Data Mining Analysis of the Relationship Between Input Variables and Hot Metal Silicon in a Blast Furnace

PIERRE WIKSTRÖM

MASTER OF SCIENCE PROGRAMME
Chemical Technology

Luleå University of Technology
Department of Chemical Engineering and Geosciences
Division of Process Metallurgy
Abstract

The purpose of this report is to get a better understanding of the relationships between input variables and silicon's behaviour in BlueScope Steel's Port Kembla No. 5 Blast Furnace. Over a period of 13 years stretching from 1991 to 2004, potentially important variables affecting the hot metal silicon content of the hot metal produced in the furnace were collected on a daily basis. A process analysis was performed using the data mining program, 'Weka', which is a freeware program developed by the University of Waikato, New Zealand. The data collected were altered, and different datasets compiled, in order to analyse the data without interferences and to reveal previously unknown and hidden relationships. A post-pulverized coal injection dataset was compiled to remove historical variations of limited relevance to current operations. An important variable affecting the hot metal silicon was found to be the amount of Quartz charged into the furnace. After finding the importance of Quartz charge, a short term, high frequency dataset was collected and analysed to find the true correlations. Quartz is charged into the furnace to change the slag basicity, but this project shows that a significant fraction of the Quartz transfers into the hot metal as silicon instead of transferring into the slag as silica as intended. The original data was compiled into a day-to-day difference dataset where each instance contained a difference between two consecutive days. In the difference dataset analysis, the long term variation was removed and the impacts of operational variables on hot metal silicon were revealed. The most significant variable affecting hot metal silicon was hot metal temperature. Hot metal silicon was corrected by hot metal temperature and the analysis was continued. The analysis showed that the relationship between hot metal sulphur and hot metal silicon was not causal, and could be explained by the impact of hot metal temperature. A data mining framework brought valuable new insight into hot metal silicon control in the blast furnace and further effort in this area is strongly recommended.
Table of contents

1 INTRODUCTION ................................................................................................................. 1

1.1 PURPOSE ....................................................................................................................... 1

2 THEORY .............................................................................................................................. 2

2.1 Fundamentals of the iron making process ................................................................. 2
2.2 Transfer of silicon into hot metal ............................................................................. 3
2.2.1 Generation of SiO gas ......................................................................................... 4
2.2.1.1 From coke ash ............................................................................................ 4
2.2.1.2 From slag ................................................................................................... 7
2.2.1.3 From pulverized coal ................................................................................. 8
2.2.2 SiO gas Adsorption ......................................................................................... 9
2.2.2.1 SiO gas to Hot metal ................................................................................ 9
2.2.2.2 SiO gas to slag ......................................................................................... 11
2.2.2.3 Slag to metal ............................................................................................ 11
2.3 Relationship between silicon and other species ............................................... 12
2.3.1 Titanium’s relationship to silicon ...................................................................... 12
2.3.2 Sulphur’s relationship to silicon ...................................................................... 13
2.3.3 Sulphur and Manganese relationships with temperature ......................... 14
2.4 Data mining ............................................................................................................... 14
2.4.1 Missing Values .............................................................................................. 15
2.4.2 Classification ................................................................................................... 15
2.4.3 Clustering ......................................................................................................... 16
2.4.3.1 Partitioning methods .............................................................................. 17
2.4.3.2 Hierarchical methods ............................................................................. 17
2.4.3.3 Density-based methods .......................................................................... 17
2.4.3.4 Grid based methods ............................................................................... 17
2.4.3.5 Model based methods ............................................................................ 18

3 Implementation and Results ......................................................................................... 19

3.1 Preliminary analysis ................................................................................................. 19
3.2 Pre and post pulverized coal injection dataset ................................................. 21
3.3 Difference dataset .................................................................................................. 27

4 Discussion .................................................................................................................... 32

5 Conclusions .................................................................................................................. 35

References ...................................................................................................................... 36
Table of figures

Figure 1. Schematic picture of the inside of a blast furnace. Husslage (8)................................. 2
Figure 2. Schematic drawing of silicon transfer in the lower zone of the blast furnace.......... 3
Figure 3. Partial pressures of SiO for univariant equilibria. Turkdogan (20)......................... 6
Figure 4. Equilibrium conditions for the evolution of SiO (g) from molten slag.
Tokuda (9)................................................................................................................................. 7
Figure 5. Relationship between C/S of sample and amount of Si transfer.
Yoshiyuki et al (24).................................................................................................................. 9
Figure 6. 2 - log P_{O2} – T diagram for Si – O system. Tokuda et al (18)............................... 11
Figure 7. The dependence of titanium partition between slag and iron on the silicon content of iron. Biswas (3).............................................................. 13
Figure 8. Schematic picture that describes the structure of data mining............................... 15
Figure 9. Classification problem. Dunham (5)........................................................................ 16
Figure 10. A 2 – D plot of data showing three cluster. Each cluster is marked with a cluster center +. Han and Kamber (6)................................................................. 16
Figure 11. Clustered instances plotted over time............................................................... 21
Figure 12. Hot metal silicon plotted over time...................................................................... 22
Figure 13. Production plotted over time.................................................................................. 22
Figure 14. Efficiency plotted over time................................................................................ 23
Figure 15. Variable plot made in statistica............................................................................ 25
Figure 16. Correlations between HMSi and Quartz, and between Quartz and Quartz at different time lags............................................................................ 26
Figure 17. Variable plot made in Statistica showing correlations between important variables.................................................................................. 28
Figure 18. The comparison between the correlation of HMSi and hot metal Sulphur and corrected HMSi and hot metal Sulphur...................................................... 30
Figure 19. The theoretical predictions of hot metal sulphur.................................................. 31
1 Introduction

The most important operation parameters of the iron making blast furnace are the amount of hot metal produced and the capability of utilizing inserted fuels. Varying with hot metal and fuel prices, these parameters will determine in what way the furnace is operated. Apart from these parameters, the hot metal quality is an essential part of the process. Hot metal silicon (HMSi) tends to get most attention because of the relative potential profit that is waiting if any decrease in hot metal content could be accomplished.

Over the years, a lot of research has been done on the subject. This has given us a better understanding of silica’s behaviour in the furnace, from insertion of SiO\textsubscript{2} in ore, coke ash and raw quartzite, to descent through the furnace and the produced silicon in hot metal and silica in slag. Despite all the knowledge, HMSi often remains above optimum levels in today’s blast furnaces. This is because iron making is such a complex process and the variables that affect the HMSi are too many, too unknown and vary between different furnaces. One thing that does not vary is that all furnaces collect thousands of different data to be able to operate at demanded levels. These data, after being analysed by the operator, are stored and often left to be. This is where data mining (DM) is suitable for developing knowledge out of the information hidden in the collected data. Analysing data with multiple variables over a long time period is time consuming. Most important is that the data contains hidden patterns between variables that cannot be seen with the naked eye. The intention of using DM tools is to solve these problems and help find unknown patterns and relationships.

1.1 Purpose

The purpose of this report is to get a better understanding of the correlation between input variables and silicon’s behaviour in the blast furnace, with particular reference to BlueScope Steel’s Port Kembla No. 5 Blast Furnace.
2 Theory

2.1 Fundamentals of the iron making process

The main purpose of the iron making process in a Blast furnace is to reduce iron ore from Fe$_2$O$_3$ (s) to Fe (l) and to remove the gangue present in the raw materials. The principal technique is to insert porous ore agglomerates at the top in forms of pellets and sinter or pure iron ore, so called lump ore. Coke and pulverized coal (PC) serve as fuel to heat up and reduce the iron ore. The ores and coke are fed into the furnace via a transfer chute at the top, which distributes them into separate layers. Hot air is blasted into the lower zone of the furnace through the tuyeres, which are cooled with PC carried by nitrogen. The high-pressure blast, combined with combustion of coke, causes a cavity next to the tuyeres called the raceway. In the raceway, coke and PC react with oxygen and combust resulting in the generation of reducing gas. This hot reducing gas travels up to the top of the furnace passing through the descending layers of ore and coke.

Figure 1. Schematic picture of the inside of a blast furnace. Husslage (8)
The raw materials are heated up as they descend down through the furnace. When the burden material reaches temperatures over 1100 °C it starts to soften and completely melts at 1450 °C. The region of the furnace where softening and melting occurs is called the cohesive zone. The metal, gangue and other slag components liquefy and drip through the mobile coke bed. When the coke particles arrive at tuyere level they either combust at the raceway or they settle down in the stationary coke pile called the Deadman. The liquids descend further through the deadman until they reach the hearth where iron settles down at the bottom and slag floats on top. Slag and metal are tapped through the tapholes.

2.2 Transfer of silicon into hot metal

There are three major sources of SiO$_2$ in the blast furnace, ore gangue, coke ash and pulverized coal. SiO$_2$ contents in ores vary considerably, but those in coke ash are roughly similar, though the ash content itself can vary substantially from coke to coke. The contribution of each source is dependent on the kinds of ore and coke, as well as on operational conditions of the furnace. Coke ash is often thought as the largest contributor to the high silicon content of hot metal although the biggest amount of SiO$_2$ comes with the ores. This is because of the high activity of silica in coke compared to the ores, where the silica also often is bound to alumina and lime, and also the higher temperatures experienced by the coke. Coke is used in the furnace as fuel. Ore and coke is inserted in separate layers into the furnace. When the coke descends through the furnace and reaches below the dripping zone down to the raceways where hot reducing gas is blown into the furnace at high speed and high pressure, the coke burns and produces CO, CO$_2$ and leaves the ash free for reduction.

![Figure 2. Schematic drawing of silicon transfer in the lower zone of the blast furnace.](image)

The ash with 50 % SiO$_2$ reacts under the high pressure and temperature and some of the SiO$_2$
Theory

gasifies to SiO. The ore that has been layered out between the coke layers descends through the furnace to a temperature above melting temperature and start to liquefy. As the ore starts to melt into liquid slag and metal it meets the up flowing gas containing SiO. The metal and slag absorbs the gas and silicon dissolves into the hot metal. The iron containing silicon in varying amounts descends to the hearth where silicon behaves as a poor reducing agent for slag oxides and fails to attain equilibrium with the slag components.

2.2.1 Generation of SiO gas

In this section the generation of SiO gas from the different sources are explained.

2.2.1.1 From coke ash

When coke enters the raceway it burns and produces carbon as CO and CO₂. The remains after combustion of coke are coke ash, which has a high content of SiO₂. In the blast furnace under high temperature combustion conditions some fraction of the SiO₂ content of the ash gasifies to SiO gas.

Yamagata et al (22), (23) reported experiments where the weight losses of cokes heated in CO/Ar atmosphere were carried out. The range of 1350 °C – 1750 °C was used to verify a mathematical simulation of silica-coke reactions. The chemical processes included in the model were two simultaneous intraparticle reactions:

\[
\text{SiO}_2 \text{(s)} + C \text{(s)} \rightarrow \text{SiO(g)} + \text{CO(g)} \quad (2.1)
\]

\[
\text{SiO}_2 + 3C(s) \rightarrow SiC(s) + 2CO(g) \quad (2.2)
\]

The physical processes included were intraparticle diffusion of the generated gas, and diffusion of gas in the peripheral coke gas film. An increase in total pressure and partial pressure of CO reduced the extent of both reactions. The formation of SiC and generation of SiO increased with increased temperature.

Mathieson et al (10) studied the formation of silicon monoxide in the raceway by calculating equilibrium phase diagrams for the combustion of high and low ash cokes under conditions of high and low raceway pressures. When coke ash was assumed to consist of SiO2 only, the coke completely gasified to SiO and CO under restricted conditions of coke to blast ratio and at temperatures greater than 1600 – 1650 °C, depending on system pressure. When coke ash was assumed to contain SiO2 and Al₂O₃ the generation of SiO (g) was lowered because of the formation of solid and liquid aluminosilicate compounds. For this case the minimum temperature of complete SiO₂ gasification was raised to 1700 °C.

Turkdogan et al (20) have done experimental investigations and thermodynamic analysis of the
Theory

coke ash reaction and expressed reactions (2.1) and (2.2) as:

$$\log \left( \frac{P_{\text{SiO}} \cdot P_{\text{CO}}}{a_{\text{SiO}_2}} \right) = -\frac{35896}{T} + 17.957$$  \hspace{1cm} (2.3)

$$\log \left( \frac{(P_{\text{CO}})^2}{a_{\text{SiO}_2}} \right) = -\frac{31600}{T} + 17.760$$  \hspace{1cm} (2.4)

The authors also expressed the reactions equilibria for the case when SiC is the stable phase rather than SiO\textsubscript{2}. SiC and C were assumed to be of unit activity and the partial pressures of the gaseous species were in atmospheres and the activity of the SiO\textsubscript{2} in coke ash was 0.5.

$$\text{SiC}(s) + \text{CO}(g) \rightarrow \text{SiO}(g) + 2\text{C}(s)$$  \hspace{1cm} (2.5)

$$\log \left( \frac{P_{\text{SiO}}}{P_{\text{CO}}} \right) = -\frac{4276}{T} + 0.197$$  \hspace{1cm} (2.6)

$$2\text{SiO}_2(s) + \text{SiC}(s) \rightarrow 3\text{SiO}(g) + \text{CO}(s)$$  \hspace{1cm} (2.7)

$$\log \left( \frac{(P_{\text{SiO}})^3 \cdot P_{\text{CO}}}{(a_{\text{SiO}_2})^2} \right) = -\frac{76088}{T} + 36.110$$  \hspace{1cm} (2.8)

At CO pressures of 1 and 4 atmospheres and for $a_{\text{SiO}_2}=0.5$, $a_{\text{c}}=1$, and $a_{\text{SiC}}=1$, SiC(s) was the stable phase above 1527 and 1671 °C respectively. To the left of curve a – c SiO\textsubscript{2} is the stable phase whilst SiC is stable between the curves c and b as can be seen in figure 3.
Theory

From the picture:

a: \( \text{SiO}_2 \text{ + C} \rightarrow \text{SiO} + \text{CO} \)  \hspace{1cm} (2.9)
b: \( \text{SiC} + \text{CO} \rightarrow \text{SiO} + 2\text{C} \)  \hspace{1cm} (2.10)
c: \( 2\text{SiO}_2 + \text{SiC} \rightarrow 3\text{SiO} + \text{CO} \)  \hspace{1cm} (2.11)

The conclusion that was drawn was that much of the silica in coke ash may have been transferred to the descending metal droplets via \( \text{SiO} \) by the time the coke has reached the raceway. \( \text{SiO} \) generation in a CO rich atmosphere starts at \( 1400^\circ \text{C} \) and the rate accelerates with increasing temperature. Besides \( \text{SiO} \) generation, there is formation of \( \text{SiC} \) which later reacts with \( \text{CO} \) or \( \text{SiO}_2 \) to yield \( \text{SiO} \).

The speculation that silicon transfer could be controlled by gas phase mass transfer was made by Turkdogan et al (20). Ozturk and Fruehan (11) state in a study that the reaction between \( \text{SiO} \) and carbon dissolved in iron can be controlled by gas phase mass transfer and chemical kinetics on the metal surface. The flux of \( \text{SiO} \) \( (J_{\text{SiO}}) \) to the surface is given by:

\[
J_{\text{SiO}} = \frac{m}{RT} (P_{\text{SiO}}^B - P_{\text{SiO}}^S)
\]

(2.12)

If gas phase mass transfer of \( \text{SiO} \) controls the rate, then the rate should vary with gas phase mass transfer coefficient \( m \) and \( P_{\text{SiO}} \). Further the rate should not be a strong function of temperature or the composition of the melt. \( P_{\text{SiO}}^B \) and \( P_{\text{SiO}}^S \) is the \( P_{\text{SiO}} \) in the bulk gas and at the surface. \( m \) is given by Ozturk and Fruehan et al (12) according to \( m = \frac{D}{X} \) where \( X \) is the distance between the bottom of the crucible in which the \( \text{SiO} \) gas is generated and to the top of
the melt. $D$ is the interdiffusivity of SiO. The rate of the transfer is expected to increase with decreasing value of $X$ if gas phase mass transfer is controlling. If chemical kinetics were controlling then the transfer should not be a strong function of the experimental geometry $X$, but instead be a strong function of temperature and vary with the melt composition. The conclusion was that the transfer of silicon was controlled by gas phase mass transfer. They based the conclusion with the fact that in the experiments the rate was independent of temperature and affected by $P_{\text{SiO}}$ and by experimental geometry. Chemical kinetics was neglected because sulphur did not decrease the rate. If chemical kinetics was important then sulphur which is a surface active substance should decrease the rate of transfer.

### 2.2.1.2 From slag

Turkdogan et al (20) stated that the SiO is generated at a gas-slag interface by the reaction:

$$\text{(SiO}_2\text{)} + \text{CO(g)} = \text{SiO(g)} + \text{CO}_2\text{(g)}$$  \hspace{1cm} (2.13)

Ito and Tokuda (9) shows from figure 4 that for a partial pressure of SiO (g) of $10^{-4}$ atm that the shaded area represents the conditions where SiO (g) can be formed. They concluded that a low basicity slag with a silica activity of 0.1 is always a source for SiO (g) formation.

![Figure 4. Equilibrium conditions for the evolution of SiO (g) from molten slag. Tokuda (9)](image)

Ozturk and Fruehan (12) have done studies of SiO (g) generation from slag. Experiments on pure silica, CaO – SiO2 and CaO - Al2O3 – SiO2 at up to 1700 °C were carried out. The formation of SiO by the reaction of silica in slags with CO is not controlled by gas or liquid phase mass transfer according to Ozturk and Fruehan (12), instead chemical kinetics at the gas-slag interface appears to be controlling.

The generation of SiO (g) through the reaction of graphite with a CaO - Al2O3 – SiO2 slag has been described by Yaginuma et al (21). They indicated that addition of SiC increases the
interfacial area between the slag and graphite following an increase in the rate and quantity of \( \text{SiO} \; (g) \) generated. The partial pressure of \( \text{CO} \) was observed to have an insignificant effect.

### 2.2.1.3 From pulverized coal

The \( \text{SiO}_2 \) reduction reaction of pulverized coal is reduction of ash \( \text{SiO}_2 \) by carbon and is assumed to be simultaneous reactions:

\[
\text{SiO}_2(s) + \text{C}(s) \rightarrow \text{SiO}(g) + \text{CO}(g) \quad (2.14)
\]

\[
\text{SiO}_2(s) + 3\text{C}(s) \rightarrow \text{SiC}(s) + 2\text{CO}(g) \quad (2.15)
\]

Yoshiyuki et al (24) derived the \( \text{SiO} \; (g) \) generation rate from pulverized coal under intensive coal injection. The experiments on \( \text{SiO} \; (g) \) generation from pulverized coal were carried out with char samples packed in a crucible under flow of reducing gas. Table 1 shows the chemical composition of pulverized coal char used for samples. In order to investigate effects of pulverized coal ash composition on the reaction rate for generating \( \text{SiO} \; (g) \), they added an aqueous solution with calcium hydroxide dissolved in aqueous nitric acid solution to the raw char. Figure 5 shows the effects of ash basicity on $\Delta [\text{Si}]_{PC}$ at 1600 °C holding temperature and for 3 hours holding time. $\Delta [\text{Si}]_{PC}$ in raw char is susceptible to the \( \text{SiO}_2 \) concentration, and in particular, char A with high \( \text{SiO}_2 \) concentration indicates a large reduction rate due to basicity.

**Table 1. Experimental conditions. Yoshiyuki et al (24)**

<table>
<thead>
<tr>
<th>Sample</th>
<th>C (mass%)</th>
<th>Ash (mass%)</th>
<th>Ash composition (mass%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>SI(\text{O}_2)</td>
</tr>
<tr>
<td>Char A</td>
<td></td>
<td>68.2</td>
<td>27.0</td>
</tr>
<tr>
<td></td>
<td>+ Flux</td>
<td>50.5</td>
<td>20.0</td>
</tr>
<tr>
<td></td>
<td>+ Flux</td>
<td>34.2</td>
<td>25.8</td>
</tr>
<tr>
<td>Char B</td>
<td></td>
<td>77.2</td>
<td>7.6</td>
</tr>
<tr>
<td></td>
<td>+ Flux</td>
<td>60.0</td>
<td>21.3</td>
</tr>
<tr>
<td></td>
<td>+ Flux</td>
<td>46.6</td>
<td>23.4</td>
</tr>
<tr>
<td>Char C</td>
<td></td>
<td>89.8</td>
<td>4.3</td>
</tr>
<tr>
<td></td>
<td>+ Flux</td>
<td>89.8</td>
<td>4.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Flux</th>
<th>Ca((\text{OH})_2) + (\text{HNO}_3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atmosphere</td>
<td>Temperature : 1500, 1600°C</td>
</tr>
<tr>
<td></td>
<td>CO gas flow rate : 2 L/min</td>
</tr>
<tr>
<td></td>
<td>Heating time : 1,2,3 hour</td>
</tr>
</tbody>
</table>
The results showed that the rate constant of pulverized coal char is three times larger than the rate of coke that Yamagata et al (22) reported. This indicated that pulverized coal char is likely to generate SiO (g) because the contact area with carbon is large as compared to coke.

2.2.2 SiO gas Adsorption

In this part the absorption of SiO gas by the slag and the metal are explained. The direct transfer of SiO\textsubscript{2} from slag to Si in metal is also described.

2.2.2.1 SiO gas to Hot metal

As iron droplets trickle down through the dripping zone it meets raceway gases containing SiO gas. SiO is reduced by the carbon dissolved in iron and silicon dissolves into the iron according to the following reaction.

\[ \text{SiO}(g) + C = \text{Si} + \text{CO}(g) \] (2.16)

When the iron is carbon-saturated, the extent of this reaction is limited by the amount of generated SiO gas so the P\textsubscript{SiO} has an effect on the adsorption. Batra (2) showed the effect of PSiO with the reactions equilibrium equation.

\[
\log \frac{a_{\text{Si}} \cdot P_{\text{CO}}}{a_{C} \cdot P_{\text{SiO}}} = \frac{-3925}{T} + 4.25 \tag{2.17}
\]

The temperature affects the adsorption of silicon into iron in the way that increasing temperature increases the extent of the reaction. The reaction time is one of the most important factors that affect the rate of adsorption. The longer time it takes for the iron droplet to descend from the cohesive zone to the hearth, the longer time the droplets are exposed to SiO gas. This
Theory

makes production a relatively good measure of reaction time. A high productivity gives short reaction time for the iron droplet and SiO and consequently a lower silicon content of the hot metal produced. The size of the iron droplet is important because the diffusion distance as well as the surface area affects the reaction. When iron is not fully saturated with carbon the amount of carbon in iron has an effect on the reaction (2.16). This does not play an important role in the blast furnace because of the fact that iron often has sufficient amounts of carbon when reaction (2.16) takes place.

Tsuchiya et al (19) claim that the transfer rate of silica is regulated by the chemical reaction between the gas and the surface of the molten iron and that the adsorption phenomenon on the surface of the liquid iron plays an important role. They estimated the amount of silicon transfer under the condition that the iron was close to fully saturated by carbon. The expression used was:

\[
\frac{d\%Si}{dt} = k_f P_{SiO} - k_b P_{CO} \%Si
\]  
(2.18)

Tsuchiya et al (19) assume the residence time of iron below the cohesive zone to be 1 min for every meter down to the hearth. Because of no transport in backward direction and arbitrary surface area the expression becomes as below where \(A\) is the interfacial area (cm\(^2\)) between gas and liquid iron and \(M\) is weight in grams of liquid iron:

\[
\frac{d\%Si}{dt} = k_f \frac{A}{M} P_{SiO}
\]  
(2.19)

According to the Iron and Steel Institute of Japan (17) the temperature dependence is:

\[
k_f = 1.1 \cdot 10^{-5} \cdot \exp\left(-\frac{65000}{RT}\right)
\]  
(2.20)

Figure 6 is a log \(P_{O_2} - T\) diagram of the Si-O system from Tokuda et al (18). The hatched part indicates the region where SiO is stable with partial pressure of more than \(10^{-3}\) atm. Arrow I shows where SiO gas is oxidized back to SiO\(_2\) as it travels upward from the tuyere region. When SiO reacts with liquid iron as shown by arrow II in the figure SiO is reduced to silicon in iron.
2.2.2.2 SiO gas to slag

From a study by Ozturk et al (13) the following reactions takes place at the gas-slag interface:

\[
\text{SiO}(g) + \text{CO}_2(g) = (\text{SiO}_2) + \text{CO}(g)
\]  
(2.21)

The results from a study by Ito et al (9) is that a high basicity slag can always absorb SiO (g), therefore the slag chemistry is important for the slag's absorption capacity of SiO (g). SiO (g) transfer to the surface of a calcium aluminate slag has been investigated by Ozturk and Fruehan (13). Slag with compositions of CaO, Al2O3 and CaO (Al2O3) were reacted at 1550 and 1650 °C respectively. The rate of increase of SiO2 in the slag phase was consistent with rate control by gas phase mass transfer of SiO (g) across the gas-slag boundary. The rate of absorption of SiO (g) by calcium aluminate slags may also increase as the oxygen potential of the reacting gas mixtures increases. Rist et al (16) have reported an increase in the absorption rate of SiO(g) by calcium aluminate slags at 1600 °C in going from a CO + SiO gas mixture to a CO + 0.5 % CO2 +SiO gas mixture, \( P_{SiO} = 1.2 \times 10^{-3} \text{ atm} \) at constant total flow rate. The argument put forward by these authors is that the activity of SiO\(_2\) (at the slag surface) is higher in the gas mixture with the higher oxygen potential.

2.2.2.3 Slag to metal

It has been proposed by Turkdogan et al (20) and Pompfret and Grieveon (15) that the reaction in slag with carbon in iron follows the reaction:

\[
(\text{SiO}_2) + 2\text{C} = \text{Si} + 2\text{CO}(g)
\]  
(2.22)
Theory

The silicon content in hot metal is a good measure of the heat conditions in a blast furnace. The rate of silicon transfer through slag-metal reaction strongly depends on temperature. According to Ashizuka et al (1) the rate of silica reduction in a graphite crucible increases with increasing slag – graphite interfacial area. The ratio \( r = \frac{S_{s-g}}{S_{s-m}} \) was used where \( S_{s-g} \) is the interfacial area between the slag and the graphite crucible and \( S_{s-m} \) is the interfacial area between the slag and the metal. They found a relationship between silica reduction and ratio \( r \). The rate of silica reduction was proportional to the activity of silica at \( r \) values smaller than 30 and the reaction was considered to be chemically controlled. At greater \( r \) values the diffusion of silicate ions in slag is considered to be the rate determining step. Because of this the rate of silicon transfer from the stagnant molten slag into the molten iron bath in the blast furnace hearth is very slow. The maximum rate at 1600 ºC is \( 5 \cdot 10^{-5} \) moles/min cm\(^2\) which gives a rise in the silicon content of 0.02-0.03 pct Si. This is one order less than the silicon content in hot metal produced in industrial blast furnaces.

2.3 Relationship between silicon and other species

2.3.1 Titanium’s relationship to silicon

The reduction of silica is intimately connected with the reduction of titania. In both laboratory experiments and plant data it has been found that the titanium content of the hot metal increases with increasing silicon content and increasing slag basicity. The equation relevant for silicon’s and titanium’s relationship is described by Delve (4) and Hess et al (7) as:

\[
(TiO_2) + Si = Ti + (SiO_2)
\]  

\[
K = \frac{a_{\text{Ti}} \cdot a_{\text{(SiO}_2)}^2}{a_{(TiO_2)} \cdot a_{\text{Si}}}
\]  

\[
\frac{a_{\text{Ti}}}{a_{(TiO_2)}} = K \cdot \frac{a_{\text{Si}}}{a_{(SiO}_2)}
\]

From the equation, the higher the temperature (\( K \) increases with temperature) and the higher the basicity the greater the reduction of titania. The extent of titania reduction from the slag depends on the rate of its reduction by silicon in iron. Silicon enters the iron above the hearth level via gaseous SiO and SiS. Titanium does not form such gaseous species. Therefore the reduction of titanium from the slag occurs during the passage of silicon bearing iron droplets through the slag layer. However since Si in iron is a poor reducing agent in the hearth, the titania reduction is restricted. Hence the apparent linear relationship between the titanium and
silicon contents of the metal.

Figure 7. The dependence of titanium partition between slag and iron on the silicon content of iron. Biswas (3)

### 2.3.2 Sulphur’s relationship to silicon

The equilibrium partition ratio of Sulphur can, according to the iron and steel institute of Japan (17), be expressed as:

\[
\log_{10}^{0} = 1.35 \cdot 1.79 \cdot (\% \text{CaO}) + 1.24 \cdot (\% \text{MgO}) \\
- 1.66 \cdot (\% \text{SiO}_2) - 0.33 \cdot (\% \text{Al}_2 \text{O}_3) - \log P_{\text{CO}} - \frac{8130}{T} + 4.15
\] 

(2.26)

The units of CO partial pressure \((P_{\text{CO}})\) and temperature \((T)\) are atmospheric pressure (atm) and absolute temperature (K) respectively. CO partial pressure is estimated by the below equation.

\[
P_{\text{CO}} \text{ (atm)} = [1 + 0.967 \cdot \text{blast\_pres\_sure (kg/cm}^2, \text{gauge}) + 0.192 \cdot \text{distance\_f\_rom\_tuyere\_to\_taphole (m)} - 0.75]
\]

(2.27)

On the other hand the practical partition ratio \((L_s)\) is calculated from the chemical composition of slag and metal tapped out of the furnace:

\[
L_s = (\% \text{S})f_s \cdot [\% \text{S}]
\]

(2.28)

\[
\log f_s = 0.11[\% \text{C}] - 0.026[\% \text{Mn}] + 0.063[\% \text{Si}]
\]

(2.29)

\(F_s\) represent the activity coefficient. Operational condition index \(R_s = (L_s/L_s^0) \cdot 100\) is the amount of FeO which has descended to the hearth in a furnace with deteriorated condition.
2.3.3 Sulphur and Manganese relationships with temperature

Pochvisnev and Kurunov (14) have determined in statistical analysis that the highest correlation coefficient with hot metal temperature is possessed by the (Mn)/[Mn], closely followed by the [Si]/[S] and (S)/[S] ratios. The reduction of slag MnO by silicon is a function of temperature and the metal sulphur is intimately related with metal Mn. Hence it is according to Biswas (3) expected that the metal temperature will be more closely related with the metal Mn and S.

In case an excess of unreduced iron oxides or re oxidized pig iron enters the hearth due to accelerated stock descent or inadequate reduction, their reduction in the hearth will decrease the metal temperature, retard the reduction of slag MnO by silicon and hence increase the metal sulphur. In the blast furnace, silicon, manganese and sulphur do not reach equilibrium with respect to carbon saturated iron. They instead are close to equilibrium with each other. Both silicon and sulphur enter the metal at a faster rate than manganese and reach maximum values at places away from the hearth. In the hearth, silicon in iron behaves as a poor reducing agent for manganese reduction, which controls sulphur removal.

2.4 Data mining

Society produces huge amount of data e.g. business, science, medicine, economics, sport, geography, and environment. From business processes, process control and increasingly computerized manufacturing with automation systems loads of data are produced. These data lose their value from failure to understand and extract as much knowledge as possible. This data is a potentially valuable resource however raw data is useless. Techniques for extracting information from this raw data become important. Data Mining is the extraction of implicit, previously unknown and potentially useful information from data. The information in the data is patterns underlying the data which often are complex and in low concentration of the whole dataset. There is a famous data mining saying, computers promised fountains of wisdom but delivered floods of data.

Programs are needed to detect patterns and regularities in the data and strong patterns can be used to make predictions. However most patterns are not interesting and patterns may be inexact or spurious if data is missing. The technical basis for data mining is algorithms for getting structural descriptions from examples. The structural descriptions represent patterns, which can be used to predict the outcome of new situations or understand and explain how the prediction is derived. The knowledge discovery process consists of several stages, data selection, data cleaning, transformation, analysis, interpretation and evaluation. The importance of selecting the data for analysis should not be underestimated. Choosing the wrong attributes can make the patterns you are looking for concealed behind patterns of less interest, often because of unwanted dependencies between attributes. Noise and missing data is another problem that distorts the results and it’s important to clean the data before analysis. Data mining can be described as a combination of three related topics, statistics, machine learning AI and data base ware housing.
2.4.1 Missing Values

Data is often incomplete, noisy and inconsistent. Data cleaning routines attempt to fill in missing values, smooth out noise while identifying outliers, and correct inconsistencies in the data.

One way is to fill in the values manually, in general a very time consuming work given a large dataset with many missing values. Another way is to fill in the missing value with the attribute’s mean or the attribute’s mean in a certain group of instances with resemblance in attribute values.

The most popular way of solving the problem is by using the most probable value. Han and Kamber (6) state that the most probable value can be determined with regression, inference-based tools using Bayesian formalism or decision tree induction. This method uses the most information from the present data to predict missing values.

Noise is an error or unreal variance in a measured value. The solution is to smooth out the data and remove the noise using different techniques. One of these techniques is clustering where variances can be detected when similar instances are organized in groups, and outliers gather in the same cluster.

2.4.2 Classification

The most familiar and the most popular data mining technique is classification. Classification maps data into predefined groups or classes. The classes are often determined before examining the data. The algorithms used for making the classification requires that the classes are defined based on attribute values. The algorithms look at the characteristics of data already known to belong to the classes. Estimation and prediction may be viewed as classification techniques. Below in figure 9 we have a basic classification problem where the data is of the form $t = (x, y)$ where $0 < x < 8$ and $0 < y < 10$. In the figure (a) shows the predefined classes by
dividing the reference space, (b) provides sample input data and (c) shows the classification of the data based on the defined classes.

Figure 9. Classification problem. Dunham (5)

2.4.3 Clustering

Clustering is the process of grouping the data into classes or clusters so that objects within a cluster have high similarity in comparison to one another, but are very dissimilar to objects in other clusters. Dissimilarities are assessed based on the attribute values describing the objects. Often distance measures are used. There are several different clustering techniques such as, partition methods, hierarchical methods, density-based methods, grid-based methods and model based methods.

The choice of clustering algorithm depends both on the type of data available and on the particular purpose and application. If cluster analysis is used as a descriptive or exploratory tool, it is possible to try several algorithms on the same data to see what the data may disclose.

Figure 10. A 2 – D plot of data showing three cluster. Each cluster is marked with a cluster center +. Han
2.4.3.1 Partitioning methods

Partitioning methods construct k partitions of the data where each partition represents a cluster, which must contain at least one object. Each object must belong to exactly one group. When given the amount of partitions to construct, a partitioning method creates an initial partitioning. It then uses an iterative relocation technique to improve the partitioning by moving objects from one group to another. The general criterion of a good partitioning is that objects in the same cluster are close or related to each other, whereas objects of different clusters are far apart or very different.

There are two popular heuristic models: k-means algorithm where each cluster is represented by the mean value of the objects in the cluster and k-medoids algorithm where each cluster is represented by one of the objects located near the center of the cluster. These heuristic clustering methods work well for finding spherical-shaped clusters in small to medium-sized databases. To find clusters with complex shapes and for clustering very large data sets other methods are preferred.

2.4.3.2 Hierarchical methods

A Hierarchical method creates a hierarchical composition of the given set of data objects. This method suffers from the fact that once a merge or split of a cluster is made it can never be undone. This makes the computation easier by not worrying about a combinatorial number of different choices.

2.4.3.3 Density-based methods

Most partitioning methods cluster objects based on the distance between objects. Such clusters can only find spherical shaped clusters and encounter difficulty discovering clusters of arbitrary shapes. Other clustering methods have been developed based on the notion of density. Their general idea is to continue growing the number of clusters as long as the density (number of objects or data points) in the “neighborhood” exceeds some threshold; that is for each data point within a given cluster, the neighborhood of a given radius has to contain at least a minimum number of points. Such a method can be used to filter out noise (outliers) and discover clusters of arbitrary shape.

2.4.3.4 Grid based methods

Grid based methods quantize the object space into a finite number of cells that form a grid structure. All of the clustering is performed on the grid structure. Wave clustering is one algorithm using both grid-based and density based methods.
2.4.3.5 Model based methods

Model based methods hypothesize a model for each of the clusters and find the best fit of the data to the given model. A model based algorithm may locate clusters by constructing a density function that reflects the spatial distribution of the data points. It is also a way of automatically determining the number of clusters based on standards statistics, taking outliers into account and thus yielding robust clustering methods. Some clustering algorithms integrate the ideas of several clustering methods.
3 Implementation and Results

The process analysis was performed with the data mining program Weka which is a free program given out by the University of Waikato, New Zealand. Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can be applied directly to the dataset or called from a java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. The program is also well suited for developing new machine learning algorithms but that lies outside the limits of this project.

3.1 Preliminary analysis

From the thousands of different data measured and collected from number 5 blast furnace at Port Kembla, the most relevant variables was chosen. At first both input and output variables was chosen as well as independent and dependent variables. The data was collected and then processed into variables compatible to Weka. The final dataset consisted of 250 variables collected from blast furnace number 5 during the years 1991-2004. Each variable contains one mean value of each day which over 14 years gives a total of 5000 days times 250 variables, that is 1,250,000 measurements. The following variables were chosen:

Input variables

Coke
- Weights
- Size distributions
- Coals
- Coal reactivity’s
- Drum indexes
- Coal size distributions
- Ash chemistries

Sinter
- Weights
- Size distributions
- Chemistries

Pellet
- Types (kg/tHM)
- Types (%)

Fluxes
- Types (kg/tHM)

Pulverized Coal
- Coal types (kg/tHM)

Blast
- Humidity
- Volume
Implementation and Results

- Air rate
- Coke ovens gas rate

Output variables

<table>
<thead>
<tr>
<th>Hot metal Chemistry</th>
<th>Carbon (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Manganese (%)</td>
</tr>
<tr>
<td></td>
<td>Phosphor (%)</td>
</tr>
<tr>
<td></td>
<td>Silicon (%)</td>
</tr>
<tr>
<td></td>
<td>Sulphur (%)</td>
</tr>
<tr>
<td></td>
<td>Titania (%)</td>
</tr>
<tr>
<td></td>
<td>Standard deviation of Silicon</td>
</tr>
<tr>
<td></td>
<td>Standard deviation of Sulphur</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Temperature</th>
<th>Hot metal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hot metal standard deviation</td>
</tr>
<tr>
<td></td>
<td>Blast</td>
</tr>
<tr>
<td></td>
<td>Top</td>
</tr>
<tr>
<td></td>
<td>Flame</td>
</tr>
<tr>
<td></td>
<td>Stave heat loads</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Slag</th>
<th>Chemistries</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Volume</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Miscellaneous</th>
<th>Furnace availability (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Production</td>
</tr>
<tr>
<td></td>
<td>Liquid volume</td>
</tr>
<tr>
<td></td>
<td>Deadman cleanliness index</td>
</tr>
<tr>
<td></td>
<td>Carbon solution loss</td>
</tr>
<tr>
<td></td>
<td>Permeability</td>
</tr>
<tr>
<td></td>
<td>Tuyere heat flow ratio</td>
</tr>
<tr>
<td></td>
<td>Eta CO</td>
</tr>
<tr>
<td></td>
<td>Gas indexes</td>
</tr>
</tbody>
</table>

The full dataset contain missing data and incorrectly measured data which is a result of measuring equipment failures. The data also contain days with process shutdowns or with serious problems with the process. All this erroneous data brings problems to the further analysis. Therefore cluster analysis is implemented on the dataset. To handle process problems the dataset was clustered considering the process capability as the three parameters: hot metal quality, productivity and efficiency. Silicon in hot metal represents the hot metal quality, whilst Eta CO explained the efficiency and finally the production of hot metal represented the productivity. The clustering technique used was a maximisation minimisation technique and the resulting plot of clusters can be seen in figure 11:
Figure 11. Clustered instances plotted over time.

As can be seen from the figure no clusters are divided in time after 2000. This can be explained by the difference in process stability comparing 1991-2000 to 2000-2004 where the latter period has a more stable process. This and the fact that a new source of silica was used in the furnace in 2002, led to a dataset divided in 2 parts. The new source of silica was pulverized coal which started in May 2002, therefore the dataset was divided in pre and post PCI:

3.2 Pre and post Pulverized coal injection dataset

The dataset was split according to the following dates:

Pre PCI - August 1991 - May 2002

Post PCI - May 2002 - October 2004

Termination of bad process days on the post PCI dataset was accomplished with cluster analysis. The results below in figure show the days divided into 7 clusters, each cluster with similarities in production, efficiency and hot metal quality.

Post PCI from 4 May 2002 - 20 October 2004
Figure 12. Hot metal silicon plotted over time.

Figure 13. Production plotted over time.
Implementation and Results

Figure 14. Efficiency plotted over time

Table 2. Mean values of clusters in figures 12, 13 and 14.

<table>
<thead>
<tr>
<th>Clusters</th>
<th>Production</th>
<th>Efficiency</th>
<th>HMSi</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>6566</td>
<td>49.1</td>
<td>0.73</td>
</tr>
<tr>
<td>1</td>
<td>6910</td>
<td>50.2</td>
<td>0.60</td>
</tr>
<tr>
<td>2</td>
<td>7327</td>
<td>51.1</td>
<td>0.50</td>
</tr>
<tr>
<td>3</td>
<td>2185</td>
<td>35.4</td>
<td>0.94</td>
</tr>
<tr>
<td>4</td>
<td>5898</td>
<td>48.5</td>
<td>1.00</td>
</tr>
<tr>
<td>5</td>
<td>2913</td>
<td>49.7</td>
<td>0.53</td>
</tr>
<tr>
<td>6</td>
<td>7094</td>
<td>50.0</td>
<td>0.43</td>
</tr>
</tbody>
</table>

Clusters 3 and 5 contain days with furnace shutdowns and major problems with production whilst cluster 4 includes under average production with high silicon content. These three clusters are considered to be unnecessary for the further analysis. The poor process days and all output variables are then removed from the dataset to simplify the further analysis.

The remaining data was then analysed for relationships with hot metal silicon using a classification technique from Weka called decision stump. This technique calculates the most significant variable affecting hot metal silicon. That variable is then removed from the dataset and the calculations are repeated to get the second most significant variable. The reason for using the decision stump technique instead of a decision tree for example, is because it is known that some of the variables are dependent and that would skew such an analysis. In table 3 the results from decision stump can be seen with the possible reasons for its significance.
Table 3. Decision stump list over the most significant variables affecting HMSi in the post PCI dataset.

<table>
<thead>
<tr>
<th>Significant variables</th>
<th>Possible reasons</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Ore base tonne</td>
<td>Related to production</td>
</tr>
<tr>
<td>2. Quartz addition kg/tHM</td>
<td>SiO$_2$ activity</td>
</tr>
<tr>
<td>3. Limestone addition kg/tHM</td>
<td>Related to slag basicity</td>
</tr>
<tr>
<td>4. Coal A</td>
<td>Ash</td>
</tr>
<tr>
<td>5. Load SiO$_2$ kg/tHM</td>
<td>Source</td>
</tr>
<tr>
<td>6. Heat Flow Ratio</td>
<td>Cohesive zone height</td>
</tr>
<tr>
<td>7. AM Index Temperature wall</td>
<td>Radial liquid</td>
</tr>
<tr>
<td>8. Sinter SiO$_2$</td>
<td>Source</td>
</tr>
<tr>
<td>9. Coals other</td>
<td>Ash SiO$_2$ activity</td>
</tr>
</tbody>
</table>

- Ore base shows up high because of its relationship to the production rate which is an output not used in this analysis and can therefore be neglected.

- Quartz in its raw form SiO$_2$ is added to the furnace as a slag basicity controller when the basicity becomes higher than wanted.

- Limestone is also used as a slag basicity controller but additions are made when the basicity is lower than wanted.

- Coal A and Coal Others are often seen as long term variables and they probably show up because of a coincidence, when the silicon rises over a long period of time (years) and the coal properties are changed within that period.

- Total load of Silica shows up as significant which is expected.

- The heat flow ratio is a calculation of the inner furnace relationships and can be seen as a measure of cohesive zone height. A lower cohesive zone height is often considered to result in a lower silicon content of the hot metal caused from a shorter iron dripping height hence a shorter contact time between iron droplet and silica.

- The wall temperature is a result of cohesive zone height and shape.

- The amount of SiO$_2$ in the Sinter is also expected since it’s a source of silicon to the hot metal.

The relationships between some of the variables from the decision stump result and some output variables are below plotted to gain further insight.
Implementation and Results

Figure 15. Variable plot made in statistica.

In figure 15 Ore base and Production relates to each other which strengthens the assumption that ore base mostly shows up high in the list because it affects or is affected by the production, whilst the production itself correlates strongly with hot metal silicon content hence a higher production rate lowers the reaction time for silica adsorption into the hot metal dripping down from the cohesive zone. Quartz is strongly correlated to silicon and an increase in Quartz gives an increase in silicon content, but at the same time as slag basicity is correlated to both Quartz and silicon. The slag basicity (CaO/SiO$_2$) increases when more of the charged silica goes to the metal instead of the slag. When the slag basicity increases and becomes higher than expected, Quartz (SiO$_2$) is charged to the furnace with the intention of lowering the basicity in the slag. So if Quartz is related to the silicon, is that only because of the fact that Quartz is charged when more silica goes to the metal instead of the slag or does a significant amount of this Quartz end up in the hot metal? The importance of the relationship between Quartz and hot metal silicon can not be verified by daily data alone so a short term dataset containing two month time series was collected. The short term time series dataset contained two variables, Quartz charged and silicon in hot metal. To explain the earlier problem of what drives what, the Quartz charge was time shifted against the HMSi. The results from the time shifting can be observed in figure 16:
Figure 16. Correlations between HMSi and Quartz, and between Quartz and Quartz at different time lags.

Figure 16 shows the correlation between HMSi and Quartz at different time lags (labelled as “HM Si”). The best correlation occurs at a time lag of 9 hours, which tells that 9 hours after the charge of Quartz, HMSi and Quartz has a correlation peak. It takes about 9 hours for the Quartz charged into the furnace to descend through the furnace and reach the hot metal, transfer the silica into the hot metal as silicon, tapping of the metal and analysis of the metal chemistry. This analysis shows that the silicon content of the hot metal follows the amount of Quartz that reaches the bottom of the furnace where the transfer occurs. Hence an increase in Quartz reaching the furnace bottom leads to an increase in hot metal silicon content due to the fact that silica from Quartz transfers to the hot metal as silicon. This can be explained by the high activity of SiO$_2$ in Quartz about 1.0 comparing to the silica activity in coke only 0.5. This proves that Quartz drives the silicon, not the other way around. To disprove that silicon drives Quartz charge we look at the figure again. If Quartz comes up as significant only because it is charged when the silicon increases then this should be seen in the correlation graph as large peak at -1 hour which is the time it takes for the charge of Quartz after the basicity of the tapped slag has been analysed. Only a very small peak can be seen in the graph at -1 hour which only shows that this plays a small role compared to the affect of Quartz charge to hot metal silicon content.

In the Quartz curve, the Quartz is plotted against time lagged Quartz. As can be seen in the figure, the curves follow each other below minus 20 and above plus 20 hours. This explains why the HMSi curve is relatively highly correlated even at minus 48 and plus 48 hours.
3.3 Difference dataset

From the start of the campaign in 1991 to the end of the dataset, differences have arisen in the furnace that cannot be measured and are therefore not shown in the dataset. Hence the full dataset suffers from these long term variabilities, e.g. changes in hearth profile due to erosion. Also, longer term variables such as coke compositions become a problem in the analysis because they can mask or dominate the influence of short term operational variables such as blast humidity, etc. Therefore a new dataset is calculated from the original dataset. The instances in the new dataset consist of day to day differences where the new day 1 is original day 2 minus original day 1. The reason for this is to eliminate the influence of unmeasured long term variability and remove the effect of known longer term variables. Both input and output variables were used in this analysis. Next step was clustering on the same operation parameters as before to remove instances with extreme day to day differences and only keep the normal operation days. Periods where the production is very low does not show big differences from day to day and was therefore filtered out with all days below 6000 tonnes in production. The clusters given from the analysis are also used on the difference dataset. Then the difference dataset and the original dataset are compared with decision stump technique.

Table 4. Decision stump list for both the original and the difference dataset.

<table>
<thead>
<tr>
<th>Original dataset</th>
<th>Difference dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Ti/TiO₂</td>
<td>1. Hot metal Titanium, Ti %</td>
</tr>
<tr>
<td>2. Carbon saturation</td>
<td>2. Ti/TiO₂</td>
</tr>
<tr>
<td>3. Pellet A kg/tHM</td>
<td>3. Hot metal temperature</td>
</tr>
<tr>
<td>4. Pellet A %</td>
<td>4. Slag Titania, TiO₂ %</td>
</tr>
<tr>
<td>5. Stave B3 Temperature</td>
<td>5. Carbon solution loss %</td>
</tr>
<tr>
<td>6. Stave S1 Temperature</td>
<td>6. Mn/MnO</td>
</tr>
<tr>
<td>7. Coke Coal &lt;1mm</td>
<td>7. Slag Al₂O₃ %</td>
</tr>
<tr>
<td>9. Coal B</td>
<td>9. Slag Magnesia, MnO %</td>
</tr>
<tr>
<td>10. Coals non fluid %</td>
<td>10. Amount of ferrous kg/tHM</td>
</tr>
<tr>
<td>11. THT load</td>
<td>11. Standard deviation HMSi</td>
</tr>
<tr>
<td>12. Coke Coal &lt;3mm</td>
<td>12. Hot metal sulphur %</td>
</tr>
<tr>
<td>15. Hot metal Titanium, Ti %</td>
<td>15. Liquid volume</td>
</tr>
<tr>
<td>16. Coals other %</td>
<td>16. Pellet B</td>
</tr>
</tbody>
</table>

In the original dataset, silicon is more correlated with longer term variables such as pellet types, coke compositions and wall temperatures. The difference dataset has only a few burden type and composition variables but the interesting result is that the importance of temperature is revealed. Then the difference dataset was cleaned through removing the outputs and possible interrelated variables such as liquid volume, slag rate, ferrous rate and slag Al₂O₃. Carbon saturation shows up because it is calculated from HMSi according to the formulas:
%\text{C}_{\text{sat}}(\text{theory}) = 1.3 + 2.57 \cdot 10^{-3} \cdot T - 0.31[%\text{Si}] - 0.33[%\text{P}] - 0.4[%\text{S}] + 0.028[%\text{Mn}] \quad (3.1)

%\text{C}_{\text{sat}}(\text{data}) = C + 2.68 \cdot 10^{-3} \cdot T - 0.31[%\text{Si}] \quad (3.2)

Correlations between ferrous, slag rate, liquid volume and slag Al$_2$O$_3$ are visualized in Figure 17 and explain their position in the significance list. They are all interrelated which skews the analysis and increases their correlation. This problem is solved by removing these variables and repeating the analysis. The significance of the manganese and titania variables can be questioned because of their known relationships with silicon. If these variables really are important then the total load of TiO$_2$ and MnO should show up if the other manganese and titania variables are removed. The results after removing the variables and repeating the decision stump are illustrated below in table 5.

Figure 17. Variable plot made in Statistica showing correlations between important variables.
Implementation and Results

**Table 5. Cleaned Decision stump list.**

<table>
<thead>
<tr>
<th>1. Hot metal temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Carbon solution loss</td>
</tr>
<tr>
<td>3. PC Sulphur</td>
</tr>
<tr>
<td>4. Pellet B</td>
</tr>
<tr>
<td>5. Temperature top</td>
</tr>
<tr>
<td>6. Permeability factor K</td>
</tr>
</tbody>
</table>

Manganese and titania do not show up in the table as total load of TiO$_2$ and MnO which removes the possible importance of Mn and Ti. As can be seen in the table, hot metal temperature is the most significant variable affecting hot metal silicon in the whole dataset. This is expected because of the knowledge that temperature drives thermodynamics and speeds kinetics of silica transfer. Temperature is also known to be a driver of many other variables which gives a lot of unwanted dependencies in the analysis. To deal with this problem hot metal silicon is then corrected with hot metal temperature. A linear regression is calculated predicting hot metal silicon with temperature.

Linear regression:

$\text{PredHMSi} = 0.0043 \cdot \text{HMT} + C \quad (3.3)$

$\text{CorrHMSi} = \text{HMSi} - \text{PredHMSi} = \text{HMSi} - 0.0043 \cdot \text{HMT} \quad (3.4)$

The next step was to look for the most significant variable affecting corrected HMSi and compare to previous variables affecting HMSi. The results are visualized in table 6:

**Table 6. Decision stump list over the most significant variables affecting both HMSi and corrected HMSi.**

<table>
<thead>
<tr>
<th>HMSi</th>
<th>Corrected HMSi</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hot metal temperature</td>
<td>Carbon solution loss</td>
</tr>
<tr>
<td>Carbon solution loss</td>
<td>Temperature top</td>
</tr>
<tr>
<td>PC sulphur</td>
<td>Production</td>
</tr>
<tr>
<td>Pellet B</td>
<td>Number of charges</td>
</tr>
<tr>
<td>Temperature top</td>
<td>Efficiency</td>
</tr>
<tr>
<td>Permeability factor K</td>
<td>Heat flow ratio tuyere</td>
</tr>
</tbody>
</table>

The tuyere heat flow ratio is the heat required by the condensed phases (Iron, slag and coke) to the heat available from the gas stream (bosh gas volume, bosh gas composition).

$$\text{Gamma Tuyere} = \frac{\text{Coke rate} \cdot \text{Cp}_{\text{coke}} + \text{Slag rate} \cdot \text{Cp}_{\text{slag}} + \text{Metal rate} \cdot \text{Cp}_{\text{metal}}}{\text{Bosh gas rate} \cdot \text{Cp}_{\text{bosh gas}}}$$

The heat flow ratio is a relatively good measure of the cohesive zone height and is therefore important to the silicon transfer.
Implementation and Results

The most interesting difference between the two lists is that the significance of sulphur in hot metal vanishes when correcting with hot metal temperature. The correlation of sulphur with corrected silicon and silicon is compared in figure 17:

\[ y = -1.411x - 0.005 \]
\[ R^2 = 0.004 \]
\[ y = -9.365x - 0.007 \]
\[ R^2 = 0.130 \]

Figure 18. The comparison between the correlation of HMSi and hot metal Sulphur and corrected HMSi and hot metal Sulphur.

The blue plot is the correlation between hot metal sulphur and hot metal silicon whilst the red plot is the correlation between HM S and corrected HMSi. The correlation disappears completely after correcting HMSi with hot metal temperature. This tells that silicon and sulphur only are related in terms of temperature.

Two theoretical treatments were used to predict HM S one from Tsuchiya and one from Cripps Clark. The results are visualized in figure 19:
Figure 19. The theoretical predictions of hot metal sulphur.

Although there could be a correlation between sulphur and silicon hidden by more significant variables, the theoretical impact of temperature and silicon (via sulphur activity in metal) are sufficient to fully explain the variation in hot metal sulphur.
4 Discussion

The dataset contains a lot of variables over a long period of time, which makes it hard to analyse because of the shear volume of data that needs to be processed in the program, since the number of calculations that has to be made by the algorithms increases fast with increasing number of instances. Therefore most of the analysis is made on daily data with each instance containing a mean value for each day. This method gives a reasonable process time but still makes the analysis proper for finding correlations between variables. You could claim that the results from an analysis with that kind of data would be inadequate because in the analysis where input variables are used, an input mean value is compared to a mean value of hot metal silicon content which does not exactly correspond in time.

In the post PCI dataset Quartz was believed to be affecting hot metal silicon but because of the uncertainty of mean values this result could not be verified. This problem was solved by scaling the dataset down to two months time series, where each instance contained the precise value of the input variable and hot metal silicon value at a specific time. There were also uncertainties whether charge of Quartz in the furnace really affected HMSi or if the Quartz charge only showed up because of its relationship with slag basicity. When the slag basicity CaO/SiO\(_2\) increases and becomes higher than wanted, Quartz (SiO\(_2\)) is charged in to the furnace to lower the basicity. The slag basicity often increases because at that time more silica goes to the metal as silicon than in to the slag as SiO\(_2\) and hence the relationship between hot metal silicon and both Quartz and slag basicity. This problem with interrelations was also solved with the introduction of time series and using time lagged variables to explain what drives what.

If the silicon would be the driver of Quartz charge then a peak would be expected at about minus one hour in figure 16. The reason for this is that the time it takes for the slag basicity to be analysed and for the operators to respond with a Quartz charge is somewhere about one hour. As can be seen in the figure only a small peak is present at minus one hour. This shows that silicon in some extent affects the charge of Quartz but far from fully. Instead the largest peak is visible at plus nine hours in the figure, so nine hours after a charge of Quartz, silicon content of the hot metal shows the greatest response to the amount of Quartz charged in to the furnace. The reason for this response is that instead of the SiO\(_2\) in Quartz transferring to the slag as intended, it transfers in to the hot metal and gives an increase in hot metal silicon. The nine hour delay is due to furnace processing time and hearth residence time. You could speculate the cause of this phenomenon by comparing the theory of SiO gas generation from different sources where coke ash and PC ash is the largest generator of SiO gas much because of the high activity of silica in these materials. The activity of silica in coke ash and PC ash is about 0.5 comparing to the activity of Quartz which is close to 1.0 since quartz almost exclusively contains raw SiO\(_2\). Further we can compare with the generation of SiO gas from slag where the silica is bound to alumina and calcium oxide with low activity of silica which results in a generation of much less extent than from Quartz. There is also a possibility that the SiO\(_2\) from Quartz reacts with the carbon in the metal and Si transfers into the metal. This reaction is also affected by the activity of silica which makes this reaction important as well. In what extent these possible reactions takes place is hard to speculate but one thing is clear a conclusion can be drawn concerning the significance of Quartz charge. The Quartz charge
affects the silicon content in hot metal negatively hence an increase in Quartz increases the hot metal silicon content. Therefore it would be better to achieve furnace basicity requirements via the sinter or pellet composition, rather than using raw fluxes.

The preliminary dataset stretches over a period of 14 years. In the analysis, instances from the first year would be compared to instances from the last year without taking into account that there are differences in the process operation between these in time separated instances. In the beginning of the campaign in 1991 the inner walls of the blast furnace were intact, and as the campaign proceeded more and more of the inside, (mainly in the lower zone of the furnace) eroded. There are also differences in the way of operating the furnace as well as variation in the types of raw materials charged in to the furnace. Many of these variables are hard to measure and are therefore missing in the dataset. This could distort the results and block out the importance of short term variables. The problem was solved by introducing a second dataset with variable values calculated from the original dataset. Each instance of the difference dataset contained a difference between two after each other following days calculated from the original dataset. The results show that the introduction of day to day differences eliminates the long term variability and helps to reveal the importance of short term variables such as hot metal temperature and hot metal chemistries. The problem with long term variability was also solved in the pre and post PCI dataset, this time by splitting up the dataset so that the post dataset contained instances stretching over four years where raw materials and inner wall erosions were relatively similar. During these four years furnace parameters vary to some extent but the variation is assumed to be of less importance. The decision stump analysis of the difference dataset produced results that can be seen in table 4 where they are compared to decision stumps from the original dataset. These results illustrate that no long term variables show up in the results, instead short term variables as hot metal temperature, slag chemistries and hot metal chemistries shows up as the most significant variables affecting hot metal silicon. This proves that the difference dataset ignores the long term variables, hence helps revealing the affect of short term variables.

One of the most important variables affecting hot metal silicon is temperature. In the dataset different temperature variables are used: hot metal temperature, top temperature, wall temperatures and blast and tuyere temperatures. In the analysis hot metal temperature showed up as highly significant while the other temperatures didn’t. This could be a hint that absorption of silicon in to the metal controls the hot metal silicon content more than generation of SiO gas. If the generation of SiO gas is the reaction that controls silicon then another temperature variable such as raceway adiabatic flame temperature should show up as significant. This because of the fact that generation of silica occurs in the raceway. Hot metal temperature is known to affect a lot of reactions in the blast furnace, therefore when finding hot metal temperature as important the question if interrelationships between temperature and all the other variables exist must be asked. The reason for this is that if a variable is almost fully controlled by temperature while temperature is affecting hot metal silicon then the variable will show up as significant to hot metal silicon though it is not. This question was answered with correcting hot metal silicon with temperature. First hot metal silicon was predicted with hot metal temperature so that an equation was received for the part of hot metal silicon that can be explained by hot metal temperature. Then the value from the equation was subtracted from the real hot metal silicon value, hence the resulting value that explains the part of hot metal silicon
that can not be explained by hot metal temperature. The next step was to find the variables affecting corrected hot metal silicon. The results from this analysis show an interesting change in the most significant list, hot metal sulphur has vanished and is no longer important. Hot metal sulphur is only affecting the part of silicon that can be explained by the temperature which gives conclusion that hot metal sulphur does not affect hot metal silicon instead hot metal sulphur is interrelated with hot metal temperature.
Conclusions

5 Conclusions

In this masters project the relations between input variables and silicon’s behaviour in Blast furnace nr 5 at Port Kembla steel works has been studied. Real process data from the furnace has been collected and analysed with data mining techniques. From the work, the following conclusions could be drawn:

C1: The SiO\textsubscript{2} in Quartz, loaded as a slag basicity controller, transfers into the hot metal instead of into the slag as intended.

C2: It is better to achieve furnace basicity requirements via the sinter or pellet composition than using raw fluxes.

C3: Analysis of day-to-day differences proved valuable in revealing the effect of short term variables by eliminating the long term variability.

C4: Hot metal sulphur shows no correlation with hot metal silicon after removing the impact of hot metal temperature.

C5: A data mining framework brought valuable new insight into hot metal silicon control in the blast furnace and further effort in this area is strongly recommended.
References


17. The Iron and Steel Institute of Japan. "Blast furnace phenomena and modeling”, Committee on reaction within Blast furnace, Joint society on iron and steel Basic research, (1982).


