12}) are statistically significant at the 1 % significance level. Of these elements, increases in the amount of P, S and Si (x_3, x_4 , x_5) lead to an increase in the probability of scale formation, whilst increases in the other elements had a negative impact on scale formation. Further, of these statistically significant elements only P (and possible Al and solAl) seems to have predictive power for both $y_i = 0$ and $y_i = 1$ (as measured by Count R^2) above that given by purely random predictions. The estimated elasticities are however quite small (especially when compared to the temperature variables in table 11 below), with for example a 1 % increase in Phosphorous content bringing forth a 1.43 % increase in the chances of significant scale forming (when this probability is already at 0.5).

Table 11. Estimated values for the β_{1j} weights for various temperatures.

Process variables	Logit slope coefficient, β_{1j}	Elasticity (%)	Count R^2 (%)
x_{13}	0.7274 [11.84]#	22.83	63.28 (72.67, 58.72)
x_{14}	0.0299 [0.59]	1.18	55.49 (38.57, 63.72)
x_{15}	0.8002 [12.80]#	27.18	64.24 (72.67, 60.14)
x_{16}	0.7175 [11.74]#	19.84	64.55 (71.51, 61.17)
x_{17}	0.2140 [3.98]#	2.1	53.77 (68.99, 46.37)
x_{18}	0.7465 [10.76]#	19.19	64.23 (75.00, 59.00)
x_{19}	0.5033 [8.85]#	31.78	61.00 (55.81, 63.52)
x_{20}	0.2205 [4.29]#	3.16	49.78 (71.51, 39.21)
x_{21}	0.5851 [11.44]#	44.85	65.38 (66.67, 64.75)
x_{22}	-0.2130 [-3.88]#	-3.1	54.60 (41.28, 61.07)

Student t value for testing the null hypothesis that $\beta_{1j} = 0$ shown in parenthesis.

Coefficient is significantly different from zero at the 1 % significance level.

Elasticity measures the percentage change in the probability of observing scale resulting from a one percentage point change in the value of process variable x_j. This elasticity is calculated using $\pi = 0.5$ and using the values for x_j at this probability. Count R^2 shows the percentage number of correct predictions which is further broken down, in brackets, to the percentage number of $y_i = 1$ and $y_i = 0$ correct predictions respectively.

Table 11 shows the estimated values for the β_{1j} weights of the first PLS components associated with all the different temperatures. Only the minimum Rougher Mill temperature appears to be statistically insignificant at the 1 % significance level and all but the average Run Out Table temperature has the expected sign - with increases in temperature bringing forth an increase in the probability of significant scale forming. The average Finishing Mill temperature appears to have the biggest effect, with a 1 % increase in this temperature bringing forth a 44.85 % increase in the chances of significant scale forming (when this probability is already at 0.5). In terms of the Count R^2 values, the average Rougher and Finishing Mill temperatures have the best predictive capabilities for both $y_i = 1$ and $y_i = 0$.

Using the algorithm described in Section 5.2 above and the values shown in table 10 and table 11, the first PLS component, written out in terms of the standardised values for the process variables, can be computed as follows

$$\begin{aligned} t_{1i} = {} & -0.0176x_{01i} - 0.1291x_{02i} + 7.1037x_{03i} + 0.2247x_{04i} + 0.5670x_{05i} - 0.0679x_{06i} \\ & - 0.8785x_{07i} - 0.1734x_{08i} - 0.7019x_{09i} - 1.062x_{010i} - 0.0507x_{011i} \\ & - 0.7353x_{012i} + 8.6124x_{013i} + 0.0176x_{014i} + 10.2427x_{015i} + 8.4235x_{016i} \\ & + 0.8517x_{017i} + 8.0323x_{018i} + 4.4542x_{019i} + 0.9459x_{020i} + 6.6935x_{021i} \\ & - 0.6603x_{022i} \end{aligned}$$

Eq. (5.4a)

The first PLS component, written out in terms of the process variables in their original units, can also be computed as follows

$$\begin{aligned} t_{1i} = {} & -3743.73 - 4.09x_{1i} - 5.71x_{2i} + 2258.17x_{3i} + 75.07x_{4i} + 175.02x_{5i} - 11.51x_{6i} \\ & - 293.87x_{7i} - 43.21x_{8i} - 149.65x_{9i} - 575.44x_{10i} - 55.94x_{11i} \\ & - 170.32x_{12i} + 0.4819x_{13i} 0.0015x_{14i} + 0.6329x_{15i} + 0.4235x_{16i} \\ & + 0.0175x_{17i} + 0.3862x_{18i} + 0.6201x_{19i} + 0.0327x_{20i} + 1.1647x_{21i} \\ & - 0.0263x_{22i} \end{aligned}$$

Eq. (5.4b)

Then when t_{1i} is used instead of all the process variables, the following PLS logit model is obtained using iterativly weighted least squares

$$\eta_i = ln\left[\frac{\pi_i}{1-\pi_i}\right] = 0.0286t_{1i} \quad [15.69] \quad Count\ R^2 = 70.32\,\%$$

Eq. (5.4c)

The student t value given in parenthesis shows that this PLS component is significantly different from zero at the 1 % significance level and so is an important predictor variable of significant scale formation. This PLS model predicts the formation of scale or lack of scale with approximately 70 % accuracy, which given the very large variability in the data is a good outcome. Further, the Count R^2 associated with $y_i = 1$ suggests that the model can predict the processing conditions that lead to excessive scale with 80.43 % accuracy (but the Count R^2 associated with $y_i = 0$ is lower at 65.41 %).

Using the above algorithm, the next PLS component using the standardised process variables was identified as

$$\begin{aligned} t_{2i} = {} & -0.1729x_{11i} - 0.0207x_{12i} + 6.29x_{13i} + 0.2287x_{14i} + 0.1464x_{15i} - 0.1043x_{16i} \\ & - 0.4650x_{17i} - 0.2017x_{18i} - 0.4807x_{19i} - 0.5822x_{110i} - 0.0011x_{111i} \\ & - 0.4740x_{112i} - 3.1391x_{113i} + 0.00001x_{114i} - 0.7965x_{115i} - 5.1901x_{116i} \\ & - 0.3543x_{117i} - 0.3668x_{118i} - 0.8308x_{119i} + 0.0357x_{120i} - 0.0562x_{121i} \\ & - 0.7789x_{122i} \end{aligned}$$

Eq. (5.5a)

or in terms of the process variables in their original units this can be re-written as

$$t_{2i} = -74.36 - 41.14x_{1i} - 2.07x_{2i} + 2458.89x_{3i} + 91.61x_{4i} + 80.62x_{5i} - 20.03x_{6i} - 215.05x_{7i} - 59.07x_{8i} - 132.79x_{9i} - 431.94x_{10i} - 12.51x_{11i} - 144.26x_{12i} - 0.0781x_{13i} + 0.0003x_{14i} + 0.0789x_{15i} - 0.1752x_{16i} - 0.0037x_{17i} + 0.0605x_{18i} + 0.0099x_{19i} + 0.0078x_{20i} + 0.2260x_{21i} - 0.0363x_{22i}$$

Eq. (5.5b)

The extended PLS logit model encompassing this extra component is then estimated by iterative weighted least squares to be

$$\eta_i = ln\left[\frac{\pi_i}{1-\pi_i}\right] = 0.0315t_{1i} + 0.0886t_{2i} \quad [15.69]\ [11.44]\ \ Count\ R^2 = 73.30\ \%$$

Eq. (5.5c)

The student t values, given in parenthesis, shows that both these PLS components are significantly different from zero at the 1 % significance level and so are an important predictor variables of scale formation. This PLS model predicts the formation of scale or lack of scale with approximately 73 % accuracy and so the addition of the second PLS provides an additional 3 % to overall predictive accuracy of the previous model given by Eq. (5.4c). Further, the Count R^2 associated with $y_i = 1$ suggests that the model can predict the processing conditions that lead to excessive scale formation with 84.69 % accuracy (but the Count R^2 associated with $y_i = 0$ is lower at 67.77 %).

Using the above algorithm, the third PLS components was identified as

$$t_{3i} = -0.5834x_{21i} - 0.0506x_{22i} - 2.1281x_{23i} + 0.0101x_{24i} + 0.3162x_{25i} + 0.0005x_{26i} - 0.3368x_{27i} - 1.0641x_{28i} - 1.0751x_{29i} - 0.1006x_{210i} - 0.0129x_{211i} - 1.0759x_{212i} - 0.0041x_{213i} - 0.0049x_{214i} + 0.0886x_{215i} + 0.1006x_{216i} - 0.0473x_{217i} + 0.00001x_{218i} - 0.3424x_{219i} + 0.0761x_{220i} - 0.0035x_{221i} - 0.3305x_{222i}$$

Eq. (5.6a)

with the extended PLS logit model

$$\eta_i = ln\left[\frac{\pi_i}{1-\pi_i}\right] = 0.0316t_{1i} + 0.0888t_{2i} + 0.0321t_{3i} \quad [15.98]\,[11.45]\,[1.61] \quad Count\ R^2 = 73.50\ \%$$

Eq. (5.6b)

The student t values given in parenthesis shows that the first two PLS components are significantly different from zero at the 1 % significance level, but the third component is only statistically significant at the 10 % significance level. Further, this PLS logit model only has an additional 0.2 % overall predictive accuracy compared to that given by Eq. (5.5c). Taken in conjunction, these two facts suggest that the PLS logit model with just two components is a valid parsimonious model, i.e. the model given by Eq. (5.5c).

5.4 Analysis and Interpretation

Figure 8 plots the scale indicator variable y_i against the PLS component t_{1i}. Also shown are the predicted probabilities of scale formation given by the above parsimonious PLS logit model. These predictions are consistent with the spread of y_i values shown on figure 8. When the second PLS component is zero, scale formation becomes more likely than not when the first principle component exceeds zero in value. At this value for t_{1i}, a change in t_{2i} has a big impact on the probability of scale formation.

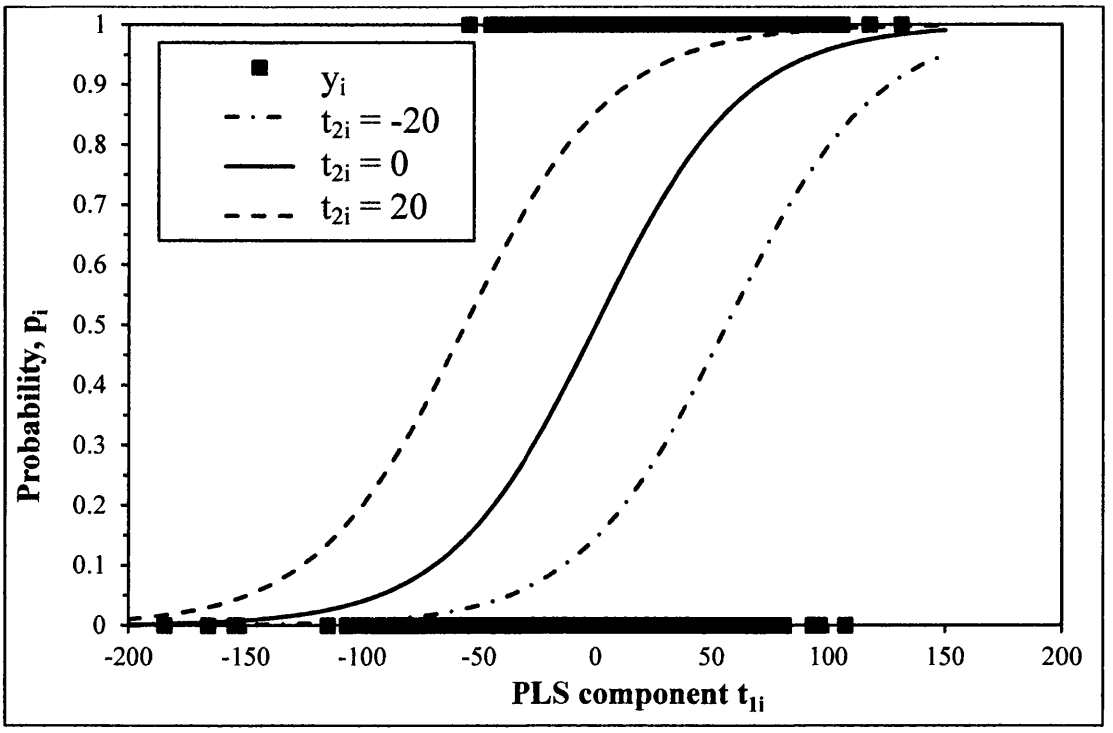


Figure 8. Predicted Probabilities of Scale as a Function of the PLS Components.

Figure 9 plots the first two PLS components against each other with the data points shaded or un-shaded depending upon whether the scale indicator variable is zero or unity. It is clear from figure 9 that samples with significant scale formation tend to correspond to those with higher t_{1i} values – more specifically with positive t_{1i} values. Variations in the values for t_{2i} are a far less indicative of whether scale has formed or not. These positive t_{1i} values can be traced back to the specific operating conditions. Clearly, many different conditions can be identified to achieve a required rate of scale formation, but figure 10 and table 12 give an indication of what can be achieved.

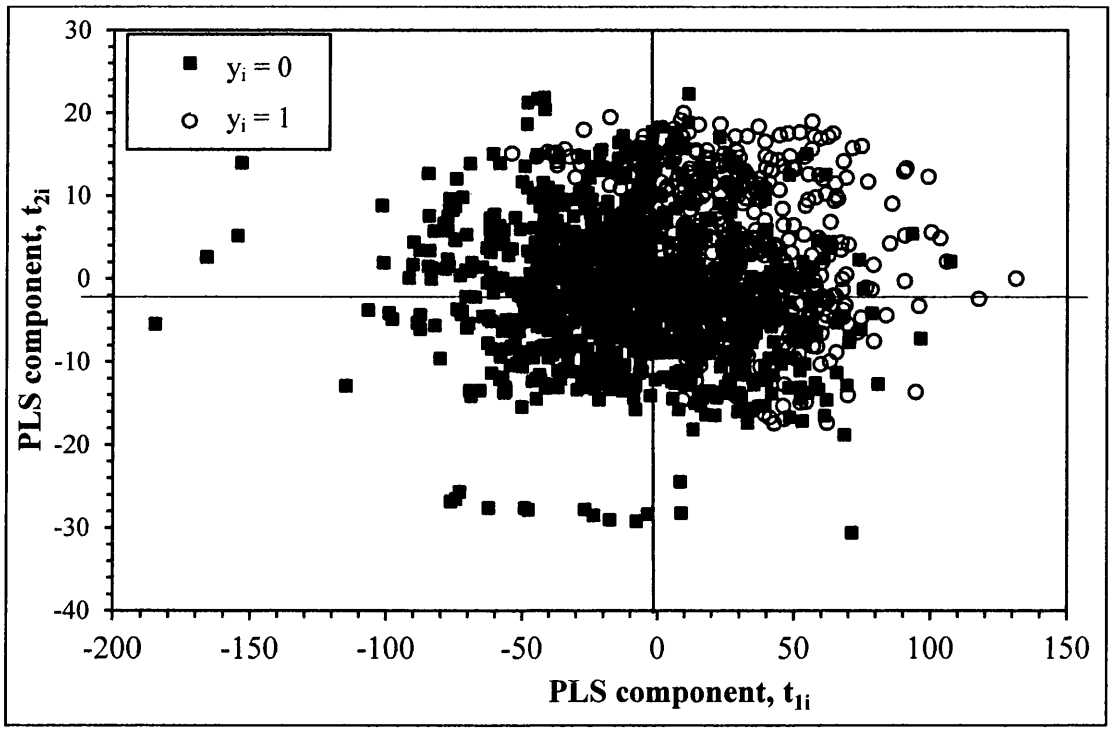


Figure 9. Cross Plot of the PLS Components.

Figure 10 shows the elasticity's associated with each process variable as calculated around the mean values for each process variables shown in table 12. These elasticity's give the percentage change in the probability of scale formation following a 1 % change in that process variable from its mean value. It can be seen from figure 10 that the most important process variable appears to be the average Finishing Mill temperature. When this temperature is increased by 1 % from its mean of 881 °C (i.e. from 881 °C to 890 °C), whilst keeping all the other process variables at their mean values shown in table 5, the chances of significant scale forming as a result of this change, increases by 25 %, i.e. from 0.5 to 0.63. Subsequently, the most important variables appear to be the average Rougher Mill temperature with an elasticity of approximately 15 %. In importance, this is closely followed by the average Crop Shear temperature and the maximum Finishing Mill temperature whose elasticity's are both approximately 10 %. The only alloying element with an elasticity above 2 % is Phosphorous. With an elasticity of 2.04 %, when this element is increased by 1 % from its mean of 0.014 % (i.e. from 0.014 to 0.01414 %), whilst keeping all the other process variables at their mean values shown in table 5, the chances of significant scale forming as a result of this change, increases by 2.04 %, i.e. from, 0.5 to 0.51. This is clearly a much smaller effect than that associated with the above mentioned temperature variables.

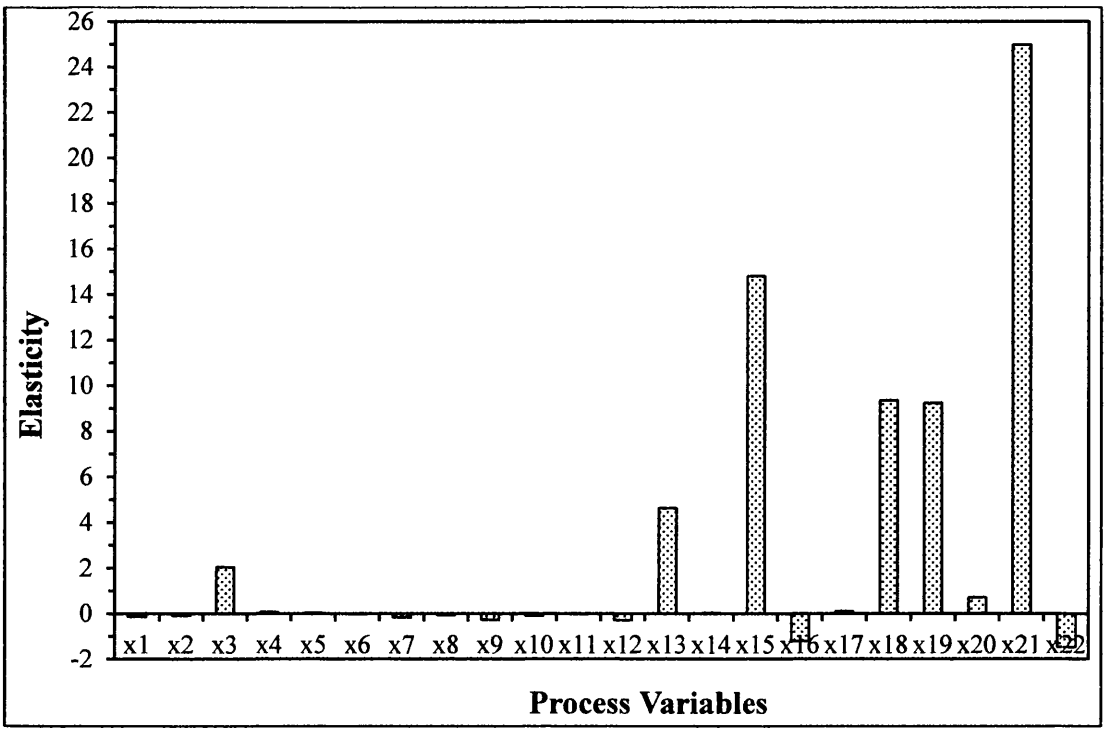


Figure 10. Calculated Elasties for Each Process Variable.

Next, table 12 shows what the process variables with the largest elasticity's should be set at so as to achieve a required low probability of scale formation when all the other process variables are set at their average values shown in table 5. The low probabilities looked at in table 12 are 0.3 down to 0.1. Looking for example at the Finishing Mill temperature (x_{21}), to reduce the probability of significant scale forming from 0.5 (when all process variables are set at the average values shown in table 5) to 0.1, this temperature must be set as low as 842 °C (compared to the sample average value of 881 °C). The same low probability of scale formation can be achieved by lowering the Phosphorous content to 0.0065 % (from the sample average of 0.0141 %) whilst holding all the other process variables at the sample average values shown in table 5. The values calculated in table 12 are theoretical values calculated using the PLS model, and should be considered in conjunction with other considerations; such as achieving the required mechanical properties and to avoid situations such as ferritic rolling.

Table 12. Examples of using the PLS logit model for process control.

Variable/Probability	x_3 (%)	x_{13} (°C)	x_{15} (°C)	x_{18} (°C)	x_{19} (°C)	x_{21} (°C)
0.1	0.0065	855.8	1017.8	944.2	799.5	842.3
0.2	0.0093	953.9	1047.9	990.5	839.2	856.6
0.3	0.0112	1019.2	1067.9	1021.2	907.2	866.1
Sample averages	**0.0141**	**1121.7**	**1099.4**	**1069.5**	**907.2**	**881**

A final insight into the estimated PLS logit model can be obtained using cumulative frequency plots. This cumulative frequency plot is constructed by sorting the sample data set, from lowest to highest, by the variable of interest and counting how many observations there are at or below various values for that variable of interest. This frequency count can then be plotted alongside the frequency count implied by the probabilities predicted from Eq. (5.5c). Figure 11 shows the cumulative frequency plot for the average Roughing Mill temperature. It can be seen that the observed and predicted values show similar trends. The model correctly predicts when scale starts to become an issue (around 1080 °C) and follows the observed frequency count until approximately 1120 °C.

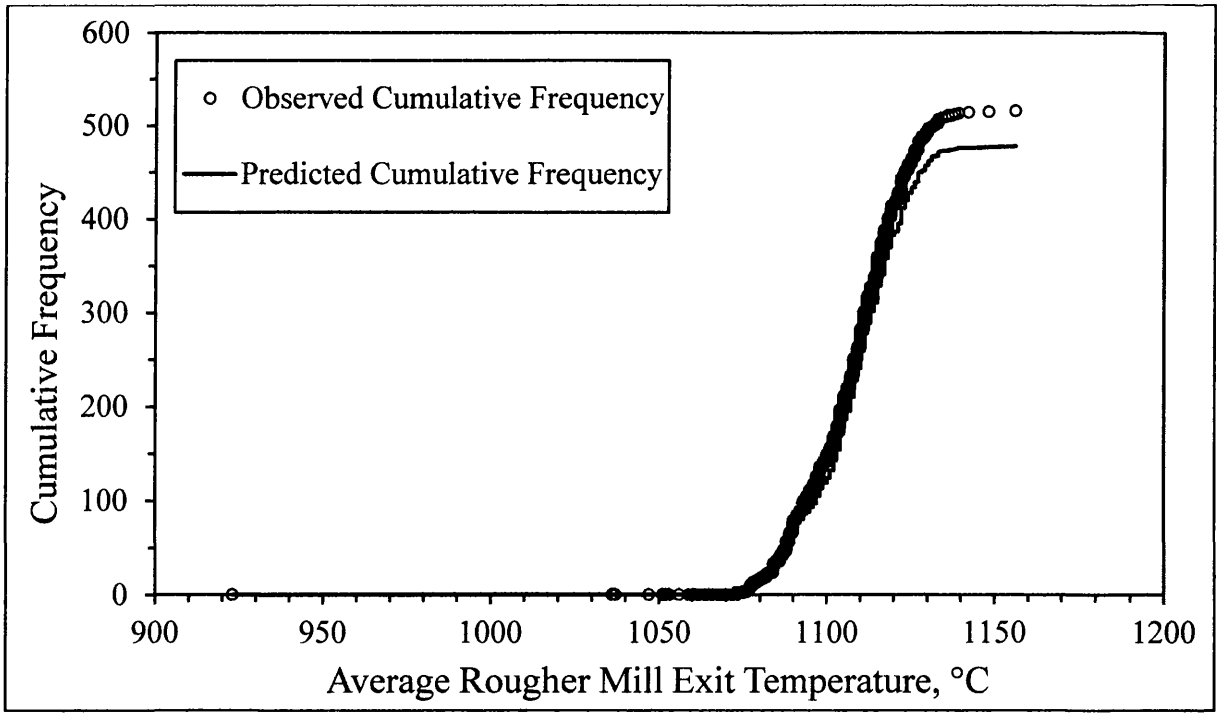


Figure 11. Cumulative Frequency Plot for the Average Rougher Mill Exit Temperature.

Figure 12 shows the cumulative frequency plot for the average Crop Shear temperature. It can be seen that the observed and predicted values do not correspond to each other in comparison with the other temperature variables. The predicted results over predict when scale starts to become an issue (1070 °C instead of approximately 1060 °C) and follows the observed line until about 1090 °C.

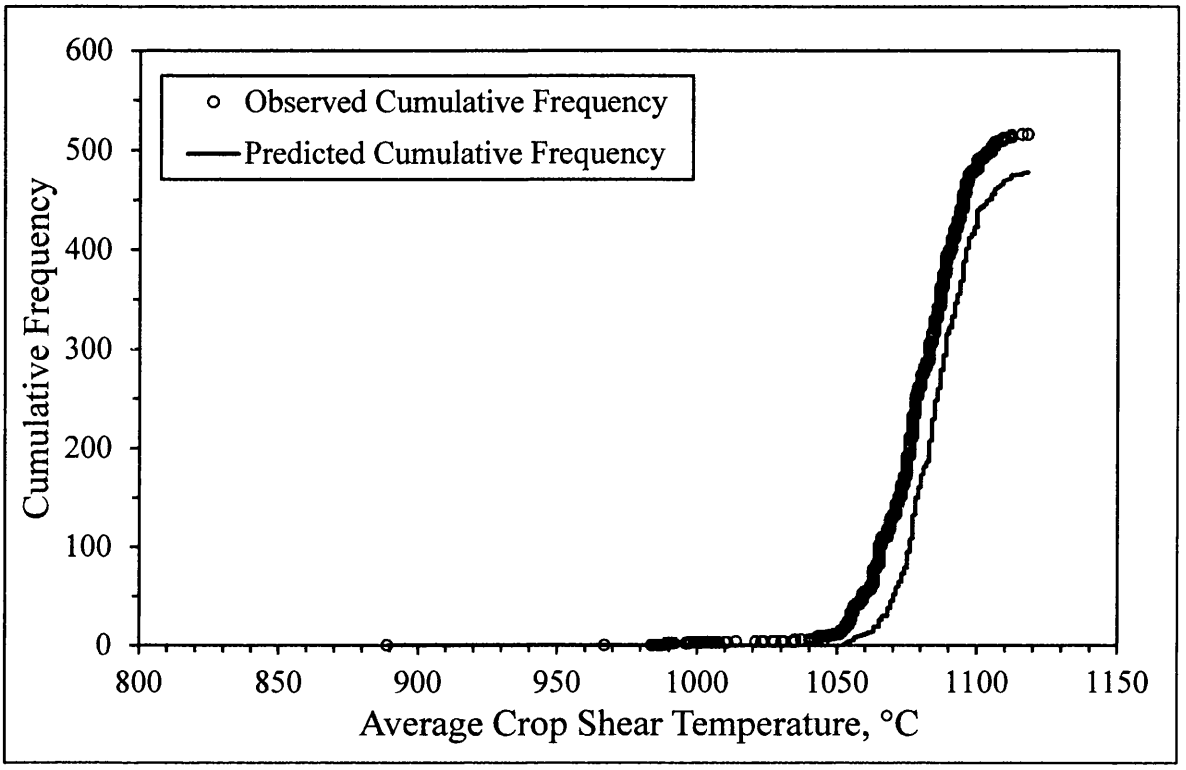


Figure 12. Cumulative Frequency Plot for the Average Crop Shear Temperature.

5.5 PLS Conclusions

Using a non-linear PLS logit model, it was found that there exists only two linear PLS components that were important and statistically significant in determining scale formation in this data set obtained from the Hot Mill at Tata Steel Port Talbot. The estimated model was capable of correctly predicting when $y_i = 1$ (i.e. when significant scale formation had occurred) 85 % of the time. The good predictive capability was also confirmed by the cumulative frequency plots, where it was observed that the predictions from the model with respect to those temperatures with a high elasticity corresponded well with the observed values. This was especially true for the temperature at which scale is first observed to occur.

The chances of scale formation also appeared to increase with increasing values of the first PLS component. This trend was then translated into suggested best practice operating conditions. For example, if the Finishing Mill temperature was set at 842 °C, with all other process variables set at the average values shown in table 5, the model predicted the chances of significant scale forming can be kept as low as 10 %. The elasticities associated with the non linear PLS logit model also showed that the main determinants of scale formation were the various temperatures. The elasticities derived from the model showed that the minimum temperatures have significantly less predictive power than the maximum and average temperatures at the various stages of operation within the Hot Mill. The most significant temperatures for scale formation were the average temperatures at the Rougher and Finishing Mills, as well as the average Crop Shear temperature. It was also found that the only chemistry element that had a substantial role to play in scale formation was the amount of Phosphorus present in the coil, which could be increasing the adhesion of the scale to the coil substrate.

5.6 References

B. D. Marx: "Iteratively Reweighted Partial Least Squares Estimation for Generalized Linear Regression". *Technometrics*, 1996, Vol. 38, No. 4, pp. 374-381.

P. McCullagh, P., and J. A. Nelder: *Generalised Linear Models*, 1989, 2nd Ed. Chapman & Hall, London.

P. Bastien and M. Tenenhaus: "PLS Generalized Linear Regression: Application to the Analysis of Life Time Data". *Computational Statistics & Data Analysis*, 2005, Vol. 48, Issue: 1, pp. 17-46.

Chapter 6. Process parameters influencing tertiary scale formation at a hot strip mill using a multinomial logit model

6.1 Introduction

An alternative to using the scale count is to classify scale count into groups. Whilst this also results in some loss of information, it can often lead to more interpretable models and give predictions in the form of probabilities of scale forming under different operating conditions - which can be useful for process control due to large variability. At one extreme is the binary classification where an indicator variable replaces scale count with this variable equalling unity when scale count exceeds the chosen value and zero otherwise. This is a simple scale – no scale classification. This classification results in the largest amount of lost information relative to the count data itself but is appropriate if scale counts less than this value is unlikely to be a problem for further processing (and all the costs associated with that).

Whilst the above binary approach can identify conditions most likely to lead to counts below 200 and therefore where no further processing is required, it cannot predict conditions leading to other amounts of scale forming. Where for example further processing is required but the cost of this is not too excessive. As the cost of further processing tends to increase with an increasing scale count, finer and more detailed process control often requires a broader classification of scale count. For example, the scale count data could be categorised as low, medium and high. Note that as the scale classification is broadened this type of approach tends to that obtained using a Poisson distribution for count data. This paper uses a broader classification than the binary one which can be seen as a balanced approach between the extremes of a binary classification and no classification – where the raw count data itself is used.

6.2 MLR data classification

Table 13 summarises the categories of scale count used for the dependent variable for the MLR model. Essentially, a coil is classified as having either a very low scale count, a low scale count, a moderate scale count or a high scale count. An indicator variable Y_i is used to summarise the category into which each of the individual manufactured coils falls into. $Y_i = 0$ when a coil has between 1 and 10 scale counts on its bottom surface, $Y_i = 1$ when a coil has between 11 and 75 scale counts on its bottom surface, $Y_i = 2$ when a coil has between 76 and

350 counts of scale on its bottom surface and $Y_i = 3$ when a coil has over 350 scale counts on its bottom surface (the subscript i denotes an individual coil). These boundaries were chosen because they resulted in the most balanced distribution of coils within each grouping – with around 25 % of coils making up each category.

Table 13. Classification of splits for the MLR model.

Indicator Variable, $Y_i = k$	Boundaries for scale count	Number in category	% of Dataset in category
k = 0; Very Low Scale	0 - 10	367	23.30
k = 1; Low Scale	11 - 75	462	29.30
k = 2; Medium Scale	76 - 350	369	23.40
k = 3; High Scale	351+	379	24.00

6.3 Principle Component Analysis

For an explanation of PCA, see Section 4.1.

Once all p principle components have been derived two additional questions require answering. First, how many principal components are essential for representing all the process variables? Cattell (1952) has suggested that only those principal components having an Eigen value greater than 1 should be considered as essential and therefore retained in the analysis. Secondly, do all the process variables need to go into making up each principle component? Frequently, extraction communalities are used to answer this second question. An extraction communality measures the variance in the process variables (once standardised) that are accounted for by each principle component. As such, extraction communalities are equal to the squares of the loadings described above. A common rule of thumb is that the loadings should be 0.7 or higher to confirm that a process variable is well represented by a particular principal component. This paper will therefore use a cut off of 0.6 for the squared loadings and extracted communality values below this number are taken to indicate variables that do not fit well with the spectral decomposition and should not be included in the construction of the principal components.

6.4 MLR model

The multinomial model can be written as

$$Prob(Y_i = k) = \frac{exp\left(\beta_{0k} + \sum_{j=1}^{p} \beta_{jk} x_{ij}\right)}{\sum_{k=0}^{K} exp\left(\beta_{0k} + \sum_{j=1}^{p} \beta_{jk} x_{ij}\right)}, \quad k = 0, 1, \ldots, K; \quad i = 1, p$$

Eq. (6.2)

For the data set used in this paper, K = 3 (i.e. 4 scale categories), there are n = 1577 coils that have undergone a scale count performed and there are p = 39 process variables. However, there is an indeterminacy in the model as it is written in Eq. (6.2). This indeterminacy takes the form of the probabilities given by Eq. (6.2) being not unique because, with the probabilities having to sum to unity over the K categories at a given set of process variable values, only K-1 parameter vectors would be needed to determine the K probabilities associated with each scale category. A convenient normalisation that solves this problem is to set $\beta_{0k} = \beta_{1k} = \ldots. = \beta_{pk} = 0$ when k = 0, i.e. for the very low scale category. Then the probabilities are

$$Prob(Y_i = k) = P_{ik} = \frac{exp\left(\beta_{0k} + \sum_{j=1}^{p} \beta_{jk} x_{ij}\right)}{\sum_{k=0}^{K-1} exp\left(\beta_{0k} + \sum_{j=1}^{p} \beta_{jk} x_{ij}\right)}, \quad k = 0, 1, \ldots, K-1$$

Eq. (6.3)

This multinomial model implies that K -1 log – odds can be computed as

$$ln\left[\frac{P_{im}}{P_{ik}}\right] = \left(\beta_{0m} + \sum_{j=1}^{p} \beta_{jm} x_{ij}\right) \quad if \quad k = 0$$

Eq. (6.4)

Within this framework the binomial logit model used in chapter 4 results if K = 1.

6.5 Estimating a Multinominal Logit Model

From the point of view of estimation, it is useful that the odds ratio P_{im}/P_{ik} is not dependent on the other scale categories. The log likelihood to be maximised when estimating the β_{jm} parameters can be derived by defining for each coil, $d_{ik} = 1$ if coil (i) has a scale count placing it in the kth scale category and zero otherwise for the K possible scale categories. Then for each coil (i) one and only one of the d_{ik}'s is unity. Then the log likelihood is

$$\ln L = \sum_{i=1}^{n} \sum_{k=0}^{K-1} d_{ij} ln\{Prob(Y_i = k)\}$$

Eq. (6.5)

The estimated values for the parameters are taken to be those that maximise Eq. (6.5) which can be easily achieved using a standard nonlinear optimisation algorithm such as SOLVER in Excel 2010.

6.6 MLR Results and Discussion

6.6.1 Principle component testing procedure

Using principal component analysis the correlations existing between all the process variables was investigated. The analysis was performed to reduce both the number of process variables and to remove the correlations among the process variables used in the model to predict scale count. With the intention of gaining an understanding of how to run the Hot Mill so as to minimise scale formation. The temperature variables in table 5 were highly correlated with each other and so it was decided to restrict the principal component analysis to these variables. For example, the correlation coefficient between the average and maximum Rougher Mill temperatures was 0.92.

It was decided to use different variables in the construction of this PC, compared to the one for Chapter 4. This was undertaken to gain a greater understanding of the relationship between the temperatures. The variables used in the construction of the PC are listed in table 14.

Table 14. Variables used in the construction of the principle components.

Important Temperature Variables	
x_{26}	Furnace Aim Temperature Average Temperature, °C
x_{27}	Furnace Centre/Surface Difference, °C
x_{28}	Furnace Surface Temperature, °C
x_{29}	Furnace Centre Temperature, °C
x_{13}	Maximum RM Temperature, °C
x_{15}	Average RM Temperature, °C
x_{14}	Minimum RM Temperature, °C
x_{16}	Maximum CS Temperature, °C
x_{19}	Maximum FM Temperature, °C
x_{21}	Average FM Temperature, °C

An iterative procedure was adopted to identify the important principal components and the temperature process variables present within them. First, all the temperature variables were used and all principal components were extracted. If a variable had a squared loading below 0.6 in any of the principal components having an Eigen value above 1, that process variable was removed from the analysis. Second, with the reduced number of temperature variables, the principal component analysis was repeated and the same procedure as above used to eliminate further temperature variables. This process was repeated until there were no longer any squared factor loadings below 0.6 in value. Table 14 lists the ten important process variables identified by this iterative procedure and table 15 shows the Eigen values associated with these principal components. It can be seen from table 15 that the average and minimum Crop Shear temperatures, the minimum Finishing Mill temperature, the average ROT temperature and all the coiling temperatures appear to be unsuitable for inclusion in any of the principal components.

Table 15. Principal component results for the temperature process variables.

Components	Eigen values	Variation, %	Cumulative Variation, %
PC_1	3.977	39.771	39.771
PC_2	2.165	21.649	61.42
PC_3	1.283	12.826	74.247
PC_4	1.058	10.583	84.829
PC_5	0.807	8.074	92.903
PC_6	0.289	2.889	95.792
PC_7	0.227	2.265	98.057
PC_8	0.098	0.979	99.036
PC_9	0.069	0.693	99.73
PC_{10}	0.027	0.27	100

Further, table 15 reveals that the first four principal components accounted for nearly 84 % of the variability present in all ten temperature process variables. Using Cattell's rule (Cattell (1952)) of only needing to use those principal components having an Eigen value of around unity or more, the ten temperature process variables can therefore be replaced by the first four principal components. These four principal components had the following form;

$$PC_1 = 0.0168 * (x_{26} - 4.78) + 0.0067 * (x_{27} - 44.96) + 0.0153 * (x_{28} - 1227.13) + 0.0065 * (x_{29} - 1182.48) + 0.0242 * (x_{13} - 1121.75) + 0.0259 * (x_{15} - 1099.43) + 0.0014 * (x_{14} - 918.34) + 0.0219 * (x_{16} - 1099.88) + 0.0489 * (x_{19} - 907.16) + 0.0636 * (x_{21} - 881.05)$$

Eq. (6.6a)

$$PC_2 = -0.0361 * (x_{26} - 4.78) + 0.0212 * (x_{27} - 44.96) + 0.0029 * (x_{28} - 1227.13) - 0.0366 * (x_{29} - 1182.48) + 0.0059 * (x_{13} - 1121.75) + 0.0026 * (x_{15} - 1099.43) + 0.0021 * (x_{14} - 918.34) + 0.0059 * (x_{16} - 1099.88) + 0.0049 * (x_{19} - 907.16) - 0.0119 * (x_{21} - 881.05)$$

Eq. (6.6b)

$$PC_3 = -0.0217 * (x_{26} - 4.78) - 0.0176 * (x_{27} - 44.96) - 0.0331 * (x_{28} - 1227.13) - 0.0069 * (x_{29} - 1182.48) + 0.0025 * (x_{13} - 1121.75) + 0.0029 * (x_{15} - 1099.4337) - 0.0091 * (x_{14} - 918.34) + 0.0081 * (x_{16} - 1099.88) + 0.04807 * (x_{19} - 907.16) + 0.0602 * (x_{21} - 881.05)$$

Eq. (6.6c)

$$PC_4 = 0.0044 * (x_{26} - 4.78) + 0.0058 * (x_{27} - 44.96) + 0.0092 * (x_{28} - 1227.13) - 0.0008 * (x_{29} - 1182.48) - 0.0101 * (x_{13} - 1121.75) - 0.0184 * (x_{15} - 1099.44) - 0.0729 * (x_{14} - 918.34) - 7.63E - 006 * (x_{16} - 1099.88) + 0.0256 * (x_{19} - 907.16) + 0.0297 * (x_{21} - 881.05)$$

Eq. (6.6d)

6.6.2 Multinominal logistic model results

A simple general to specific procedure was next utilised to obtain the structure of a parsimonious Multinomial Logistic Model. First, all the variables in table 5 were included in the model, but with all the temperature variables shown in table 14 replaced by the four principal components shown in the previous sub section. Using all these process variables and principal components the log likelihood given by Eq. (6.5) was maximised using SOLVER in Excel. All process variables that were not significant at the 5 % significance level were then removed from the model. This significance of a process variable was judged by calculating twice the ratio of the log likelihoods calculated with and without this process variable. Under the null that this process variable is unimportant, this statistic is Chi square distributed with 3 degrees of freedom (as each variable enters the equation for predicting the log odds in the four different categories of scale). Using this reduced number of process variables, the log likelihood given by Eq. (6.5) was maximised again using Excel's SOLVER. This iterative procedure was repeated until all the process variables in the Multinomial Logit Model were statistically significant at the 5 % significance level.

Table 16. The estimated values for the parameters in Eq. (6.4).

	Scale category		
Variable	Low Scale; $Y_i = 1$	Medium Scale; $Y_i = 2$	3 High Scale; $Y_i = 3$
Constant	8.98 [1.2]*	43.02 [4.7]	30.89 [2.7]
x_{38}	-119.76 [33.6]	-191.43 [-5.0]	-303.39 [-6.7]
x_{39}	-0.0004 [-0.3]*	-0.005 [-2.5]	-0.009 [-4.7]
x_1	-54.75 [-3.0]	-42.79 [-2.0]	-151.18 [-5.9]
x_3	84.33 [2.8]	384.79 [10.1]	702.25 [15.4]
x_6	27.52 [1.8]	50.09 [2.7]	-8.80 [-0.4]*
x_7	18.13 [0.6]*	-72.52 [-2.0]	-135.25 [-3.1]
x_8	-48.69 [-2.4]	-120.66 [-4.6]	-162.81 [-4.9]
x_9	568.61 [3.4]	532.48 [2.7]	375.19 [1.7]
x_{12}	-632.02 [-3.5]	-680.09 [-3.2]	-572.31 [-2.3]
x_{31}	-0.229 [-1.3]*	-0.888 [-4.2]	-1.13 [-4.5]
x_{18}	0.016 [3.1]	0.011 [1.8]	0.037 [4.6]
x_{22}	-0.023 [-3.2]	-0.054 [-7.3]	-0.055 [-7.2]
PC_1	0.116 [2.2]	0.625 [9.0]	1.03 [11.9]
PC_2	0.276 [5.0]	0.401 [6.2]	0.303 [4.1]
PC_3	-0.001 [0.1]*	0.204 [2.5]	0.321 [3.1]

Ln(L) = -1598. Base Ln (L) = -2179. Student t values in parenthesis.* Statistically insignificant at the 10 % significance level.

Table **16** shows the estimates made for β_{jk} parameters in the parsimonious version of Eq. (6.4) as identified using the above mentioned iterative procedure. Recall that for model identification, normalisation with respect to the very low scale category was undertaken (i.e. the parameters of the model for this category are set to zero). In table 16, all variables are statistically significant in the sense that they determine the probability of scale formation in at least one category of scale severity. For example, the addition of Nickel (x_7) to the base steel does not affect the probability of low amounts of scale forming, but it does significantly determine the probability of medium and high amounts of scale forming. Exactly the same can be stated for the operation of the Interstand Spray System (x_{31}) in that the choice between automatic and manual operation does determine the probability of medium to large amounts of scale forming – but not the probability of low amounts forming. On the other hand, the addition of Copper (x_6) is statistically significant in the determination of the probability of low to medium amounts of scale forming on the steel surface, but not large amounts.

Table 17. Table of observed and predicted values for scale count.

			Predicted Category					
			$Y_i = 0$	$Y_i = 1$	$Y_i = 2$	$Y_i = 3$	Sum Observed	% of correct predictions
Observed Category		$Y_i = 0$	182	136	19	30	367	50
		$Y_i = 1$	122	242	71	27	462	52
		$Y_i = 2$	28	103	139	99	369	38
		$Y_i = 3$	10	29	68	272	379	72
		Sum Predicted	342	510	297	428	1577	

$Y_i = 0$ corresponds to a very low scale count, $Y_i = 1$ corresponds to a low scale count, $Y_i = 2$ corresponds medium scale count, $Y_i = 3$ corresponds to a high scale count.

The base log likelihood, is the log likelihood associated with a model containing only a constant term. Thus the ratio of this to the log likelihood when the process variables in table 16 are included provides a measure of explanatory power. Using these criteria, the inclusion of these process variables improves the ability to predict various categories of scale intensity by almost 30 %. Table 17 provides another way of visualising the explanatory power of this model as demonstrated by giving a cross plot of the models predictions against the actual severity of scale formation. Thus 379 of the 1577 coils had a high scale count and the model

correctly predicted 272 of these coils to have a high scale count. (If the model gives a probability of more than 0.5 for a high scale count, then that coil is predicted by the model to have high scale count formation). This is a correct prediction rate of 72 %. The model has mediocre power in predicting low scale counts, and poor predictive power in relation to moderate amounts of scale formation. It thus appears that this type of statistical model works best at identifying the process conditions leading to large amounts of scale forming compared to low amounts of scale forming. When using this model, more success in controlling scale will be achieved by ensuring the Hot Mill is not run under those conditions most likely (as predicted by the model) to lead to high scale counts, compared to ensuring the Hot Mill is run under those conditions most likely to lead to low scale counts.

There are various ways to use this model to identify these optimal Hot Mill conditions. One is to use an optimisation algorithm to search for those Hot Mill conditions the model identifies as most likely to result in high scale counts (and then avoid running at these conditions). There are likely to be many such conditions that could then be tabulated for the Hot Mill operators to refer to. An alternative is to take a partial view, i.e. look at one process variable at a time whilst running the Hot Mill at the average (or other predetermined set of values) values for all the other process variables. Using this approach there are two useful tools or statistics to refer to – probability plots and elasticities.

Whilst the parameter values in table 16 are useful for identifying the statistically significant variables that are present in scale formation, they demostrate very little about how important each variable is for controlling the process – because these parameters are related to the log odds ratio rather than the probabilities of scale forming. Elasticities are superior in this respect because they show the percentage change in the probability of each category of scale forming that results from a one per cent change in the process variables. In table 18 below, these elasticities are calculated at the average values for each process variable.

Table 18. Elasticities for each of the statistically significant process variables.

Process Variable	Elasticity (% change)			
	Scale Categories			
	$Y_i = 0$	$Y_i = 1$	$Y_i = 2$	$Y_i = 3$
Transfer Gauge (x_{38})	1.06	0.36*	-0.62	-0.81
Coil Width (x_{39})	0.45	0.80	-0.52	-0.75
% Carbon (x_1)	0.66	-0.06*	0.22*	-0.83
% Phosphorous (x_3)	-0.62	-0.82	0.66	0.79
% Copper (x_6)	-0.08	0.01*	0.13	-0.07
% Nickel (x_7)	0.07	0.21	-0.14	-0.14
% Chromium (x_8)	0.25	0.19	-0.25	-0.19
% Aluminium (x_9)	-2.66	1.75	1.13*	-0.22*
% Solute Aluminium (x_{12})	3.01	-1.26	-1.58*	-0.18*
Interstand Spray System (x_{31})	-0.06	-0.07	0.08	0.05
Average CS Temperature (x_{18})	-2.70	0.73*	-1.07*	3.04
Average ROT Temperature (x_{22})	4.38	2.75	-5.11	-2.01

$Y_i = 0$ corresponds to a very low scale count, $Y_i = 1$ corresponds to a low scale count, $Y_i = 2$ corresponds medium scale count, $Y_i = 3$ corresponds to a high scale count. * Statistically insignificant at the 10 % significance level.

Ignoring the temperature variables in the three principal components for a moment, it can be observed that irrespective of the scale category, the largest elasticities are associated with the Aluminium content of the slab and various Hot Mill temperatures. Thus it can be seen that the average Crop Shear (x_{18}) and the average ROT (x_{22}) temperatures have the highest impact on the probability of observing high scale counts on the finished coil. For example, a 1 % increase in the average Crop Shear temperature (x_{18}) will bring forth a 3.04 % increase in the probability of high scale counts occurring on the finished coil when the Hot Mill runs at the average values for all the other process variables. Then at the other end of the scale classification, it can be seen that the % Aluminium Solute (x_{12}) and the average ROT temperature (x_{22}) have the biggest impact on the probability of observing very low scale counts on the finished coil. For example, a 1 % increase in the average ROT temperature (x_{22}) will bring forth a 4.38 % increase in the probability of very low scale counts occurring on the finished coil when the Hot Mill runs at the average values for all the other process variables.

When looking across the scale categories, the elasticities associated with each variable change sign as you move from the lower scale count categories to the higher ones For example, a 1 % increase in the % Phosphorous added to the base steel will bring forth a 0.62 % decrease in the probability of very low scale counts being observed on the finished coil, but a 0.79 % increase in the probability of high scale counts being observed when the Hot Mill runs at the average values for all the other process variables. Elasticities can also be calculated using values other than the average for the other process variables.

In table 18 the elasticities for the principal components are not shown because by construction they will equal zero. This is because the mean values for each component are by construction zero. This does not imply that the process variables in each of the components are unimportant, but that under the special conditions where these variables are such that the components are zero, their influence is minimal. Under all other conditions the temperature variables in each component are significant variables as observed from the probability plots in figure 13.

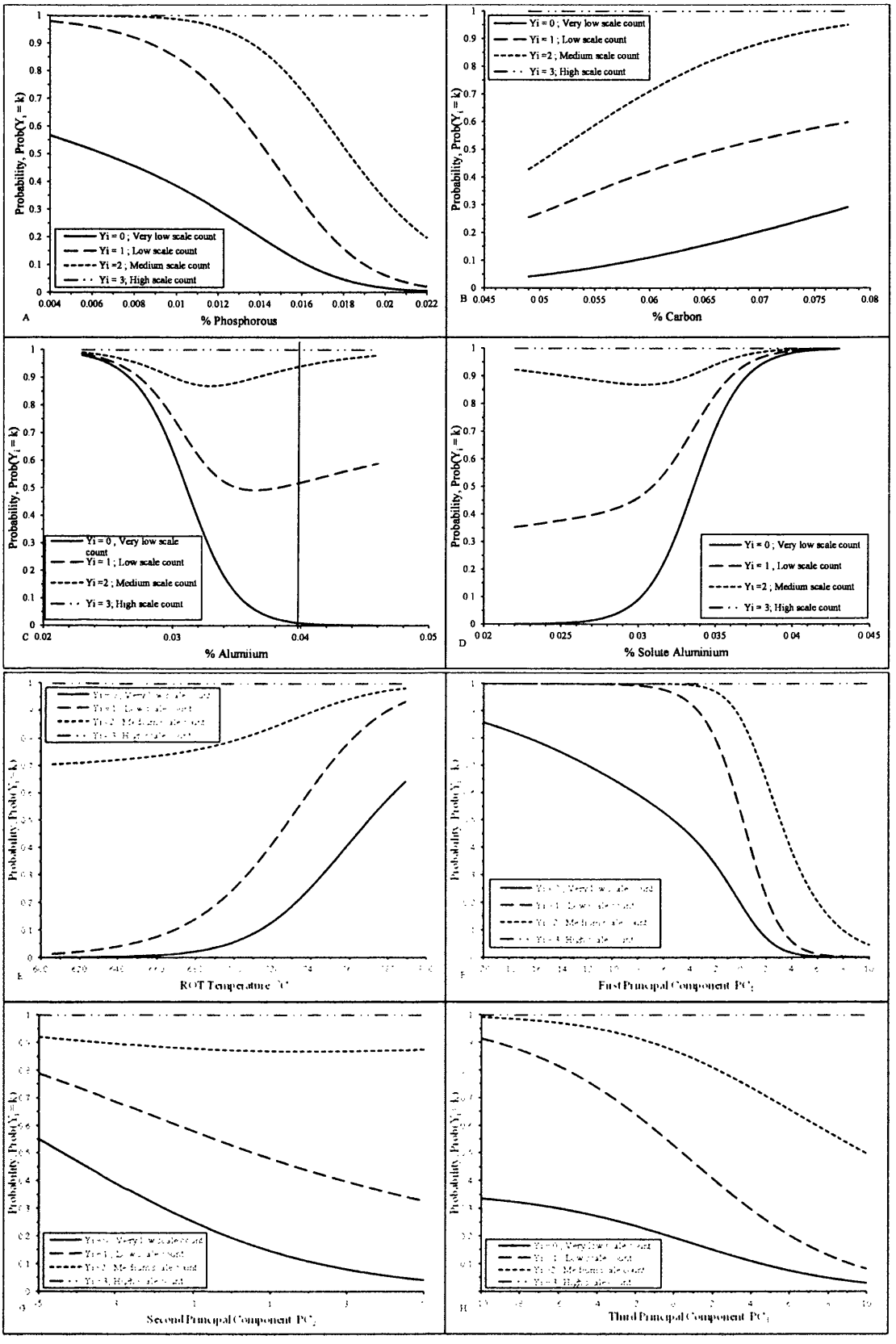


Figure 13. Probability of scale formation with variations in some of the process variables.

Figure 13 shows the effect of changing each process variable on the probability of observing the various categories of scale whilst holding all the other variables at their mean values. The three principal components of the temperature variables are shown, together with some of the other process variables that have high elasticities.

To illustrate how to use such graphs for control purposes take the amount of aluminium added to the slab. At low levels of Aluminium (0.024 % or less) it is virtually certain that the processed coil will have a very low scale count when operating the Hot Mill at the average values for all other variables shown in table 5. Contrastingly, when the incoming slab to the Hot Mill has 0.04 % Aluminium and the Hot Mill operates at average values for the other variables there is limited chance of the processed slab having a very low scale count. Under these conditions around 55 % of processed coils will have low scale counts and a further 50 % medium scale counts. The other chemistries shown in figure 13 can be interpreted in similar ways. Interestingly the effect of % Solute Aluminium (x_{12}) is similar to % Aluminium (x_9) but in the opposing direction.

Next consider the first principal component in figure 13. When this first principal component is below -3 in value (with all other process variables set at their average values), around 90 % of processed coils will have very low or low scale counts. Consequently, a small increase in this principle components value above -3 will result in a dramatic change, for example over 75 % of coils being medium or high in scale count when this principal component value is 2. Such probability plots can also be calculated using values other than the average for the other process variables.

These principal component values can also be converted into actual temperature values at various points in the Hot Mill. As an example take the 90 % or more of coils being low or very low in scale as being an acceptable outcome for the Hot Mill output. This corresponds to $PC_1 = -3$ or less as shown above. Using Eq. (6.6) and $PC_1 = -3$, a particular Hot Mill temperature can be calculated assuming the process runs at the mean values for all the other temperature variables at the Hot Mill. These calculated temperatures are shown in table 19. It was not possible to achieve a PC_1 value of -3 or less by manipulating process variables x_{26}, x_{27} and x_{14}, when running the mill at average values for all the other variables shown in table 5. The centre of the Reheat Furnace should be set at 875 °C. When observing at the average

Rougher Mill temperature the temperature should be 1020 °C or less. The corresponding temperature for the maximum Crop Shear temperatures is 1010 °C or less. The corresponding temperature for the maximum Finishing Mill temperatures is 865 °C or less. Finally, for the average Finishing Mill temperature the temperature is 850 °C or less. This type of calculation can be done for any PC_1 value, and therefore for any frequency of significant scale formation to determine what temperatures to run the Hot Mill at to achieve this required frequency. It should be noted that the values in table 19 are unlikely to be used because the material would struggle to be effectively processed at these temperatures or achieve the required mechanical properties.

Table 19. Maximum temperature settings to achieve a 10 % or less chance of very low scale forming.

	x_{28}	x_{29}	x_{13}	x_{15}	x_{16}	x_{19}	x_{21}
Temperature, °C	1095	875	1040	1020	1010	865	850

6.7 MLR Conclusions

The principal component analysis above indicated that the recorded temperatures at the Hot Mill were highly correlated. However, nearly 85 % of the variation in the ten Hot Mill temperatures could be accounted for by just four principal components that were completely uncorrelated with each other.

The above analysis showed that to be able to successfully predict scale formation many variables need to be taken into account including chemistry (most notably the addition of Phosphorous and Aluminium), temperature (that at the Run Out Table and those temperature constituting each principal component), and the descaler sprays. The Multinomial Logistical Regression model also had an overall correct prediction rate of 50 %. However, a correct prediction rate of 72 % was obtained for the high scale category. This large difference in prediction shows that there is a large inherent variability within different sections of the dataset, which implies that some scale count ranges have a significant influence from variables not being utilised within this dataset.

The main aim for estimating this multinomial model was not just to find the most important determinants of scale formation as the above identified variables are broadly consistent with what is already known. The main aim was to provide quantitative information on how to control the Hot Mill, i.e. where to set the process variables so as to minimise the number of coils produced with significant scale present on their surfaces. The model suggested that this can be done by choosing either the right chemistry for the incoming steel or through setting the correct temperatures at the Hot Mill. It was found, for example, that to keep the rate of coils containing medium or high scale counts below 10 %, the average Rougher Mill temperature should be kept at 1020 °C or more, when running the Hot Mill at the average values (shown in table 5) for all the other process variables. For chemistry it was found for example, that to keep the rate of coils containing medium or high scale counts below 10 %, the Aluminium content must be below 0.024 % wt. when running the Hot Mill at the mean values for all the other process variables.

It should be noted that the model could also be used to identify the optimum temperatures to minimise scale formation if the chemistry was required to be changed.

6.8 References

Cattell, R. B., 1952. *Factor Analysis*. New York, Harper.

Chapter 7. Development of a multi-layer ANFIS model for the prediction of tertiary scale formation

The Adaptive Neuro fuzzy inference system (ANFIS) model integrates neural network and fuzzy logic principles and uses them within a single framework. The model developed within this body of work is constructed in 3 layers. The aim is then to use the knowledge gained from this technique to better understand scale formation and how it can be reduced at the point of manufacture. Only those variables identified as being significant (see table 20) in the previous 3 chapters are considered for inclusion in this ANFIS model.

Table 20. Process Variables and their Sample Means and Standard Deviations.

Process Variable	x_j	Mean	Standard Deviation
% Phosphorus	x_3	0.014	0.0031
Maximum RM Exit Temperature, °C	x_{13}	1121.8	17.87
Average RM Exit Temperature, °C	x_{15}	1099.4	16.18
Maximum CS Temperature, °C	x_{16}	1099.9	19.89
Average CS Temperature, °C	x_{18}	1069.6	20.8
Maximum FM Exit Temperature, °C	x_{19}	907.16	7.18
Average FM Exit Temperature, °C	x_{21}	881.05	5.75
Scale Count Bottom Surface	Y	250.1	402.6

RM = Rougher Mill, CS = Crop Shear, FM = Finishing Mill and ROT = Run Out Table.

7.1 ANFIS model

The basic philosophy of the ANFIS model is to split the data up into subsets and then to fit a standard second order response surface model to the data in each subset. However, where the splits in the data occur are fuzzyfied. Figure 16 describes the structure for a simplified ANFIS model where the experimental data is split up into four sub sections.

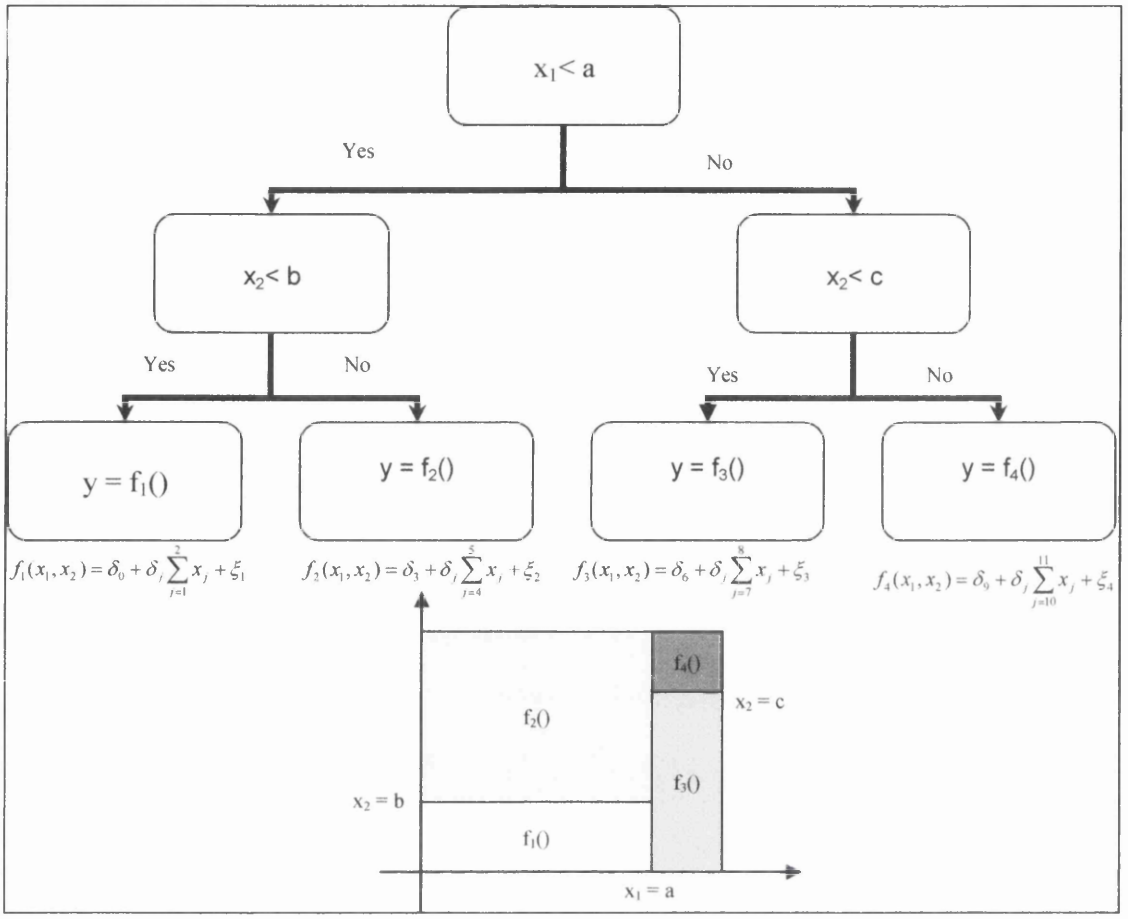


Figure 14. A typical binary regression tree with two inputs (x_1 and x_2) and one output y.

The decision tree partitions the input space into a number of non overlapping rectangular regions, each of which is assigned a label $f_i()$ to represent a predicted output value. In this illustration the tree identifies four simple if – then rules. Note that each terminal node has a unique path that starts with the root node and ends with the terminal node; the path corresponds to a decision rule that is a conjunction of various tests or conditions. Typically, each $f_i()$ will be a linear function of x_1 and x_2 and for each region there is a separate linear equation. The functional form for $f_i()$ does not have to be linear however. More generally, and as used in this paper, it can be a second order response surface model given by

$$f(x_1, x_2) = \beta_0 + \beta_j \sum_{j=1}^{2} x_j + \lambda_{jv} \sum_{j=1}^{2} \sum_{v=j+1}^{2} x_j x_v + \varphi_j \sum_{j=1}^{2} x_j^2$$

Eq. (7.1)

The problem with this approach is that the resulting modelled response surface is highly discontinuous in that it changes abruptly at the decision rules. This problem is overcome by fuzzyfying the decision rules. For example, the crisp decision rule associated with the left most branch of the binary tree in figure 14 is

If $x_1 < a$ and $x_2 < b$ then $y = f_1()$

It is established that fuzzy expert systems use membership functions to quantify possibilities (Jang *et al.* (1997)). Possibility is a fuzzy measure indicating the degree of evidence or belief that a certain value x_1 belongs to a set. A membership function has a value between 0 and 1 such that x_1 values further and further below a, have membership values closer and closer to one. A common functional form used for the membership function is the sigmoidal function

$$\mu_{x1<a} = 1 - \frac{1}{1 + exp[-\kappa(x_1 - a)]}$$

Eq. (7.2a)

where κ is a parameter requiring estimation. The value for κ determines the steepness of the membership function at a. So the further x_1 is below a, the greater will be the value for $\mu_{x1<a}$, indicating a stronger belief that that value for x_1 belongs to the set $x_1 < a$. $\mu_{x1} < a$ varies over the range 0 to 1, with 1 indicating the strongest possible belief.

Finally, the parameters of the model are fine tuned using a particular type of neural network. This network is called an ANFIS – adaptive network – based fuzzy interference system (Tsoukalas and Uhrig (1997)). There are various ANFIS architectures, but one using a first order Sugeno (1995) fuzzy model is the most common. The layers of this ANFIS network are shown in figure 15 for the decision rules shown in figure 14.

Rule 1: If $x_1 < a$ and $x_2 < b$, then $y = f_1()$

Rule 2: If $x_1 < a$ and $x_2 \geq b$, then $y = f_2()$

Rule 3: If $x_1 \geq a$ and $x_2 < c$, then $y = f_3()$

Rule 4: If $x_1 \geq a$ and $x_2 \geq c$, then $y = f_4()$

In the first layer of the ANFIS network, each of the i values for x_1 and x_2 are given membership quantities using the following sigmoidal functions

$$\mu_{x1<a} = 1 - \frac{1}{1+\exp[-\kappa_1(x_1 - a)]} \text{ with inverse (INV) } \mu_{x1\geq a} = \frac{1}{1+\exp[-\kappa_1(x_1 - a)]}$$

Eq. (7.2b)

$$\mu_{x2<b} = 1 - \frac{1}{1+\exp[-\kappa_2(x_1 - b)]} \text{ with inverse (INV) } \mu_{x2\geq b} = \frac{1}{1+\exp[-\kappa_2(x_1 - b)]}$$

Eq. (7.2c)

$$\mu_{x2<c} = 1 - \frac{1}{1+\exp[-\kappa_3(x_1 - c)]} \text{ with inverse (INV) } \mu_{x2\geq c} = \frac{1}{1+\exp[-\kappa_3(x_1 - c)]}$$

Eq. (7.2d)

where κ_1 to κ_3 are parameters requiring estimation. In layer 2 weights are determined that represent the possibility that each pairing for the i values of x_1 and x_2 belong to one of the four sets given by the decision rules above. These weights are given by

$$w_1 = (\mu x_1 < a)(\mu x_2 < b)$$

Eq. (7.3a)

$$w_2 = (\mu x_1 < a)(\mu x_2 \geq b)$$

Eq. (7.3b)

$$w_3 = (\mu x_1 \geq a)(\mu x_2 < c)$$

Eq. (7.3c)

$$w_4 = (\mu x_1 \geq a)(\mu x_2 \geq c)$$

Eq. (7.3d)

These are examples of a T-norm operator for working out the possibility, for example, that x_1 is less than a AND x_2 is less than b. In figure 15, Π stands for the use of this T-Norm. In layer

3 each value for $f_i()$ is multiplied by its w_i value so that more emphasis is placed on the linear function corresponding to the rule most likely to describe the x_1 and x_2 pairing (the form for $f_i(0$ is as shown in figure 14). Finally, in layer 4 these weighted functions are added up to give the prediction for y coming out of the ANFIS network. Conjugate gradient methods are then used to optimise the values for κ_1 to κ_3 and all the parameters of $f_1()$ to $f_4()$.

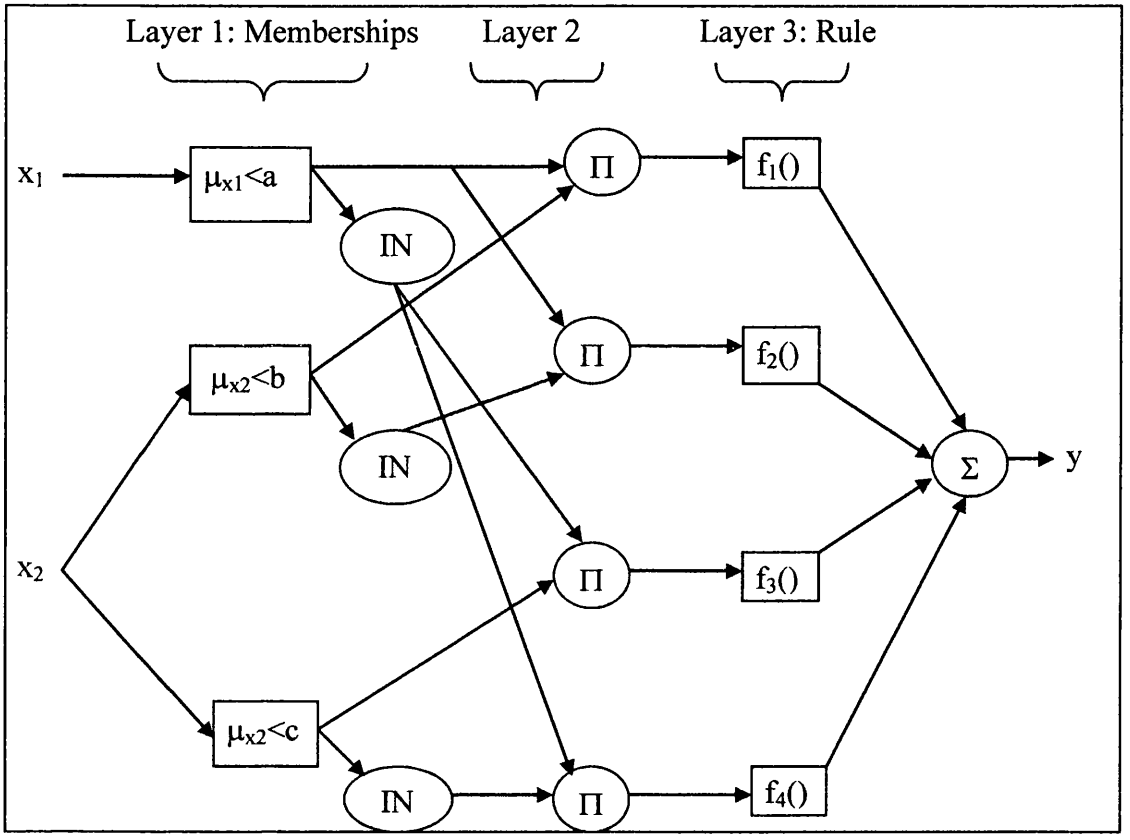


Figure 15. ANFIS architecture corresponding to the representation shown in Figure 14. Values for w given by Eq. (7.3).

The only remaining question is how to decide which variables to split by and how many variables to split. In this paper a simple binary procedure is used. The first of the p process variables shown in table 20 is selected and the above ANFIS model fitted for just this one variable. The predictions from this ANFIS model with p = 1 are then compared to the observed values and the sum squared error is computed

$$Sum\ Square\ Error = \sum_{i=1}^{N}(Predicted_i - Y_i)\ \hat{}\ 2 \qquad \text{Eq. (7.4)}$$

for all different values of κ_1. The value for κ_1 is then taken to be that value producing the smallest sum squared error. This is repeated for all p variables and the variable chosen for the first split is the one producing the smallest Sum Squared Error. The variable chosen for the second split is found in the same way and any number of splits can be made provided each sub section of the data has enough observations in it to estimate Eq. (7.1).

7.2 ANFIS Results and Discussion

Table 21 shows the sum square error of each of the stages described above. Var is a vector containing the variables used to split the data. Thus Var_1 contains the variable used to create the first split and Var_2, contains the variable used to create the second split in the data and so on. Table 21 shows that the best predictor for Var_1 was the AvgRM. For, Var_2, PerP is the best predictor of the variables investigated. While for Var_3 AvgCS was added to the ANFIS model.

Table 21. Table showing the sum of the square errors for the different variables used.

	Sum Square Error for Var_1 (AvgRM)	Sum Square Error for $Var_{1,2}$ (AvgRM, PerP)	Sum Square Error for $Var_{1,2,3}$ (AvgRM, PerP, AvgCS)
PerP	1363.95	**935.99**	N/A
MaxRM	1264.09	*1174.43*	825.16
AvgRM	**1207.79**	N/A	N/A
MaxCS	1373.91	1150.69	811.82
AvgCS	1375.78	1072.16	**758.16**
MaxFM	*1466.22*	1103.87	809.35
AvgFM	1409.06	1139.55	*826.95*

The numbers in bold are the variables that have the lowest sum of squared error for that Var. The numbers in Italics have the highest sum of square error for the Var.

Consider first the ANFIS model with just the first variable identified in table 21. Figure 16 is a graph of observed versus predicted scale count for a Var_1 (AvgRM) ANFIS model. The

values for a and κ in Eq. (7.2a) were 1.69583 and 1132.70 respectively and $f(Var_1)$ was estimated to be

$f(Var_1) = (3.97)+(-3.08)*AvgRM+(1.31)*AvgRM^2$ when $Var_1 > a$

$f(Var_1) = (-0.23)+(0.17)*AvgRM+(0.01)*AvgRM^2$ when $Var_1 < a$

When a linear trend line is added to the data in figure 16, the equation for the trend line has a slope value of near 1 and a very low intercept term. The R^2 value is 0.2336 which means that 23 % of the observed values for scale count can be explained by this ANFIS model. The line is close to a 45 degree line so there is very little systematic prediction error.

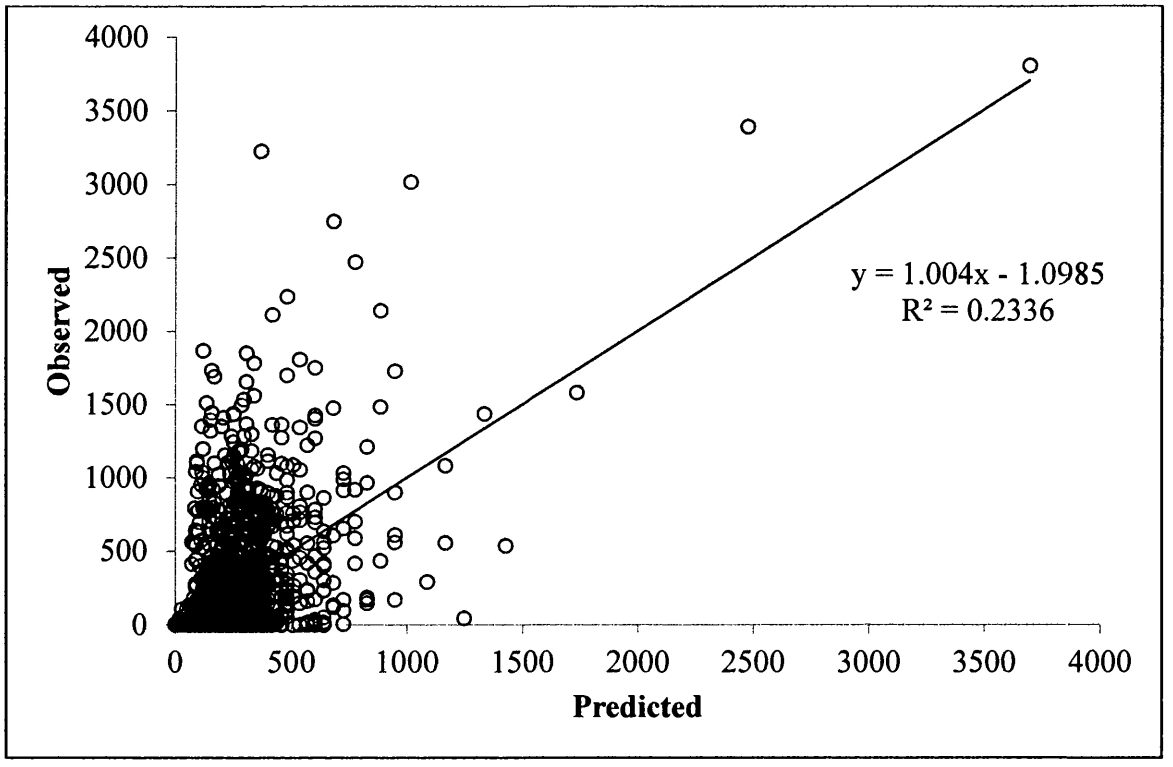


Figure 16. Observed tertiary scale versus predicted tertiary scale for Var_1 (AvgRM).

Figure 17 shows the observed versus predicted results for the $Var_{1,2}$ ANFIS model. It can be seen that in comparison to Var_1 that the slope of the trend line has improved (converged to one) by 0.0024 and the intercept has improved (converged to zero) by 0.2899. This combined with an improved R^2 value of 0.4104, indicates that the prediction obtained by this model has less systematic errors in prediction and a higher predictive capability. This ANFIS models prediction now describes 41 % of the observed dataset's results.

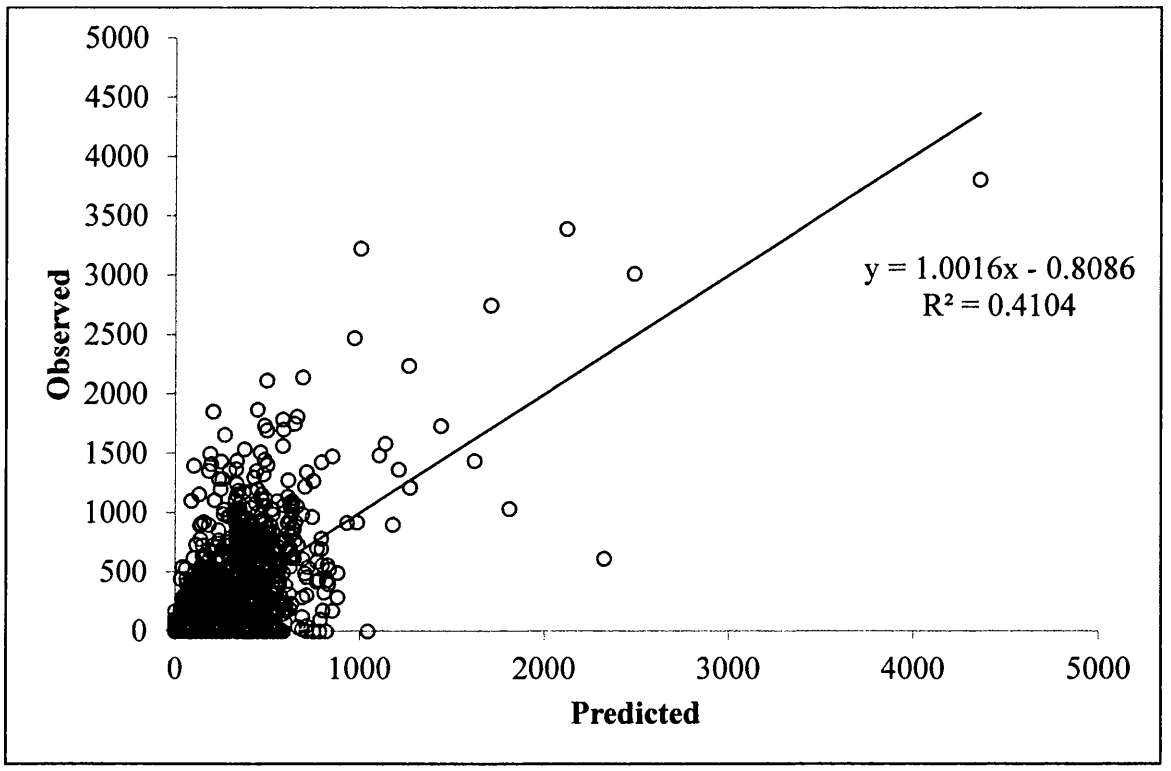


Figure 17. Observed tertiary scale versus predicted tertiary scale for $Var_{1,2}$ (AvgRM, PerP).

Figure 18 shows the observed versus predicted results for the $Var_{1,2,3}$ ANFIS model (i.e. with all 3 variables identified in table 21 included). When compared to the $Var_{1,2}$ model its trend line equation has its slope value decreased by 0.0021 and the intercept has decreased by 0.9902. However, this is not a significant change. By using the 3 best calculated predictors for the ANFIS model, the predictions explain 52 % of the observed dataset.

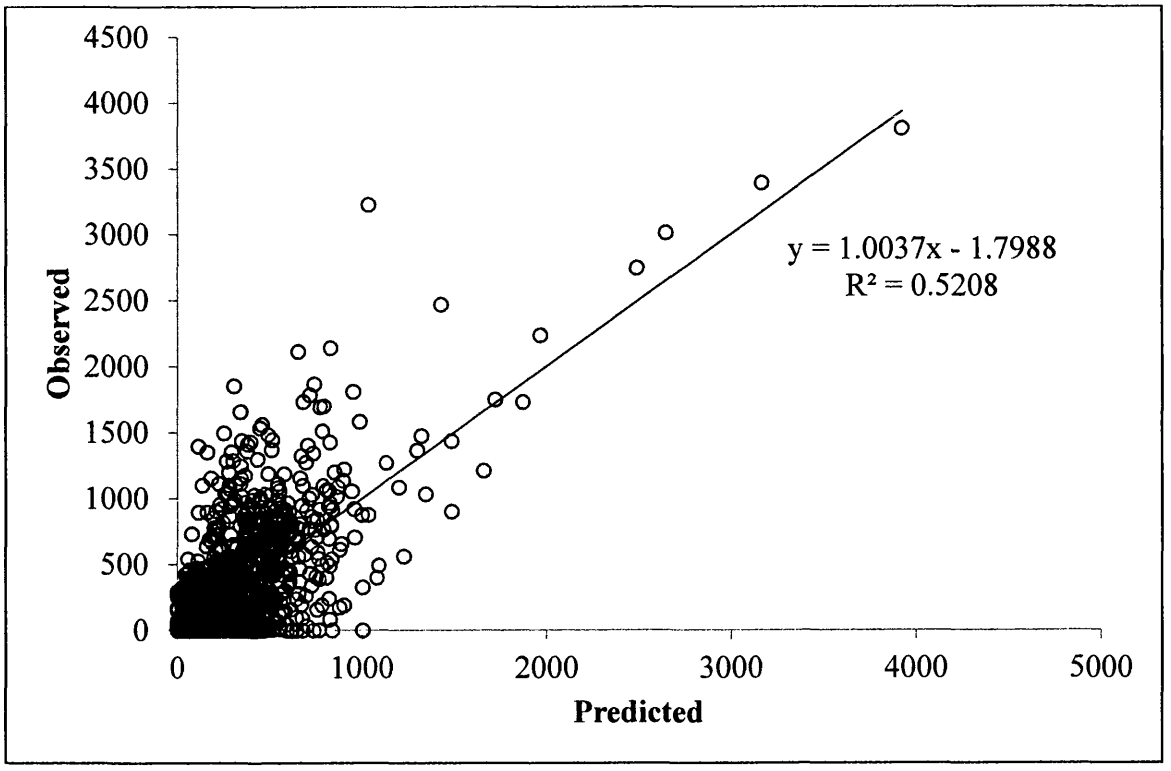


Figure 18. Observed tertiary scale versus predicted tertiary scale for $Var_{1,2,3}$ (AvgRM, PerP, AvgCS).

Table 22 illustrates the regression parameters for this last ANFIS model, with the most significant coefficients for each of the interactions highlighted in bold. Observations from table 22 indicate the value with the highest regression parameter is X_2 (PerP) with 3 out of 8 values highlighted. This is followed by X_1 (AvgRM) and the constant with 2 out 8 each, followed by X_3 (AvgCS) with 1 out of 8. This indicates that the variable that has the greatest influence on scale count within a sub region of the data is X_2 (PerP), and the general trend is for variables to have greater significance than the interactions between the variables or the squared values of the variables.

Table 22. Table showing the regression parameters for $Var_{1,2,3}$. Bold numbers represent the most significant parameters for that function.

	Regression parameters							
	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8
const.	-216.27	67.83	-89.15	67.83	13.43	**-4.47**	14.74	**-0.36**
X_1	168.70	**-45.54**	55.90	**-45.54**	10.24	-2.15	8.12	-0.19
X_2	**334.64**	-18.57	**97.07**	-18.57	**-28.31**	1.58	-1.38	0.18
X_3	-229.92	-14.60	85.77	-14.60	9.68	0.45	**-21.42**	0.30
X^2_1	-23.04	8.61	-13.55	8.60	0.38	-0.73	0.51	0.02
X^2_2	-64.08	4.72	19.47	4.72	8.11	-0.30	0.10	0.02
X^2_3	69.61	-2.43	-22.29	-2.43	-5.56	0.29	7.89	0.01
X_{12}	-44.87	4.26	-17.87	4.26	-3.56	-0.63	-4.21	-0.15
X_{13}	-6.65	5.70	-2.39	5.70	2.61	0.51	-5.95	-0.03
X_{23}	-25.81	-2.31	2.12	-2.31	3.73	1.15	0.25	0.24

Table 23 shows the calculated κ and a values for $Var_{1,2,3}$. The κ represents the steepness of the membership function, a low value will give a shallow slope while a higher value will give a steeper slope. A steeper slope will tend to yield an ANFIS model very similar to a CART model. The data shows that the κ value is lowest for AvgRM and highest for AvgCS. This demonstrates that the steepness of the membership function improves with every new variable added to the model.

The a value (from table 23) represents where the fuzzy split in the membership function will occur relative to the variable. The split for AvgRM is over two standard deviations away from the mean value, so is at the higher end of the range. The a value for PerP is within half a standard deviation of the mean, so is comparable to the average value of the dataset. For AvgCS the a value is greater than one standard deviation from the mean. These are the three most important determinants of scale count at the Hot Mill.

Table 23. Table showing the calculated κ and a values at $Var_{1,2,3}$.

	AvgRM	PerP	AvgCS
κ	1.69583	3.2322981	4.3772184
a	1132.70	0.0155	1094.03

Table 24 illustrates coils at different scale counts with the predicted scale counts and the three variables (AvgRM, PerP, and AvgCS). The variables show an overall increase with increasing predicted values as one would expect. The difference in the observed and predicted count in the sampled coils varies from being off by 20 to 200 counts. In general the scale count prediction decreases in accuracy the higher the scale count becomes (excluding the highest values). This is likely to be due to the influence of conditioning sprays in the Finishing Mill.

Table 24. Table showing selected coils at different scale counts

Observed Bottom Scale Count	Predicted Bottom Scale Count	AvgRM	PerP	AvgCS
0	87.3	1094	0.014	1066
199	175.8	1109	0.012	1077
1003	712.4	1084	0.02	1072
3806	3924.2	1156	0.017	1093

7.3 ANFIS Conclusions

From the variables that were selected it can be concluded that scale formation is mostly governed by the temperature of the steel as it enters the early stands of the Finishing Mill (AvgRM & AvgCS). With 23 % of the dataset being predicted by AvgRM and 11 % (difference in R^2 between $Var_{1,2}$ and $Var_{1,2,3}$) of the dataset being explained by AvgCS. This gives a combined value of 34 % of the dataset explained by pre-Finishing Mill temperatures.

There is also an element of chemistry that impacts upon the scale count. With PerP being the best predictor for $Var_{1,2}$ accounting for 18 % of the observed scale counts.

The ANFIS model gives good predicted scale counts at low and high scale. This is most likely a product of the process, because if the temperature going through the Finishing Mill is low, then a low amount of scale will be generated. While if the coil is very hot coming into the Finishing Mill then a high volume of scale will be generated. However, for mid scale counts (approximately > 750) the effectiveness of the conditioning sprays might have a

significant influence. These sprays are situated on the entry into the Finishing Mill and will suppress scale formation. If these conditioning sprays are running less effectively than normal for a period of time, then this could explain why the predicted values usually underestimate the scale count. Also, if the sprays do not come on/off at the correct time then the head/tail of the coil will not be conditioned which will lead to increased scale on the missed section. The other reason for this weakness in the model could be having average temperature values across the coil.

Overall the 3 level ANFIS models prediction rate was 52 % of a high variability dataset (i.e. variation in scale count).

7.4 References

L. H. TSOUKALAS and R. E. UHRIG. "Fuzzy and Neural Approaches in Engineering" (John Wiley & Sons Inc, New York, Chapter 2, 1997).

J. S. R. JANG, C. T. SUN and E. MIZUTANI. "Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence" (Prentice Hall, Upper Saddle River, Chapter 12, 1997).

T. TAKAGI and M. SUGENO. "Fuzzy Identification of Systems and its Applications to Modelling and Control", *IEEE Transactions on Systems, Man and Cybernetics*. 15 (1995) pp.116-132.

Chapter 8. General Discussion

This general discussion section will bring together the main conclusion obtained from the application of the statistical techniques employed within this research.

8.1 Binary vs. Multinomial vs. Continuous

The dependent variable has been described in a number of ways for the different data mining techniques. The three ways that have been described are as follows;

- Binary
    - Changes the continuous scale count to a binary classifier
- Multinomial
    - Changes the continuous scale count to a multinomial classifier
- Continuous
    - The dependent variable is the Parsytec scale count.

Each way of classifying the data has its own strengths and weaknesses. Each of these strengths and weaknesses need to be considered before the models can be compared to each other.

8.1.1 Binary

Models that use a binary split tend to have higher predictive capability. This is because it is easier to predict a binary split compared to a continuous dependent variable. The relevance of these results are dependent on the choice of the split.

It is also important to consider the placement of the split as it needs to be relevant to the process that is being investigated. The problem with using the scale count is that there is no strict scale count that indicates scale problems downstream, but values above the selected value for the split are more likely to have downstream issues.

The main strength of this method is in classifying coils that have very low or very high scale counts. If the split is in the correct location then the binary split should give good results for low and high values of the scale count.

The main weakness of this method is the coils with scale counts around the split are going to be harder to correctly predict. With the split being at 200, scale counts of 199 will be classed as 0 while values of 201 will be classed as 1. In reality there is usually less difference between 199 and 201, than 199 and 0 or 201 and 401+. This can cause errors and incorrect predictions, which could lead to the wrong variables being selected.

Splitting the dataset into low quantity and high quantity scale has several advantages and disadvantages. The main purpose of this investigation is to identify the process parameters related to scale formation and it could be argued that splitting the dataset in a binary way will achieve this better than other methods. On the other hand the split gives low prediction for values around the binary split.

8.1.2 Multinomial

Splitting the continuous dependent variable into sets has several advantages and disadvantages over other techniques. The main advantage is that there are multiple sets that the coil can be classed into so you can better understand the severity of the scale present.

Like the binary dependent variable, problems with prediction occur around the splits. However there are more splits present and they cover less range than in the binary split so they should have less overall error associated with them.

The problem comes with the scale count not representing an increase in adherence but an increase in quantity of scale. This means that the different groups are not necessary representative of increasing scale problem. However scale count is an indicator of potential problems, so the different groups do represent increasing potential for scale problems to occur.

8.1.3 Continuous

Keeping the dependent variable as a continuous variable has several strengths and weaknesses when compared to other ways of representing the dependent variable.

The main strength is that it directly represents the observed dependent variable, unlike other methods which simplify the data. This avoids the problem associated with splitting the dataset into various groups.

The main weakness from using a continuous dependent variable is that the scale count being predicted is not a measure of adherence of the scale but frequency. As there is limited data at high scale counts this could lead to the model incorrectly predicted significant variables based on anomalous results. Also the variability is greater in the raw scale count data.

8.2 Model Comparisons

All the different models used in this investigation show slightly different significant variables. This is due to differences in how the models work and how the dependent variable is expressed. However there are some variables that keep occurring through the investigation.

In the different analysis undertaken temperature proved to be significant. As there is a high correlation between the different temperature variables, as seen by the two PCA performed in Chapter 4 and Chapter 6. It can also be observed from the PC results that there is a greater correlation between the pre-Finishing Mill temperatures than the post-Finishing Mill temperatures.

PC1 from Chapter 6 (see figure 12f) is comparable to the PC1 from Chapter 4 (see figure 6f). When Chapter 4's PC1 has all its components at their mean values (zero on the horizontal axis) there is a 15 % chance of the bottom scale count being greater than 200. While Chapter 6's PC1 (at its zero value) shows a 13 % chance of the bottom scale count being greater than 350, a 35 % chance of it being between 76 - 350, a 32 % chance of it being between 11 - 75, and a 20 % chance of the scale count being between 0 - 10. These values are very similar to each other and both show that as the temperature is increased above the mean values there is a significant increase in the chance of bottom scale being produced.

The temperature entering the Finishing Mill is the most significant variable for scale formation. This makes logical sense as tertiary scale is formed in the early stands of the Finishing Mill (see 1.2.4).

The main chemistry component that seems to contribute to scale formation is Phosphorus. Other elements have limited significance in some of the models but are not in all of them because the tinplate grade investigated has a tight chemistry specification. Phosphorus was found to be significant throughout the different analysis used, with values greater than 0.015 causing greater amounts of scale formation. Chapter 4 shows that at 0.015 % there is a 20 % chance that the coil will have a scale count greater than 200 on the bottom surface, which will increase to a 45 % chance at 0.020 % Phosphorus (see figure 6a). A similar pattern is observed in Chapter 6 in figure 12a; at 0.015 % Phosphorus there is a 19 % chance of the bottom scale count being greater than 350, a 31 % chance of it being between 76 - 350, a 35 % chance of it being between 11 - 75, and a 15 % chance of the scale count being between 0 - 10. While at 0.020 % Phosphorus there is a 66 % chance of the bottom scale count being greater than 350, a 28 % chance of it being between 76 - 350, a 4 % chance of it being between 11 - 75, and a 2 % chance of the scale count being between 0 - 10. Chapter 7 shows that 18 % of the dataset is predicted by Phosphorus, and is the second strongest variable used in the ANFIS analysis. These models both show that at Phosphorus levels greater than 0.015 % there is an increased chance of scale being a problem.

The Logit and PLS models have several similarities for comparison. The first is that they both use a binary dependent variable, which makes it easier for comparisons to be drawn. The PLS model has a greater prediction rate than the Logit model (85 % compared to 73 %), which means that the PLS model has a 12 % improved prediction rate on the dataset than the Logit model. Which could be interpreted that the PLS gives more reliable results than the Logit model. Both models show that temperature is a significant variable in scale formation. The PLS model predicts a 10 % scale rate if the average Finishing Mill is below 842 °C. The Finishing Mill temperature is dependent on the pre-Finishing Mill temperatures and the conditioning sprays present. This implies that if the conditioning sprays are working and the pre-Finishing Mill temperature is limited then better scale results will be achieved.

Even though the MLR and ANFIS models are very different there are several similarities that need to be discussed. Both models show that some areas of the dataset are more difficult to predict than others. The MLR model had an overall prediction rate of 50 % but the prediction rate for individual sets varies from 38 % to 72 %. The ANFIS model gave improved prediction at the extremes of the dataset, but had decreased performance at predicting mid-

range values. This shows that there are scale counts that are difficult to predict with the process data that is available. Both of these models show that the percentage of Phosphorus and the temperature of the strip is significant for scale formation.

8.3 Model Use

To be able to perform successful analysis a number of stages need to be undertaken. The first stage is to identify the reason the analysis is being carried out, this will influence how the problem will be structured and aid in interpreting the results. The second stage is to understand what data is available and if the data will allow the question to be answered. This should be carried out until the data collected is capable of answering the required question.

Stage 3 is the preparation of data. This is done by removing anomalous values and presenting the data in a way that is suitable for whichever models have been selected. Stage 4 is commencing the initial modelling, which should allow for a greater understanding of the data. This increased understanding will allow for improved modelling to be performed. This process will be repeated until usable results are achieved.

Stage 5 is to evaluate the results to understand if it is has met the requirements of stage 1. If the results have not answered the original problem then, the analysis should return to stage 1. If the analysis has answered the original question then the results should be deployed to obtain value from them.

8.4 Scale Formation

The temperature is the most significant variable for scale formation, particuly the temperature entering the Finishing Mill. Due to data limitations it cannot be determined the exact location of the scale formation but the data analysis and literature review confirm that the entry into the Finishing Mill is the critical location. This is highlighted in Chapter 6 and it can be observed that there is marked increase in scale formation when the average Roughing Mill temperature is greater than 1070 °C (see figure 10). With a similar increase in scale occurrence occuring when the average Crop Shear temperature increases above 1050 °C (see figure 10).

At lower temperatures little scale is generally formed. This confirms that scale formation is a temperature dependent reaction. While at high temperatures scale is generated in large amounts. Mid-way temperatures are influenced by other variables more than temperature, such as chemistry and descaling practice.

The chemistry present in the slab also has an influence on scale formation. The chemical element that is most significant for scale formation is Phosphorus. The literature states that Phosphorus becomes enriched at the metal/scale interface which suppresses the diffusion of Fe ions. This lack of Fe ions will not prevent the oxidation from arising, but will promote the formation of Magnetite and Hematite over Wustite. This will cause scale to be slightly more adherent which can lead to higher counts present on the coil.

Chapter 9. General Conclusions

It can be concluded that temperature is the most significant variable for scale formation. With the temperature entering the Finishing Mill being the most significant region to control to be able to effectively manage scale formation. Low temperatures will result in low levels of scale formation while high temperatures will result in high levels of scale formation. Mid level temperatures are more difficult to predict because it seems that other factors are influencing scale formation more significantly than temperature. In particular Phosphorus seems to have a significant impact on scale formation especially at values greater than 0.015 %.

Models should not be used in isolation, as every analysis method has advantages and disadvantages and they are often difficult to identify without several other methods to compare against.

After discussion with process specialists it was decided to implement monitors to pass the significant variables in a more visible way downstream of the Hot Mill. The plan is to continue to use the monitors to communicate between units in a productive way and to continue to evolve the monitors over time to manage changes in the system. The initial monitors are featured on table 25. The values and boundaries have been decided by comparing the results from this thesis with ongoing investigations into scale formation.

Each of the significant variables selected will be classed as RED (major risk of scale), AMBER (risk of scale), GREEN (low risk of scale). The class is selected depending on the value of the variable and the boundary conditions for the process condition. This simple method of visualising the data was selected to improve communication and improve the flow of information. There will also be an overall colour designation for the coil. The colour selected will be the highest colour (ranked; RED, AMBER then GREEN) that the coil had.

Table 25. Lead Monitor for Tinplate grades.

HOT MILL PARAMETER	METRIC	AIMS	RED	AMBER	GREEN
Exit RR temp achievement	Temp values along length of strip.	Aim temp 1060 °C ± 20 °C	If > 10 % length > 1090 °C, then RED	If > 20 % length > 1080 °C, then AMBER	If > 95 % length < 1080 °C, then GREEN
Crop Shear temp achievement	Temp values along length of strip.	Aim temp 1050 °C ± 20 °C	If > 10 % length > 1080 °C, then RED	If > 20 % length > 1070 °C, then AMBER	If > 95 % length < 1070 °C, then GREEN
Fin temp achievement	Temp values along length of strip.	Aim temp 860 °C ± 20 °C	If > 2 % length > 835 °C, then RED	If > 5 % length > 840 °C, then AMBER	If > 95 % length >= 840 °C, then GREEN
FIN MILL Scrubbers	Scrubbers used & operating pressure		low pressure < 25bar	F6&F7 bot or F5&F7 Bot	F5&F6 bot
Fin mill edge sprays	No of edge sprays used		None used	1 used	> 1 used
Fin Mill ISS Operation	Operational status		Manual > 10 % ISS restricted	Manual <10 % ISS restricted	Computer Or Manual Extra ISS
F5-F8 work roll type	Work roll type		All HSS or HSS in F5 or F6	Mixed. Not red or green	All HiCr or HiCr in F5&F6

Chapter 10. Recommendations

There are several recommendations that can be implemented from this thesis. The first is that to limit scale formation the temperature needs to be kept low. Values greater than 1070 °C for the average Roughing Mill and above 1050 °C for the average Crop Shear temperature are considered high, with values greater than this increasing the chance of scale formation. As the temperature increases more scale suppression measures are required to limit scale formation, with high temperatures more likely to generate a greater amount of scale even with fully functional scale suppression systems in place.

Chemistry is also a significant factor in scale formation, with Phosphorus being the most significant of the chemistry variables. It is recommended that the chemistry specification for Phosphorus be limited to a maximum value of 0.015 % rather than 0.020 % to limit scale formation. Slabs with higher values should be treated with particular care when being processed through the Hot Mill to limit scale formation.

It is recommended that the monitor that has been deployed continue to be utilised and updated, to aid in the removal of scale as a significant defect and to improve communication between business units.