OULU 2021
C 777

ACTA UNIVERSITATIS OULUENSIS

Tero Vuolio

MODEL-BASED IDENTIFICATION AND ANALYSIS OF HOT METAL DESULPHURISATION

UNIVERSITY OF OULU GRADUATE SCHOOL;
UNIVERSITY OF OULU,
FACULTY OF TECHNOLOGY

ACTA UNIVERSITATIS OULUENSIS
C Technica 777

TERO VUOLIO

MODEL-BASED IDENTIFICATION AND ANALYSIS OF HOT METAL DESULPHURISATION

Academic dissertation to be presented with the assent of the Doctoral Training Committee of Technology and Natural Sciences of the University of Oulu for public defence in the OP auditorium (L10), Linnanmaa, on 29 January 2021, at 12 noon

UNIVERSITY OF OULU, OULU 2021

Acta Univ. Oul. C 777, 2021

Supervised by
Professor Timo Fabritius
Docent Ville-Valtteri Visuri
Doctor Timo Paananen

Reviewed by
Professor Nirupam Chakrabortti
Professor Rob Boom

Opponent
Docent Mikko Helle

ISBN 978-952-62-2835-8 (Paperback)
ISBN 978-952-62-2836-5 (PDF)

ISSN 0355-3213 (Printed)
ISSN 1796-2226 (Online)

Cover Design
Raimo Ahonen

PUNAMUSTA
TAMPERE 2021

Vuolio, Tero, Model-based identification and analysis of hot metal desulphurisation.
University of Oulu Graduate School; University of Oulu, Faculty of Technology
Acta Univ. Oul. C 777, 2021
University of Oulu, P.O. Box 8000, FI-90014 University of Oulu, Finland

Abstract

Sulphur is considered one of the main impurities in steel. Hot metal desulphurisation serves as the main unit process for sulphur removal in the production of steel. The main objective of this thesis is to identify the relevant phenomena and attributes needed to construct a mathematical model suitable for online use. The study also includes a detailed literature review on the modelling of hot metal desulphurisation, which considers a categorisation of the existing models for the process, but also outlines the main uncertainties in the process that may decrease the prediction performance of the existing models.

In this study, model-based process identification techniques are studied. More specifically, the objective is to study different techniques, both to explain the variance and to predict the end content of sulphur in the process. To do this, a modelling framework exploiting data-driven and mechanistic modelling techniques is proposed. The model identification procedure is divided into variable construction, variable selection, model structure selection, and model parameter identification steps. The model identification procedure considers both manual and automatic model identification techniques. The thesis focuses on grey box and black box model structures. In automatic model identification, the focus is on evolutionary search strategies, particularly genetic algorithms.

The results of this study show that in the case of lime-based hot metal desulphurisation, the major factors inducing variance in the end content of sulphur are related to the properties of the reagent, i.e. to the rate of the transitory contact reaction. If the particle size distribution is known a priori or can be assumed constant, the prediction accuracy of the models can be improved considerably. In addition, the parameterisation of the reaction models improves the prediction performance. It was also found that physically meaningful descriptions for the uncertain phenomena may help to constrain the search of parameters. In addition, in-depth phenomena-based analysis and automatic model identification strategies may assist in model selection.

Keywords: data-driven methods, genetic algorithm, hot metal desulphurisation, mathematical modelling, model selection

Vuolio, Tero, Raakaraudan rikinpoiston mallipohjainen analyysi.
Oulun yliopiston tutkijakoulu; Oulun yliopisto, Teknillinen tiedekunta
Acta Univ. Oul. C 777, 2021
Oulun yliopisto, PL 8000, 90014 Oulun yliopisto

Tiivistelmä

Rikki on keskeisimpiä raakarautaan liuenneita epäpuhtauksia. Hiiliteräksen valmistusketjussa raakaraudan rikinpoisto on prosessi, jossa rikki pääasiallisesti poistetaan. Tämän työn tavoitteena on tunnistaa prosessin kannalta merkityksellisiä ilmiöitä ja tekijöitä, joita tarvitaan on-line käyttöön soveltuvien matemaattisten mallien luomiseen. Työ sisältää myös yksityiskohtaisen kirjallisuusselvityksen, jonka tavoitteena on kategorisoida kirjallisuudessa esitetyt mallit, mutta myös tarkastella mallien suorituskykyyn liittyviä epävarmuustekijöitä prosessin näkökulmasta.

Menetelmällisesti työ perustuu prosessin mallipohjaiseen analyysiin ja mallien valintaan. Tarkempana tavoitteena on tarkastella systemaattisia tapoja selittää prosessin loppurikkipitoisuuden vaihtelua, mutta myös ennustaa loppupitoisuutta luotettavasti saatavilla olevan aineiston perusteella. Tätä varten tehtiin mallinnuskehys, joka hyödyntää sekä täysin datapohjaisia, mutta myös mekanistisiin ilmiöihin pohjautuvia dataa hyödyntäviä malleja. Mallin vallinta jaotellaan ennustemuuttujien rakenteluun, ennustemuuttujien valintaan, mallin rakenteen valintaan sekä malliparametrien estimointiin. Valinnassa käytetään sekä automaattisia, että asiantuntijatietoon perustuvia tekniikoita. Mallit ovat rakenteellisesti joko harmaa- tai mustalaatikko filosofiaan pohjautuvia. Automaattisessa mallien valinnassa tarkastellaan eniten erityisesti geneettisten algoritmien toimintaa.

Tämän työn tulokset näyttävät, että reagenssin ominaisuuksilla kuten partikkelikokojakaumalla sekä kaasua injektoivien lisäaineiden määrällä on vaikutus rikkipitoisuuden vaihteluun erityisesti partikkelien ja rautasulan välillä tapahtuvan reaktion nopeuden näkökulmasta. Yleisesti mallien suorituskykyä voidaan parantaa, kun partikkelikokojakauman vaihtelu tunnetaan, tai sen voidaan otaksua olevan vakio. Malliparametrien optimointia helpottavat fysikaaliset reunaehdot ja prosessituntemus. Automaattisten mallin valintatekniikoiden käyttäminen voi auttaa mallintajaa tarkoituksenmukaisen mallin valinnassa, mutta asiantuntijatiedon merkitystä mallinnuksessa ei voi kuitenkaan korostaa liikaa.

Asiasanat: datapohjaiset menetelmät, geneettiset algoritmit, mallin valinta, matemaattinen mallinnus, raakaraudan rikinpoisto

To my family.

Acknowledgements

This work was carried out at the Process Metallurgy Research Unit, University of Oulu from November 2017 to December 2020. The research was conducted within the Flexible and Adaptive Operations in Metal Production (FLEX) and Symbiosis of Metal Production and Nature (SYMMET) funded by Business Finland. In addition, the research has been supported by personal grants awarded by The Technology Industries of Finland Centennial Foundation, Tauno Tönning Foundation, Walter Ahlström Foundation and Finnish Foundation for Technology Promotion.

The support of the supervisors of the work is greatly appreciated. First, I would like to thank my principal supervisor Professor Timo Fabritius for enabling me to carry out the research work at the research unit. The acknowledgement for the interesting research topic goes to Doctor Timo Paananen, who also hired me at the first place as a master's thesis worker at SSAB Europe Oy, and by that provided valuable practical knowledge of the topic. The support of Docent Ville-Valtteri Visuri has been invaluable. Docent Visuri has transferred me a vast amount of knowledge in various subject areas, but especially in modelling. Without him, this work would certainly not be nearly as good as it is now.

As for the other co-authors, I would like to express my gratitude to Doctor Aki Sorsa, who was willingly helping me throughout the work, and provided support especially in the model selection studies. Mr. Sakari Tuomikoski from SSAB Europe Raahe provided his expertise and aid in arranging and executing the industrial trials at the hot metal desulfurization site to collect the validation data. In addition, Mr. Seppo Ollila provided some new data on the secondary desulfurization, which enabled the finetuning of the topic and the finalization of the work.

I would also like to thank my family and other relatives for their love, support, and encouragement. Finally, I would like to thank my beloved Sara and our hilarious dachshund Hertta for their love and support, and of course for forming our goofy trio.

11 December 2020, Oulu Tero Vuolio

List of abbreviations and symbols

Abbreviations

ANN	Artificial neural network
BCGA	Binary-coded genetic algorithm
CV	Cross-validation
GA	Genetic algorithm
LMO-CV	Leave-multiple-out cross-validation
LOO-CV	Leave-one-out-cross-validation
MAE	Mean absolute error
MLR	Multivariable linear regression
MSE	Mean squared error
PSD	Particle size distribution
RCGA	Real-coded genetic algorithm
RMSE	Root mean-squared error
SSE	Sum of squared error
HMD	Hot metal desulphurisation

Symbols

A	Area [m^2]
a	Activity of a species [-]
b_i	Regression coefficient for a variable i [-]
C_S'	Sulphide capacity based on slag-metal equilibrium [-]
C_S	Sulphide capacity based on slag-gas equilibrium [-]
d_{80}	Diameter of a particle corresponding to 80 V-% in a cumulative particle size distribution [m]
d_{ka}	Mean particle diameter according to surface area approximation and Sh = 2. [m]
d_{32}	Sauter mean diameter of a particle size distribution [m]
D	Diffusion coefficient [m^2/s]
K	Equilibrium constant [-]
k_i	Time constant for a reaction i [1/s]
L_S	Sulphur partition ratio [-]
M	Molar mass [kg/mol]

N	Normal distribution [-]
N	Number of cross-validation splits [-]
n_{eff}	Molar efficiency of a reagent [%]
n_{pop}	Number of individuals [-]
$\dot{m}$	Reagent feed rate [kg/s]
p_C	Crossover probability [-]
p_M	Mutation probability [-]
Q	Volumetric gas flowrate [m^3/s]
R	Universal gas constant [J/(mol·K)]
R_i	Reaction rate for mechanism i [%/s]
R^2	Coefficient of determination [-]
r	Uniformly distributed random number [-]
t	Time [s]
t_{res}	Residence time [s]
V	Volume [m^3]
x_i	Input variable i [-]
y	Output variable [-]
$\hat{y}$	Predicted output variable [-]
w	Mass fraction [-]
X	Data matrix [-]
α	Confidence level [-]
β	Mass transfer coefficient [m/s]
ρ	Density [kg/m^3]
γ	Individual in the population of a genetic algorithm [-]
[ ]	Species dissolved in hot metal
( )	Species in slag phase
{ }	Species in gas phase
< >	Solid species

List of original publications

This thesis is based on the following publications, which are referred to throughout the text by their Roman numerals (I–V):

I Vuolio, T., Visuri, V.-V., Tuomikoski, S., Paananen, T., & Fabritius, T. (2018). Data-driven mathematical modeling of the effect of particle size distribution on the transitory reaction kinetics of hot metal desulfurization. *Metallurgical and Materials Transactions B*, *49*(5), 2692–2708. https://doi.org/10.1007/s11663-018-1318-4

II Vuolio, T., Visuri, V.-V., Paananen, T., & Fabritius, T. (2019). Identification of rate, extent and mechanisms of hot metal resulfurization with CaO-SiO_2-Na_2O slag systems. *Metallurgical and Materials Transactions B*, *50*(4), 1791–1807. https://doi.org/10.1007/s11663-019-01600-5

III Vuolio, T., Visuri, V.-V., Sorsa, A., Paananen, T., & Fabritius, T. (2019). Genetic Algorithm-based variable selection in prediction of hot metal desulfurization kinetics. *Steel Research International*, *90*(8), 1900090. https://doi.org/10.1002/srin.201900090

IV Vuolio, T., Visuri, V.-V., Sorsa, A., Ollila, S., & Fabritius, T. (2020). Application of a genetic algorithm based model selection algorithm for identification of carbide-based hot metal desulfurization. *Applied Soft Computing Journal*, *92*, 106330. https://doi.org/10.1016/j.asoc.2020.106330

V Visuri, V.-V., Vuolio, T., Haas, T., & Fabritius, T. (2020). A review of modeling hot metal desulfurization. *Steel Research International*, *91*(4), 1900454. https://doi.org/10.1002/srin.201900454

Contribution of the author

Publications I and III

The author of this thesis prepared the research plan, designed and conducted the experiments with the plant personnel, designed and implemented the studied algorithms, analyzed the results with the co-authors, and wrote the draft version of the manuscript.

Publication II

The author of this thesis prepared the research plan with the co-authors, designed and conducted the experiments with the laboratory staff, analysed the results of the experiments and previous studies with the co-authors, designed and implemented the mathematical models, implemented the model-training strategies, and wrote the draft version of the manuscript.

Publication IV

The author of the thesis designed and implemented the algorithm presented in the article, conducted the literature part with the co-authors, designed and carried out the experiments, and wrote the draft version of the manuscript.

Publication V

The author of the thesis analysed the literature with the co-authors and wrote the draft version of the manuscript with the first author.

Table of contents

1 Introduction

1.1 Hot metal desulphurisation in steel production

In steelmaking, sulphur is considered one of the main impurities in hot metal. The sulphur originates from the raw materials used in blast furnace ironmaking. The existence of sulphur in the final steel product causes brittleness, decreasing its weldability and corrosion resistance. The sulphur content after the blast furnace is typically around 0.030–0.045 wt-%. The appropriate final sulphur content depends on the final steel product, but it can be as low as 0.001 wt-% (10 ppm) (Schrama et al. 2017). In steel production, the sulphur is removed either in secondary metallurgical processes such as hot metal desulphurisation (HMD) and ladle furnace (LF), or in primary metallurgical processes such as blast furnace (BF) or Basic Oxygen Furnace (BOF) (Schrama et al. 2017). Figure 1 illustrates the hot metal desulphurisation and its role in ironmaking at SSAB Europe's Raahe steelworks.

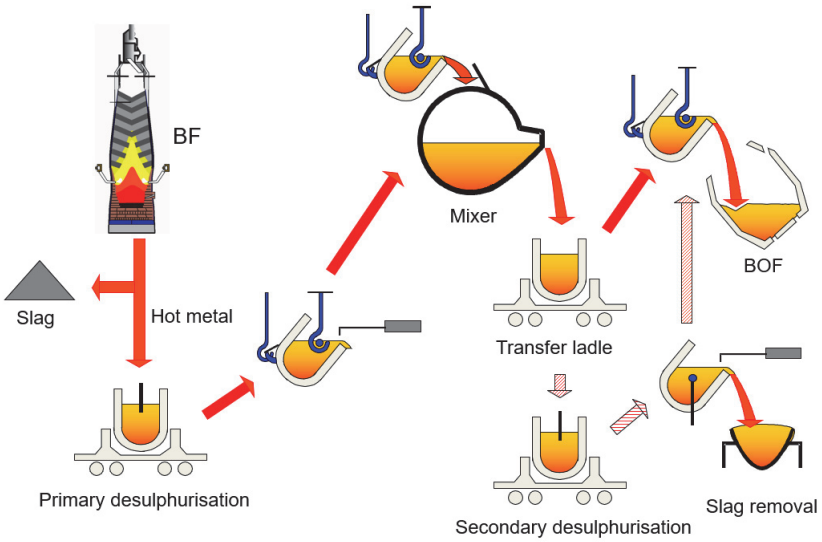

Fig. 1. Hot metal desulphurisation in ironmaking (Reprinted, with permission, from SSAB Europe [Raahe steelworks] 2019 © SSAB).

Hot metal desulphurisation serves as a unit process for external sulphur removal prior to decarburisation in the basic oxygen furnace. Desulphurisation is more

efficient prior to converting the hot metal to steel, because it demands reductive conditions, which prevail in melts with high silicon or carbon content. In this context, external sulphur removal entails desulphurisation outside the blast furnace process (Oeters 1994).

There are three primary techniques for conducting hot metal desulphurisation: the lance injection technique; the Kanbara process; and the permanent contact reaction with bottom stirring (Schrama et al. 2017). Of these, this study most considers the full-scale processes conducted using lance injection. The Kanbara process is a common technique for hot metal desulphurisation in east Asia, mainly in Japan and China (Schrama et al. 2017). In European steelmaking, lance injection is considered a standard practice (Schrama et al. 2017). Less attention is therefore paid to the mechanisms of the Kanbara process and bottom stirring in this thesis, although Publication V also reviews the models available for Kanbara. For a comprehensive review of other hot metal desulphurisation techniques and reagents, the review article by Schrama et al. (2017) provides a detailed and accessible summary.

In lance injection-based hot metal desulphurisation, sulphur is extracted from the metal phase and transferred to the slag phase by injecting a fine-grade desulphurisation reagent, with the aid of a carrier gas via an immersed lance. A typical gas used in the injection is nitrogen (N_2), but the use of other gases such as natural gas is also possible (Oeters 1994). The desulphurisation reagent can be composed of a single chemical compound, or it can be a mixture of several compounds. Suitable chemical compounds include calcium oxide (CaO), calcium carbide (CaC_2), calcium carbonate ($CaCO_3$), sodium carbonate (Na_2CO_3), magnesium (Mg), and zinc oxide (ZnO), as well as their mixtures (Schrama et al. 2017). Of these, this study considers CaO, $CaCO_3$, Na_2CO_3, and CaC_2.

The main benefits of lance injection are the easy controllability of the reagent injection, and that it allows the co-injection of various compounds that enhance the reaction conditions (Irons 1988). Compared to the desulphurisation conducted with top slag, the reaction mechanisms in lance injection are considerably faster due to the substantial interfacial area available for mass transfer. In addition, the injection gas tends to keep the hot metal well-mixed (Irons 1988; Irons 1989). However, the utilisation ratio of the hot metal desulphurisation reagent is relatively low, mainly due to the short residence time of particles injected into the melt (Oeters 1994).

The reagent can be injected with or without gas-forming compounds that may form substantial amounts of gas when decomposing due to the heat prevailing in the hot metal. From the previously mentioned, the $CaCO_3$ is nowadays injected as

a gas-forming auxiliary compound. The use of auxiliary compounds with the primary reagent has been proposed to promote direct reagent-metal contact in the three-phase system, i.e. in lance injection (Irons 1989), but also to enhance the scattering of the agglomerated reagent particles in the two-phase system (Lindström, Nortier, & Sichen 2014). However, the effect of the gas formed in decomposition was found to have a negligible effect on the stirring of the metal bath (Irons 1989).

1.2 Outline and contribution of the thesis

In this study, the hot metal desulphurisation process and related phenomena are investigated with experimental and mathematical modelling. The emphasis is on the data-driven models, which are briefly categorised in Section 3 and in more detail in Publication V. The research questions to be answered in this thesis are as follows:

1. What are the main attributes that affect the efficiency of hot metal desulphurisation?
2. What is the effect of reagent particle size distribution on the rate and efficiency of lime-based desulphurisation?
3. What types of model might be considered applicable in explaining and predicting the behaviour of the process in online use?

With reference to the research questions, the objective of this thesis is twofold. The thesis provides a discussion of the relevant phenomena that need to be considered in the construction of prediction models for the process under study. The emphasis is on building parameterised models suitable for online use, but also for explanatory analysis of the contributing factors in the system. The guidelines proposed in the thesis and in the publications can therefore also be applied to systematic process development, along with the use of models in process control and optimisation. However, the thesis also introduces some automatic model selection techniques for generic model structures, thus providing alternatives for the use of domain knowledge in model selection.

It should be noted that Section 3 summarises some fundamental considerations that are not reported in Publications I–V, but that are used as a basis for the construction of the model selection strategies and the models presented in the publications, and in the Results and discussion section. More specifically, the articles focus on methods and literature surveys, whereas this thesis opens the

background and motivation for the studies. The derivation of the models yet to be included in any of the listed publications is described more exhaustively in the Results and discussion section, in Sections 5.3, 5.4., 5.5., and 5.7. The meta-analysis of the sulphide capacity models and the collected sulphide capacity data has been discussed in the case of $CaO\text{-}SiO_2\text{-}Na_2O$ in Publication II and is extended in this thesis in Section 5.7. The results and discussion section also covers discussion and smaller studies beyond the scope of Publications I–V, which they therefore do not report. The objectives and the outline of the publications are presented in Table 1.

Table 1. Objectives and outline of the thesis.

Publication	Objectives and contents
Publication I	The article aims to plug the literature gap on the effect of particle size distribution on the rate of hot metal desulphurisation by exploiting explanatory and predictive modelling.
Publication II	The primary objective of this study is to systematically investigate the industrially relevant problem of re-sulphurisation. The re-sulphurisation kinetics and mechanisms are evaluated through systematic explanatory modelling.
Publication III	As in Publication I, the primary objective is to select a prediction model for primary hot metal desulphurisation. Here, the proposed model structure is generic, and the proposed selection strategy employs the genetic algorithm.
Publication IV	The objective of this study is to extend prediction model identification studies to neural network models. The models are employed to predict carbide-based hot metal desulphurisation.
Publication V	The objective of this paper is to provide a comprehensive review of the modelling of hot metal desulphurisation and discuss the future needs, outlines, and trends in the field.

2 Modelling of hot metal desulphurisation

The general desulphurisation reaction is described as a reaction in which the oxygen acts as an electron donor for sulphur to form a negatively charged anion (Oeters 1994):

$$[\mathrm{S}] + (\mathrm{O}^{2-}) \rightleftarrows (\mathrm{S}^{2-}) + [\mathrm{O}]. \quad (1)$$

The main reaction is considered to occur via three different mechanisms, which are as follows (Pal & Patil 1986; Chiang et al. 1990; Oeters 1994):

i. The transitory contact reaction, which occurs between the injected reagent particles and the metal phase, mainly during the ascending of particles towards the bath surface.
ii. The permanent contact reaction between the top slag and the metal phase.
iii. The reaction between the particles entrapped inside the bubbles formed due to the injection of the carrier gas.

In the literature, the overall desulphurisation reaction is often assumed to follow first-order reaction kinetics, and the main reaction control mechanism is attributed to mass transfer (Oeters 1994). Provided that the particles come into contact with the metal phase, the transitory contact reaction is often considered to be controlled by two sequentially occurring steps, which are the boundary layer and solid-state diffusion (Oeters 1994). The situation in which the particles remain inside the carrier gas bubbles, preventing them from coming into direct metal contact, is referred to here as contact control, as in Chiang et al. (1990). The general form for the rate of the individual mechanism is

$$R_j = -k_j([\mathrm{S}]_t - [\mathrm{S}]^*), \quad (2)$$

where R_j is the rate of reaction scheme *j,* being now either *i, ii,* or *iii*, k_j is the time constant for the reaction scheme *j*, $[\mathrm{S}]_t$ is the sulphur content at time instant *t*, and $[\mathrm{S}]^*$ is the equilibrium content of sulphur in the bulk or the sulphur content at the reaction interface for a reaction scheme *j*. As outlined in Publication V, a huge number of models is available for the hot metal desulphurisation process. In most, the change of sulphur in the metal bath is considered to follow first-order kinetics, as above. Referring to the existing models, the overall rate for desulphurisation can

be written as the sum of the above rates (Deo & Boom 1993; Seshadri, Da Silva, Da Silva, & von Krüger 1997)

$$R_{tot} = R_i + R_{ii} + R_{iii}, \quad (3)$$

where R_{tot} is the overall rate, R_i is the rate of transitory reaction, R_{ii} is the rate of permanent contact reaction, and R_{iii} is the rate between the metal phase and the particles entrapped in the bubbles.

2.1 Transitory reaction

As stated, the reaction that occurs between the reagent particles and the melt is referred to as the transitory reaction. The main desulphurisation reaction with CaO-based reagent particles with sulphur is (Oeters 1994)

$$< \mathrm{CaO} > + [\mathrm{S}] \rightarrow < \mathrm{CaS} > + [\mathrm{O}]. \quad (4)$$

With CaC_2 based reagents, the considered reaction is

$$< \mathrm{CaC_2} > + [\mathrm{S}] \rightarrow (\mathrm{CaS}) + 2[\mathrm{C}]. \quad (5)$$

In this study, these reactions are referred to as *transitory reactions*, because they are considered to occur during the ascent of the injected particles through the metal bath to the top slag. The transitory desulphurisation reaction can be generalised as an extraction process, in which the sulphur that is dissolved in the metal phase is bound to a suitable cation (Oeters 1994). In the above reactions, the Ca^{2+} acts as the binding cation. Of the rate expressions, the rate of transitory reaction is often considered the most significant (Chiang et al. 1990; Rastogi et al. 1994). However, there are some contradictions to this. For example, Hara et al. (1986) reported that if the $CaCO_3$ was used as the reagent, the contribution of the transitory reaction to the overall rate was only around 25%, whereas in the modelling study of Seshadri et al. (1997), the particle size specific contributions to the overall rate varied from 0 to 60%, depending on particle size. However, the model of Seshadri et al. (1997) was not validated with experimental data but was a good semi-quantitative illustration of the effects of different attributes on reaction rates.

As indicated in Publications I and V, there are many uncertainties regarding the description of the relevant phenomena in the models, which may result in the

expressions for the rate laws suffering from uncertainty. In addition, the datasets available for the identification of such full-scale mechanisms are often scarce or may not contain all the relevant data. These uncertainties complicate the evaluation of the relative contribution, because the rate of transitory reaction depends on the type and properties of the reagent, such as particle size distribution (Coudure & Irons 1994; Lindström & Sichen 2015). In addition, as the residence times and the fraction of particles in contact with the metal phase are not accurately known, there can be numerous solutions for relative contributions that satisfy the observed rate. It is assumed that the actual contribution depends on at least (Irons 1989; Chiang et al. 1990; Zhang & Irons 1994; Oeters 1994; Coudure & Irons 1994; Lindström et al. 2014; Lindström & Sichen 2015):

- the solid surface area of the particles in contact with the melt;
- the feed rate of the particles;
- the local oxygen activity;
- the rate of solid-state diffusion;
- the average residence time of the reagent particles;
- the mass transfer in the metal-reagent boundary layer.

Many of these factors are extremely difficult to quantify, not only in the full-scale process but at the laboratory scale. There are therefore several degrees of freedom in the system. Depending on the process and the drawn assumptions, the expression can be simplified by neglecting mechanism *iii*, because the contribution and the actual reaction mechanism between the metal phase and the particles entrapped in the carrier gas bubbles is difficult to estimate. This is mainly because the experiments provide only direct or indirect information concerning the *fraction of entrapped particles inside the carrier gas bubbles.* For example, Zhao and Irons (1994) used model-based analysis and the heat transfer correlations to determine this attribute (Zhao & Irons 1994).

As many studies outline, in both CaC_2 and CaO reagents, the effective solid surface area in contact with the metal is an important factor to be considered (Coudure & Irons 1994; Lindström & Sichen 2015), and the surface area to volume ratio of particles is theoretically proportional to particle size, because for ideal spheres, the following holds:

$$\left(\frac{A}{V}\right)_p = \frac{6}{d_p}, \tag{6}$$

where A is the surface area of a spherical particle, V is its volume, and d_p is its diameter. Due to the surface forces prevailing in the complex three-phase injection system, the *effective surface area of particles* is assumed to be smaller than the nominal surface area. This assumption is supported by Lee and Morita (2004), for example. They propose that for fine-grade particles (100 μm), the penetration velocities can be as high as 100 m/s. Some findings supporting this assumption concerning particle size distribution have been made by Coudure and Irons (1994), who find that the effect of mass transfer, determined by exploiting the equivalent diameter of a spherical particle in a diffusion-controlled process, is smaller than the theoretical effect. A regression analysis with logarithmic effects shows that the coefficient corresponding to the equivalent diameter for <CaC_2> reagent was $d_{ka}^{1.31}$, whereas the authors propose that the theoretical effect is squared, i.e. d_{ka}^{2} (Coudure & Irons 1994). The squared effect is based on the assumption that the characteristic length of diffusion is equal to the particle diameter (Coudure & Irons 1994). Publication I provides similar findings, as does this thesis.

Another factor that complicates the analysis of the reaction rate is the penetration behaviour and the residence times of reagent particles. There is no consensus in the literature concerning the values to use for these attributes. Nevertheless, it is almost impossible to directly quantify these attributes in industrial conditions by measuring, and only numerical estimates can therefore be made with dedicated computational fluid dynamics models. However, for online applications, these models provide only descriptive information on the boundary conditions or on the system attributes, because the governing equations for momentum, mass, and heat transfer are computationally very expensive to solve, and their use in process control is therefore impractical. This study relies on explanatory analysis with parameterised and generic data-driven models of quantifying the effect on CaO. However, this attribute is very difficult to determine precisely both in full-scale and at laboratory scale, and only semi-quantitative results are therefore available (Engh et al. 1979; Farias & Irons 1985; Farias & Irons 1986).

2.2 Permanent reaction

The permanent reaction has often been assumed to be mainly a function of specific surface area, the thermodynamic driving force, and the overall mass transfer rate in the boundary layer. The driving force of the slag-metal reaction is often taken as the concentration difference between the bulk-metal and metal-slag boundary

layers (Deng & Oeters 1990). The concentration in the boundary layer can be expressed by the sulphur partition ratio (Seshadri et al. 1997). The sulphur partition ratio can be defined between the slag and metal phases as follows

$$L_{\mathrm{S}} = \frac{(\%\mathrm{S})}{[\%\mathrm{S}]}. \tag{7}$$

As indicated in Publication II, the direction of the reaction can be expressed with the aid of the following inequalities of the sulphur partition ratio:

– $L_{S,t} < L_{S,eq}$ – is a sufficient condition for desulphurisation;

– $L_{S,t} > L_{S,eq}$ – is a sufficient condition for re-sulphurisation.

As discussed in Publication II, resulphurisation is considered a major problem in the industry. The rate of re-sulphurisation was found to be controlled by the magnitude of the thermodynamic driving force. Furthermore, the magnitude of the thermodynamic driving force can be associated with the sulphide capacity of the slag phase, which can be defined either on the basis of the slag-metal reaction presented above or on the basis of the slag-gas reaction of (Nzotta et al. 1998)

$$\frac{1}{2}\{\mathrm{S}_2\} + (\mathrm{O}^{2-}) \rightarrow (\mathrm{S}^{2-}) + \frac{1}{2}\{\mathrm{O}_2\}. \tag{8}$$

Based on the equilibrium constant of the above reaction, the expression for the sulphide capacity can be given as

$$C_{\mathrm{S}} = (\mathrm{S})\sqrt{\frac{p_{\mathrm{O}_2}}{p_{\mathrm{S}_2}}}, \tag{9}$$

where p_i is the partial pressure of a gaseous species i. Further, the equilibrium partition ratio of sulphur between the slag and metal phases can be expressed as (Oeters 1994)

$$\frac{(\%\mathrm{S})}{[\%\mathrm{S}]} = \frac{C_S'}{a_{[\mathrm{O}]}^{\mathrm{H}}} f_{\mathrm{S}}^{\mathrm{H}}, \tag{10}$$

where $C_S^{'}$ is the sulphide capacity on slag-metal reaction basis, and $a_{[O]}^{H}$ is the Henrian based activity of oxygen and f_S^H is the activity coefficient of sulphur in the metal. From this definition, it can be seen that (Oeters 1994)

$$C_S' = \frac{K a_{(O)}^{R}}{\gamma_{(S)}^{R}}, \tag{11}$$

where $\gamma_{(S)}^{R}$ is the Raoultian activity coefficient of sulphur in the slag. The sulphide capacity thus depends on the composition and temperature of the slag phase. As reported in Publication II, the functional dependency of the sulphide capacity on the temperature and slag composition has been studied quite extensively. Sections 4.3. and 5.7.2. discuss the concept and modelling of sulphide capacity with meta-analysis, and extend the studies conducted in Publication II.

According to the formulation of Deng and Oeters (1982), the overall mass transfer coefficient in steady flow conditions also depends on sulphide capacity (Deng & Oeters 1982). In a steady flow system, Choi et al. (2001) and Tong et al. (2017) propose a linear relationship between the overall time constant and slag's sulphide capacity. However, the mass transfer rate for the slag-metal reaction also depends on the composition and temperature through viscous mass-transfer and the diffusion rate, which cannot be considered independent of the composition, making the system non-linear in nature. Moreover, the mass transfer in the ladle depends on the convective flows in the system (Oeters 1994). Indeed, the convection dominates diffusion-related mass transfer due to the large gas flowrate, and the reaction volume is therefore often assumed to be perfectly mixed, allowing the lumping of the spatial dimensions. In practice, this assumption allows the use of mathematical models with a phenomenological basis online due to the reasonable computational time. Due to the large gas flowrate and bubble bursting phenomenon (Oeters 1994), during the injection, a large number of metal droplets are induced from the metal to the slag phases, in which they get entrapped. In theory, these droplets might react with the slag phase, further promoting the reaction. However, as the exact droplet size distribution and the behaviour of the slag phase surrounding the droplets is not well known, this mechanism is usually ignored in the models and in this work.

As the area for the slag-metal interface is significantly smaller than that of the surface area of the fine-grade powder, the overall rate for the permanent contact reaction is often considered significantly slower. Rastogi et al. (1994) estimate that

the contribution of the slag phase is half that of the reagent contribution, whereas Seshadri et al. (1997) estimate that the permanent contact reaction essentially defines the overall rate. However, these attributes greatly depend on the chosen residence time for the transitory contact reaction. As Oeters (1994) proposes, the emulsification of the slag droplets in the metal phase in a liquid slag cause the system behaviour to approach an emulsified reaction system, which drastically increases the reaction rate. However, due to the relatively low temperature, the solubility of the solid CaO in the slag phase is very limited, making the slag a heterogeneous system, consisting of the liquid slag phase and the solid particles, especially with high reagent injection rates. Chiang et al. (1990) study the effect of different slag conditions on calcium carbide-based desulphurisation, finding that the dry slag in fact provides a slower rate than the liquid slag. This property is attributed to the capacity of liquid slag to dissolve the reaction products. Similar behaviour was observed when the metal phase was not initially surmounted by the slag phase (Chiang et al. 1990). The computational complexity for the heterogeneous systems are naturally quite high, and the derived models often therefore consider the slag phase to be fully liquid, and the mass transfer coefficients are expressed with empirical correlations, such as the correlation of Riboud and Olette (1982), which is given as

$$\beta_{tot} = \tau \left(D_{[S]} \frac{Q_{tot}}{A} \right)^{0.5}, \tag{12}$$

where τ is an empirical describing the relation between the gas flowrate Q_{tot} through a specific surface area A in bath surface conditions (temperature and pressure) and $D_{[S]}$ is the diffusion coefficient of sulphur in the hot metal. The authors suggested a value of 500 ($m^{-0.5}$) for τ to be suitable for converter and ladle metallurgical operations (Riboud & Olette 1982). In hot metal desulphurisation, due to the injection of solid species, the composition of the slag phase evolves dynamically during the injection. Practical observations have revealed that the slag is initially liquid at the operating temperature, but during the injection, it approaches a heterogeneous system due to the high content of non-dissolved CaO particles. The compositional details of the slag phase are discussed in Section 4.1., which describes the experimental data used in this study.

2.3 Comparable models presented in the literature

As a detailed review of the models proposed for hot metal desulphurisation has been given in Publication V, this section focuses on comparable models, i.e. those that use a similar methodology to that of this study. The focus of this section in terms of the presented categorisation is illustrated in Figure 2. It can be seen that data-driven modelling techniques often rely either on parameterised reaction or parametric regression models, including neural networks, or a multiple linear regression and its variants. Overall, data-driven models are a minority in both the modelling of hot metal desulphurisation and more broadly, in the metallurgical context. The reasoning for this is provided in Publication V and in Sections 2 and 3. The next subsections provide a more detailed description of these models.

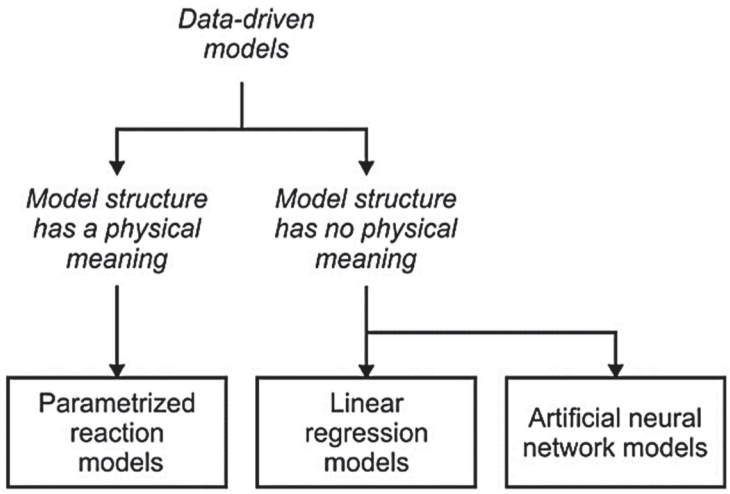

Fig. 2. Categorisation of the existing data-driven models for hot metal desulphurisation. (Reprinted [adapted] under CC By 4.0 lincense from Publication V © 2019 Authors).

2.3.1 Rastogi et al. (1994)

In the study of Rastogi et al. (1994), a phenomena-based model proposed by Deo and Boom (1993) is used as an identification basis to study the contribution of different desulphurisation mechanisms in carbide-based hot metal desulphurisation. The model belongs to the category of parameterised reaction

models, because it relies on the phenomenological description of the reaction rates that are corrected with explicitly estimated empirical parameters. The model parameters are identified with a binary-coded genetic algorithm, which is compared with the analytical solution of the least-squares objective function. The genetic algorithm is found to provide a more realistic estimate of the contributing mechanisms than the analytical solution. It is also found that the error for the training set of the parameterised model is 58% smaller than the prediction error for the non-parameterised model. The study's authors evaluate the figures of merit for the test set as $R^2 = 0.69$ and MAE = 19.1 ppm. As the authors do not report the prediction error for the end sulphur content for the external validation set, they evaluate the model performance from the dimensionless quantities given in the study.

2.3.2 Datta et al. (1994)

Datta et al. (1994) present a neural network model for hot metal desulphurisation. The model is manually selected, and trained using the backpropagation and gradient descent methods. The figures of merit for the external validation dataset, consisting of 11 treatments, are $R^2 = 0.60$ and MAE = 27 ppm. The model structure is selected manually, and the input variables are sulphur content before the treatment, treatment time, the mass of the hot metal, and the gas and CaC_2 flowrates. However, the gas flowrate is constant in the data, and there is therefore no covariance between the gas flowrate and the observed sulphur content, which means the effect of input gas flowrate is non-observable in the data.

2.3.3 Deo et al. (1994)

The study of Deo et al. (1994) presents a neural network model whose learning rate is meta-optimised with various methods. The model uses the parameterised model presented by Rastogi et al. (1994) as one of the input variables. The model structure is identified using trial and error. In addition, the authors propose that the logistic transform of the output enhances the convergence of the network. The model performs similarly to the previous model, but for a larger test set of 45 treatments.

2.3.4 Dudzic and Zhang (2004)

Dudzic and Zhang (2004) apply an adaptive partial least-squares (PLS) regression modelling algorithm for the online control of hot metal desulphurisation. They use the poor knowledge of the underlying mechanisms as an argument for constructing a fully data-driven model. The constructed models use 14 input variables, and the data is divided into two parts, based on the used reagent. To achieve adaptivity, the transformed input data is treated with moving average transformation, and the new model parameters are updated accordingly. The authors do not report the quantitative estimation error for the end content of sulphur, but the model can provide a well-correlating estimate of the reagent consumption with the true value of the consumed reagent. The article of Quinn et al. (2002) considers the use of the proposed algorithm in plant practice.

2.3.5 Bhattacharya et al. (2004)

The study of Battacharya et al. (2004) considers the use of PLS regression to predict carbide consumption in the lance injection process. The study finds that the PLS model performs slightly better than the multivariable linear regression (MLR) model. However, the dimensions of the MLR model are not reduced, and a detailed variable selection is not considered. The study finds that the most influential variables are the initial, final, target, and turndown sulphur content.

2.3.6 Vinoo et al. (2007)

In the study of Vinoo et al. (2007), a piecewise linear regression model is constructed with manual selection techniques to predict the consumption of CaC_2 in hot metal desulphurisation. The model constitutes four linear equations that are identified for pre-classified temperature ranges. The figures of merit $R^2 = 0.48$ and MAE 15.3 ppm are defined for 15 treatments in online use. The properties of the reagent are not accounted for.

2.3.7 Dan et al. (2008)

In the approach of Dan et al. (2008), model selection is conducted manually to predict three attributes of a Kanbara process. The output variables of interest are the consumption of the reagent, stirring speed, and stirring time. To this end, three

different configurations of artificial neural networks are used. The authors use a semi-empirical relation to determine the optimal number of hidden neurons instead of selecting the hyperparameters with cross-validation.

2.3.8 Feng et al. (2019)

In this work, the authors use a case-based reasoning model and a parameterised model to predict the end content of sulphur in a Kanbara hot metal desulphurisation. The final estimate of the end content of sulphur is constructed based on the weighted and biased Euclidean norm of the predictions of the case-based reasoning and the parameterised model. The case-based reasoning model works similarly to the k-nearest-neighbour (*kNN*) algorithm, because it regresses the future outcome as the mean, median, or equivalent of the similar observations, i.e. those closest in the feature or the variable space (Altman 1992).

3 Data-driven mathematical modelling

Mathematical models are a set of equations that describe the behaviour of a static or dynamic system, or a specific phenomenon prevailing in it (Szekely 1988). Two very general purposes of mathematical models are to explain the mechanisms or predict the future outcome of the system (Shmueli 2010). With mathematical models, the output variable Y is to be predicted or explained based on a set of input variables $X = [x_1, x_2 \ldots x_n]$. The relation between the output and input variables is linked with a mechanistic or a generic statistical function, i.e. $Y = f(X)$, where $f(X)$ is the assumed functional relation between the input and the output variable. In the context of regression problems, Y is defined in continuous space, but the variables in X can be either discrete or continuous. Especially in the context of statistical models, the models usually contain parameters that are estimated based on the available data. Thus, the form of the prediction model becomes (Hastie et al. 2009)

$$Y = f(X, \theta) + \varepsilon, \tag{13}$$

where $f(X, \theta)$ is the parametric functional form of the model, θ is the model parameter vector, and ε is the modelling residual (or modelling error), following a certain distribution. The distribution of error term is usually assumed homoscedastic, i.e. $\varepsilon = N(0, \sigma^2)$. The model itself is denoted as $f(X, \theta)$, which can be interpreted as the conditional expectation of Y with given X and θ, i.e. $f(X, \theta) = E(Y \mid X, \theta)$ (Hastie et al. 2009).

As Publication V outlines, *numerous mathematical models are available for hot metal desulphurisation*, and a categorisation of models is therefore necessary, especially as a huge number of models is developed for different purposes with different approaches. In the mathematical modelling of systems and processes, there are several default examples of model categorisation. A rough categorisation distinguishes between model types based either on data or domain knowledge, i.e. quantitative information or prior knowledge (Sohlberg & Jacobsen 2008). However, this categorisation is quite narrow concerning the current state of the art in the field of modelling, especially because the field of *machine learning* and *data-driven modelling* has recently gained increased popularity mainly because of the results achieved with deep learning in pattern recognition applications (LeCun et al. 2015). A more detailed categorisation would be the frequently referenced division based on the model structure in which the different types of mathematical model are often differentiated as follows (Sjöberg et al. 1995):

- *White box models* are models, which are constructed based on prior knowledge and physical principles, by making argumentative assumptions concerning the system and its phenomena. In such a case, the model functions ($f(X)$) tend to have a physical significance, but the function itself contains no parameters that fit the available data on the system to be the model, i.e. the expected model output does not depend on θ.
- *Black box models* have no physically significant structure, and the model is identified solely based on data. Black box models are therefore usually generic, providing flexibility for modelling. In this study, black box models are also referred to as data-driven models.
- *Grey box models* have a physically meaningful structure but contain parameters (θ) identified from the data-driven estimation of the relation between the *system inputs* and *system outcomes* based on the data to be explained or predicted. In the literature, grey box models may also be called semi-mechanistic models (Te Braake et al. 1998). The parameters can have a physical significance or meaningless units. In the statistical model context, the estimation process is referred to as model training, parameter estimation, etc., whereas in the field of process control and system identification, the term that is often used is model parameter identification. Another approach to grey box modelling is given in Sohlberg and Jacobsen (2008) and Stosch et al. (2014), who introduce serial and parallel configuration of white box models with black box models, resulting in an overall grey box model structure.

Publication V provides an overview of model types in the context of hot metal desulphurisation. The categorisation relies on the above categorisation but distinguishes between the different model subcategories. For example, *grey box* models in the context of hot metal desulphurisation are referred to mostly as *parameterised reaction models,* whereas, for example, the category of *black box models* sprouts in *linear* and *non-linear regression,* of which MLR, PLS, PCR, and ANN are special examples. As the literature concerning the use of black box models to predict hot metal desulphurisation is relatively scarce, the categorisation is not sprouted further, because it would not support the outline of this thesis, and much more influential reviews and books are available for this category's models. More detailed descriptions and reviews of neural networks can be found in LeCun et al. (2015) and Haykin (1994), and the fundamentals of the random forests can be found in Breiman's study (2001), not to mention the huge number of other available algorithms. For example, model ensembles (bagging and boosting) are very

influential, but the algorithms are not found to serve this study's purposes. For efficient boosting and bagging algorithms in pattern recognition, the reader is advised to begin with the studies of Chen and Guestrin (2016) on XgBoost, Breiman (2001) on random forests, and Hastie et al. (2009) on AdaBoost. It should also be noted that there are many more ways to categorise prediction models. For example, regression models can be further divided into nonparametric and parametric regression (Altman 1991). However, this categorisation is impractical in this context, because there are very few examples in the literature that consider non-parametric data-driven methods in the modelling of hot metal desulphurisation. Indeed, the only study referenced in this thesis that uses non-parametric estimators in the modelling of hot metal desulphurisation is the study of Feng et al. (2019). Furthermore, the data-driven approaches can be categorised as non-supervised and supervised learning methods (Hastie et al. 2009), but because the methods considered in this thesis and in the previous literature on hot metal desulphurisation mostly consider supervised learning, this categorisation is not discussed further.

3.1 Applications of mathematical models

A fundamental question determining the methodology to select a mathematical model and the suitability of a model per se for a problem under study is "*Does this study consider explanatory or predictive modelling?"* The division between explanatory and predictive modelling originates in the field of statistical modelling, but the principles of model selection also intersect with the principles of mechanistic white box and grey box model selection, especially concerning the evaluation of the *model performance* (Sargent 2010), but also of model training and validation. In explanatory modelling, the general objective is to find a model that helps to answer the formulated causal hypotheses, whereas in predictive modelling, the models are built to predict the future outcomes of the system based on the examples drawn from historical data (Shmueli 2010). The methodology for evaluating the explanatory and predictive models is dissimilar, and it is therefore necessary to differentiate the approaches at a very early stage. The objectives of these three types of model can be summarised as in Szekely (1988) and Shmueli (2010):

- *Descriptive modelling* uses a mathematical model to construct a description of underlying relations in the data. The objective of descriptive modelling is not

to explain the phenomenological basis by studying the causal hypothesis but to investigate the relation between X and Y.

- *Explanatory modelling* uses mathematical models to test the formulated causal hypothesis. However, the hypothesis testing can be seen very differently, depending on whether the studied model is a statistical or mechanistic model. For example, the hypothesis tests are well-defined for multivariable linear regression, but the techniques to study the causality of phenomena in the case of mechanistic models are very different (Sargent 1988; Sargent 2010).
- *Predictive modelling* uses mathematical models to predict the system's future outcomes given the set of input data (Shmueli 2010). In predictive modelling, the main emphasis is on the model's predictive performance, and less attention is usually paid to explanatory analysis. If meaningful predictions are to be made, the need for data is generally higher than in the explanatory model.

This study mostly considers explanatory and predictive modelling in the context of both fully data-driven and parameterised grey box models. However, Publication V also presents the other modelling techniques available in the context of hot metal desulphurisation. In considering Publication V, it may be asked what the purpose of the presented models is in the context of hot metal desulphurisation. A high temperature and otherwise hostile conditions constitute a solid argument for the use of mathematical models in the context of full-scale extractive metallurgy (Saxén & Pettersson 2007). As it is very difficult to measure system properties, the models can offer a phenomenological insight into the system. In some contexts, it is a commonplace to refer to such model usage as soft sensing (Lin et al. 2007). Soft sensors refer to using software and models to sense the system conditions without physical measurements. As this study does not consider the final control application, the term soft sensor is not used or explained further.

As indicated in Publication V, most models for hot metal desulphurisation are phenomena-based mathematical or computational fluid dynamics models, both of which can usually be categorised as white box models. However, referring to the problem of *the residence time of the particles*, for example, the models usually contain empirically fitted parameters, even though the parameter identification procedure is often not very sophisticated but is conducted on a manual trial-and-error basis (Oeters 1994). The generic data-driven and parameterised models used for fully *predictive purposes* are a minority in the context of hot metal desulphurisation but are also relatively rare in the field of modelling ferrous pyrometallurgical processes. However, their usability is widely acknowledged in

the literature. Good but limited examples of the use of such models are given in the context of metallurgy other than hot metal desulphurisation (Petterson et al. 2007) and in other disciplines (Odom & Shadra 1990; Ghiaus et al. 2007; Barbosa et al. 2011).

Studies often refer to the use of models in process control and optimisation purposes. For example, it is claimed the model by Ma et al. (2017) is designed for this purpose. However, with reference to the argument of Saxén and Pettersson (2007) on the hostile conditions for measurements, the validation data available for testing the predictive performance of the models is often scarce, which, according to statisticians, does not encourage the use of models to predict future outcomes, especially if the model contains any parameters fitted to the experimental data (Hastie et al. 2009). More often, the objective is to *explain the phenomena and causalities concerning the operating parameters and the process behaviour, and furthermore to study their effect on the process outcome.* For example, Chiang et al. (1990), Coudure and Irons (1994), Zhao and Irons (1994a), and Zhao and Irons (1994b) use a detailed model-based analysis to *explain* the rate phenomena during injection. In this context, the phenomenological analysis appears very useful, but as shown by the analysis of Oeters (1994) and Chiang et al. (1990), the assumptions made affect the results.

3.2 Model selection

The purpose of model selection is to select the best mathematical model from the set of candidate models. In the literature, a term that is often used is the *true model* of the system (Chatfield 1995). The true model can be interpreted as a model that explains the mechanisms of the underlying system in arbitrary accuracy. In model selection, the existence of a true model has often been questioned. Instead, it is recommended to find the best approximative model based on the data (Chatfield 1995). The criterion that determines *which model is best* depends on both the model's final application and its type (Shmueli 2010; Heinze et al. 2018).

Some very general preferred criteria that seem independent of the model type and application field can be summarised as *accuracy* and *interpretability.* In the context of data-driven models, *parsimonicity* is also a desired criterion (Burnham & Anderson 2002), but with reference to Publication V, it is poorly applicable to mechanistic models, because the models usually seek to explain the causalities with relatively exhaustive descriptions, especially in the case of CFD models. Thus, a good mathematical model could be summarised at least as an accurate but

interpretable abstraction of real-world complexities. As model selection techniques are very different in mechanistic and data-driven models, the next subsections are structured with this in mind.

3.2.1 Data-driven model selection

In statistical models, data-driven models, machine-learning models, etc., the semantics may differ, but the basic principles are very similar. The objective is to select an explanatory or predictive model by *using a collection of observations, i.e. a set of experimental data.* If the dataset on whose basis the prediction models are to be constructed has already been collected and pre-processed, data-driven prediction model selection (i.e. *black box* or *grey box* model selection) can be divided roughly into five steps:

1. model structure selection;
2. variable construction or feature extraction;
3. variable or feature selection;
4. model parameter identification or model training;
5. model validation or model performance assessment.

Some of these introduced steps can be gone through sequentially or in parallel, depending on the components implemented to carry out the task. For example, in the case of model-based variable selection algorithms (i.e. wrapper algorithms), the model must be trained during the selection (Kohavi & John 1997). As it is more common to call computational variables "features", especially in the context of signal analysis, this study mainly uses the term "variable". However, Section 3.4. shows that the modelling scheme also includes the use of several computational variables.

The selection of a data-driven predictive model is a combination of discrete and continuous optimisation problems, where the model is selected such that the specified objective function is either minimised or maximised, depending on the objective. The discrete optimisation problem arises from the selection of the variables, whereas model parameter identification is usually defined in the continuous space. In the context of predictive models, the general objective of model selection is to identify the trade-off between model bias and model variance, i.e. to minimise the risk that is defined as the sum of model variance and the squared bias. However, estimating these quantities is not straightforward, and a Monte Carlo type approximation is usually used. For example, Baumann and Baumann

(2014) give a detailed description of such an approach. The bias-variance trade-off can be simplified to analyse model complexity as follows (Hastie et al. 2009):

1. A more complex model reduces model bias but increases model variance. Models with high variance tend not to generalise well to external datasets, because the model output changes drastically with given inputs.
2. A simplistic model has a large bias but small variance, i.e. the model cannot explain the variance in the output and does not generalise well to external datasets.

The above cases can also be described in terms of overfitting and underfitting. The model is said to be underfitted if it fails to describe the underlying mechanisms of the system, because the model is too simplistic, whereas the overfitted model replicates the data that fits very well but fails to generalise to external datasets. In other words, the prediction errors measured with the training and external validation sets are not consistent in such a way that the external validation error is much larger (Hastie et al. 2009). However, it should be noted that underfitted models also tend to generalise poorly, because they completely fail to describe the system. However, determining the *underfitting* by comparing the performance to the *true model* is seldom possible, because the data does not describe the *true behaviour* of the assumedly ultimately complex system (Burnham & Anderson 2002). For example, Burnham and Anderson (2002) differentiate the terms *underfitting* and *overfitting* to consider only the best available approximative model instead of the *true model*, which can be considered a much more practical approach (Burnham & Anderson 2002). To select a model with the least tendency either for over or underfitting, the *rule of parsimonicity* (Occam's razor) is often applied, which means that of the model candidates with equal prediction performance, the simplest should be selected. (Rasmussen 2001) However, it is common in practical applications to trade some variance of a model for model bias (Heinze et al. 2017). Thus, in model selection one often arrives at a principle of Occam's razor that advises the selection of a model that performs similarly to the other solution candidates but has the least complexity (Rasmussen 2001).

A comparison of the available approximative models can be conducted by using a performance metric. The objective function for data-driven model selection is usually formulated by using a performance assessment criterion. According to Baumann (2003), the objective function steers the variable selection in the right direction. Similar ideas have been expressed by Burnham and Anderson (2002) and Hastie et al. (2017), and various other authors who are not listed here, because

model selection is an extensively studied area. In regression, the objective function is usually based on the squared error function and its variants. In this study, the used variants are the sum of squared error (SSE), mean squared error (MSE), and root mean squared error (RMSE), determined either for the cross-validation set or the training set, depending on the case. In the case of statistical or data-driven models, the standard metrics for model evaluation include Akaike's Information Criterion (AIC) (Akaike 1974), the Bayes Information Criterion (BIC) (Schwarz 1978), Mallows's C_p (Mallows 2000) and cross-validation statistics (Shao 1993), and many others. The introduced information criteria degrade the estimated model performance by adding a penalty term describing model complexity, resulting in a favouring of parsimonious models over complex ones in the selection. Cross-validation estimates model variance and bias by exploiting a dataset that is independent of the model selection process (Hastie et al. 2009). In this study, the different variants of cross-validation, hypothesis testing, and information criteria are mainly used to guide model selection. The computational details of these are given in Section 3.7.

3.2.2 Mechanistic model selection

Referring to the division between the selection of mechanistic or statistical models, the steps taken in model construction differ methodologically but can be philosophically seen as overlapping with data-driven model selection. However, some steps are only suitable for generic models, whereas others apply only for mechanistic models, especially in terms of formulating the optimisation problem. For example, the term variable selection often refers to a case where a set of input variables is selected to explain or predict the system behaviour with a generic model. However, the construction of mechanistic models can be seen as a special case of variable selection and construction, because model selection involves the identification of the relevant phenomena to be described, and thus selects and constructs the necessary predictor variables within the model's derivation. This process can be formulated as the *model component selection problem*, assuming that there is a set of model components C, each describing a specific sub-model of a system, and the objective is to find the most suitable combination of components, constrained by the a priori information. Such an approach has been studied by Levy et al. (1997). However, as this analogy is somewhat shallow and tendentious, it is not proceeded with further. Mechanistic model selection, i.e. *white box model selection,* can be summarised as consisting of the following non-sequential steps

and their analogies with data-driven model selection (Sargent 1988; Levy et al. 1997; Sargent 2010):

1. formulate the hypothesis on the system behaviour based on domain knowledge;
2. select the relevant phenomena to be modelled, and how to model each phenomenon;
3. explain the phenomena with mathematical equations that accounts for the phenomena and assumptions;
4. evaluate the available system data;
5. implement the computerised model;
6. verify and validate the model.

If the model is a grey box, the variable selection and model parameter identification steps can be included in the selection framework in Step 5. With reference to the different architectures of grey box models, the variable selection can consider the selection of an auxiliary equation, sub-model, or similar. The identification of the model parameters can be conducted similarly to the training of the data-driven models, i.e. by minimising the model prediction error by using the squared error function (Tan & Li 2002).

Similarly, as in the context of data-driven models, there is often a question of the goodness of the model with mechanistic models. As the outline of this thesis is not to provide a comprehensive overview of the limitations of white box models, the next chapters cut some corners. In the case of mechanistic models, the *true model* context can be attributed to describing *all the phenomena* that define system behaviour. However, in the context of computationally complex white box models (i.e. CFD models), the endeavour towards the true model is clearly limited, for example, by the accuracy of the domain knowledge, used boundary conditions, and computational complexity, because the discretisation method applied for the computational lattice may be a significant source of error, for example. In addition, detailed white box models usually also contain sub-models (or auxiliary equations) that are constructed by means of data-driven techniques, of which chemical reaction models and turbulence models are good examples (Oberkampf & Trucano 2002). Obviously, the error associated with the data-driven auxiliary equations accumulate to the total error of the mechanistic model. As indicated in Publication V and the other literature, in the context of metallurgy and hot metal desulphurisation, the accurate description of the injection scheme requires the solving of the governing equations for transport phenomena, which are defined in the time and spatial domain (Oberkampf & Trucano 2002).

With reference to data-driven model selection, it would also be beneficial to select mechanistic models in terms of *the best explaining model among all model candidates*. However, as the computational complexity of solving the governing equations is great, this assumption of several model candidates with respect to the auxiliary equation, boundary condition, and numerical solution strategy space is indeed very ambitious and ambiguous. Instead, candidate models are often conceptual, meaning that the abstraction is constructed based on the arguments considering the system behaviour before the actual implementation. Thus, in mechanistic model selection, one often arrives at selecting a model with an *acceptable level of agreement with the domain of intended application* (Oberkampf & Trucano 2002).

3.2.3 Parameterised models

As discussed in Publication V, parameterised prediction models are referred to as models for which a physically meaningful structure is constructed by making argumentative assumptions concerning the system behaviour. The model structure is then used as a basis of the parameter identification process, i.e. the model is fitted to the experimental data using mathematical optimisation. As stated earlier, this approach is often referred to in the literature as the *grey box approach*, because it is a combination of *white box* and *black box* approaches. According to the literature review we present in Publication V, nearly all the suggested models for hot metal desulphurisation contain some experimental parameters and functions, mainly related to the determining of physical quantities, even though the model structure is physically meaningful. However, these parameters are defined independently for the data used to validate the model. Thus, in Publication V, parameterised reaction models are referred to as those that use a sophisticated optimisation strategy for parameter identification.

A non-parameterised mechanistic reaction model is usually validated with data from the actual system. Parameterised models differ from these in that they are trained according to the data gathered from the process. Thus, parameterised models are by default more accurate at explaining the variance of the studied data but are assumed to be less generalisable to other contexts than non-parameterised mechanistic models, because the parameters capture some variance that is induced by the specific process due to differences in equipment, sampling technique, analysing technique, and so on. The parametrisation strategies used in this study are discussed in more detail in Section 3.6.

Parameterised reaction models can be either dynamic or static. If the model contains several auxiliary equations, the computation procedure for a dynamic parameterised model usually consists of the computation of several constructed variables. Because of these properties, the variable selection step is poorly applicable in constructing these models, and the model structure is therefore usually selected manually. In modelling, the trade-off between the computational complexity and level of detail in the mathematical description of the system is of great importance, because the final application determines the weighting of these individual factors. Computational complexity can therefore also steer the selection of a parameterised model, albeit depending on the final application. The lumped parameter approach is often applicable to computational complexity, meaning that spatially defined model equations are simplified and presented in fewer dimensions, i.e. the model defined in t–x–y–z dimensions is described, for example, only in the time domain (Suesserman & Spelman 1993). Indeed, in Publication V, most of the mathematical models for hot metal desulphurisation are lumped parameter models that significantly simplify the computational domain.

As stated, the parameters can have a physical meaning, or they can be dimensionless. For example, Katare et al. (2004) use metaheuristic and local search methods to solve rate constants, entropies, and changes in activation energies in large-scale kinetic models. The test problems reported by the authors show that the optimisation of the reaction rate constants is common, especially in consecutive and parallel reaction systems. However, it is also found that the suitable parameter ranges are wide, and there are several equally good solutions to the problem.

3.3 Model structure selection

This section describes the model structure used in this study with reference to the earlier sections and Publication V. Model structures are selected for generic models such that the models remain interpretable with respect to input variable space. Hence, the modelling techniques that rely on projection of X into a new variable space, such as partial least squares regression (see e.g. Gelaldi & Kowalski (1986)) and principal component regression (see e.g. Jolliffe 1982) (and many others), are omitted from the thesis, even though their existence as model structure candidates is acknowledged.

3.3.1 Generic data-driven models

Generic prediction models are used in Publications I–V. In generic models, the functional form of the model includes no physically meaningful relations, which makes the models flexible and suitable for situations where no external knowledge is available about the system. However, as shown in Publications I, III, and V, the existence of a physical meaning is not straightforward to include or exclude in model categorisation. The suitable functional form depends on the relations between the selected input and the output variables. For example, linear models assume linear dependencies between inputs and outputs. The next subsections introduce the model types considered in this study.

Multiple linear regression

In this study, the multiple linear regression (MLR) models were used for both predictive and explanatory analysis. The general form of a multiple linear regression is (Harrell 2001)

$$y = b_0 + \sum_{i=1}^{k} b_i \, x_i + \, \varepsilon, \tag{14}$$

where b_0 is the bias term, and b_i is the corresponding regression coefficient for variable i. The coefficients can be interpreted such that the values define the direction and magnitude of the change in the output variable when an explanatory variable is changed by one unit. The bias of the multiple linear regression can be low only if the dependencies between the explanatory and output variables are linear. However, some transformations are useful to preserve the linearity of the parameters and to allow the analytical solution for the estimates. For example, a logistic transform of the explanatory and the predictor variables can be used (Harrell 2001). Estimates of the unknown parameters $b = [b_0, b_1 \ldots b_j]^T$ can be obtained by minimising the sum of the squared error (SSE) between the model predictions and measured outcomes. b, which minimises the SSE, can be obtained with the matrix pseudoinversion (Harrell 2001)

$$b = (\mathrm{X}^T\mathrm{X})^{-1}\mathrm{X}^T y. \tag{15}$$

Artificial neural networks

Artificial neural networks (ANN), usually called simply “neural networks”, are computational models that are inspired by the biological neurons of natural brains. Depending on the network structure, neural networks can be used for classification, regression, and clustering problems. A well-known example of the use of neural networks is the field of statistical pattern recognition, which is a typical classification problem (LeCun et al. 2015). In regression problems, the output to be modelled and predicted is continuous, whereas a classification model maps the output of the network to discrete space with the decision rule. In this study, neural networks are used to predict the sulphide capacity in Publication II and the end content of sulphur in carbide-based hot metal desulphurisation in Publication IV. Publication IV also studies the use of the hybrid encoded genetic algorithm in the simultaneous variable selection and selection of the number of hidden neurons. In this context, modelling problems are therefore defined in continuous space, which is why the classification networks are excluded from the study, even though they are very influential and extremely popular (LeCun et al. 2015).

In function approximation and regression problems, neural networks are considered universal approximators, because they can map any continuous function with arbitrary accuracy, provided that the network structure is sufficiently complex (Hornik et al. 1989). Neural networks are therefore suitable for predicting even complex and non-linear system behaviour, based solely on the available data about the system. With reference to categorisation, neural networks are themselves considered black box models, but can also be implemented in series or in parallel with a white box model, defining the overall model as a grey box type. Good examples of a parallel grey box model with a neural network in the context of chemical engineering can be found in the study of Xiong and Jutan (2002).

Although the networks are flexible and generic, the selection and training of a neural network model cannot be considered trivial. For example, even in simple multilayer feedforward networks, there are several hyperparameters to be defined, including the input variables, number of hidden neurons, fraction of connectivity, and used activation functions (Hagan et al. 1997). In a regression neural network, either a hyperbolic tangent or non-scaled sigmoid function is used as the hidden layer activation, and a continuous linear function is used as the output activation function. With reference to the network’s layered structure, the model output can be given with the following function decomposition, with a linear activation function in the output layer:

$$Y = w_0^o + \sum_{i=1}^{k} w_i^o \, g\left(W^i X + w_0^i\right), \tag{16}$$

where w_0^o is the bias term of the output neuron, w_i^o is the weight coefficient between the output layer and the hidden layer, g is the hidden layer activation function, W^i is the weight matrix between the input variables and the hidden layer, and w_0^i is the bias term of the hidden neuron. It is common to use the sigmoid type activation function in the network in regression problems. This study uses the hyperbolic tangent (i.e. the scaled sigmoid) function as the hidden layer activation. As the number of parameters in a fully connected network is usually large, the amount of data needed to select a well-generalising network is large. Consequently, in this study, neural networks are used only for Datasets 2 (secondary desulphurisation) and 3 (sulphide capacity data). Identifying these weights is referred to as the training of a neural network. In the literature, numerous training algorithms are available; indeed, several algorithms suitable for continuous function optimisation are also suitable for the training of neural networks. The general idea of training is first to calculate the network output and then to update the network coefficients iteratively, such that the training error is minimised. The iterative steps are taken towards the gradient of the error function. In most approaches, the gradient of the error is computed with the backpropagation algorithm (Lecun et al. 2015). In this study, the main training algorithm is the Levenberg-Marquardt algorithm with or without Bayesian regularisation. Exhaustive descriptions of these algorithms can be found in the studies of Hagan and Menhaj (1994) and Foresee and Hagan (1997), for example. These algorithms have proved useful in the training of feedforward neural networks, and their implementation is relatively simple.

The Levenberg-Marquardt algorithm relies on the second-order derivative of the objective function, i.e. the Hessian matrix. However, the Hessian matrix is approximated with $J^T J$, where J is the Jacobian matrix that is constituted in the first-order derivatives of the prediction residual (Hagan & Menhaj 1994). As a second-order method, the Levenberg-Marquardt algorithm is relatively fast, and often performs well, provided the network architecture is not too complex, making the computation of the Jacobian matrix computationally inefficient compared to first-order methods. The Bayesian regularisation increases the generalisation capability of the networks, because it simultaneously minimises the sum of squared weight coefficients and the training error. The regularisation parameter can also be specified prior to model identification, as in the classic Lasso regression (Tibshirani

1996). In the Bayesian regularisation, Foresee and Hagan (1997) propose optimal regularisation parameters α and β, which are estimated in the Bayesian sense by using the information concerning the number of parameters that are effectively used in the minimisation of the objective function (Foresee & Hagan 1997).

If a neural network is used as the model basis in wrapper-based variable selection, the computational complexity of the iterative network training needs to be addressed. To tackle this issue, the wrapper algorithm employed in Publication IV uses the extreme learning machine (ELM) architecture originally proposed by Huang et al. (2006). In the ELM model, the hidden layer weight matrix is left untrained, and the weight coefficients between the hidden and the output layers are solved with the Moore-Penrose inversion described earlier. A more detailed description of this approach is given in Publication IV. Some variants of ELM conduct the training of the hidden layer weight coefficients with a metaheuristic algorithm such as particle swarm optimisation (PSO) (Zeng et al. 2017) or the real-coded genetic algorithm (Suresh et al. 2010), although this approach is not considered here.

3.4 Variable construction

Variable construction entails generating a set of useful variables from a set of raw data. Suitable variable construction techniques are somewhat case-specific and depend on the available data (Guyon & Elisseef 2006). However, many standard techniques and procedures are available, especially in the field of signal and image analysis. In the context of image analysis, variable construction is quite intuitive, because the images themselves contain objects and patterns that can be easily distinguished by human perception, but without *extracting the features* from the image, the image data itself is useless for *pattern recognition*, because it contains spatial variation, thus requiring local transformation operations. It should be noted that image analysis is used here only as an analogy for illustrative purposes.

On the other hand, variable construction can refer to data pre-processing. Pre-processing techniques include *standardisation, scaling*, *normalisation*, *noise reduction,* and *discretisation*. The purposes of these operations are further discussed in Guyon and Elisseef (2006). Of these, this study uses the standardisation and scaling operations, and some variants of the Box-Cox transformation (Box & Cox 1964). Standardisation is given as (Han et al. 2011)

$$X' = \frac{X - \overline{X}}{\sigma_X}, \tag{17}$$

where X' is the transformed column vector that has a zero mean and unit variance. The scaling of data means scaling the data to a pre-specified interval, i.e. such that $X' = \{a, b\}$ (Han et al. 2011). As a default, this study uses $X' = \{0, 1\}$ as the scaling interval.

In the context of this study, an obvious example of a constructed variable is *the sulphide capacity* (C_S) of the slag, which is estimated based on the chemical composition of the slag and other system conditions. In this analogy, data pre-processing is linked to *dimensional reduction*, because the original dimensionality of the data would be from $n+k$ to 1, where n is the number of slag components, and k is the number of system conditions. The sulphide capacity is then used to determine the equilibrium stage of the permanent contact reaction through to the evaluation of the sulphur partition ratio between the slag and metal phase. The construction of the sulphide capacity from the raw data is explained in detail in Publication II and Section 4.3. In the metallurgical context, it is common to estimate the mass-transfer of the available mass-transfer correlations that use the *Reynolds'*, *Schmidt,* and *Sherwood* numbers, for example (Clift et al. 1978). Similarly, the whole particle size distribution is not applicable in a generic data-driven model, because the dimensionality of the model would explode. Instead, deriving a meaningful characteristic value such as the *Sauter mean diameter* (Pacek et al. 1998)*, mean volume-based size*, or *median area-based size*, all of which describe the properties of the distribution, is an example of constructing a suitable predictor based on a raw measurement.

The importance of domain knowledge cannot be underestimated in variable construction. Indeed, most of the complex simulation models use variable construction techniques that are based on theoretical, empirical, or semi-empirical models. However, there is no actual need to draw a solid analogy between dynamic simulation and data-driven prediction models, for which the concept of variable construction was originally introduced. Instead, acknowledging and understanding this connection may help in deriving more meaningful data-driven models.

From the examples above, it is obvious that the constructed variables may be correlated, raising the problem of multicollinearity, i.e. a situation in which the explanatory variables have a strong linear relation with each other. This is detrimental, especially for the *explanatory analysis of the models*, because the multicollinearity increases the error of the parameter estimates and may saturate

the individual effects (Harrell 2015). In addition, as the dimensionality of the data increases within the number of constructed variables, there is potential to increase the number of redundant, irrelevant, and noisy variables (Guyon & Elisseef 2006). To distinguish these variables from the relevant ones, input variable selection is needed (Guyon & Elisseef 2003).

3.5 Variable selection

Input variable selection plays a crucial role in data-driven model selection (Heinze et al. 2018). The need for variable selection arises within the increased number of available input variable candidates. Variable selection can be seen as the most computationally intensive step in model selection, especially when there is a large number of variable candidates in the dataset, and no domain knowledge, either of the suitable number of predictor variables or the suitable predictor variables themselves, is available. This is because the number of available models increases exponentially within the number of variables (Kohavi & John 1997). Variable selection has been discussed as a way to improve model performance and interpretability. The improved performance can be attributed to the improved generalisation of the model through to decreased variance, because decreasing the number of input variables reduces the model complexity (Hastie et al. 2009). In addition, in very large models, variable selection also decreases the computational load of the model parameter identification if each variable is associated with a corresponding parameter (Guyon & Elisseef 2003).

The primary objective of variable selection is to find a suitable subset of variables that maximises model performance in terms of the assessed criteria. The criteria themselves in variable and model selection are similar. In addition, as in model selection, the criteria can be formulated as objective functions to be maximised or minimised. The importance of the objective function in the context specifically of variable selection has been discussed and illustrated by Baumann (2003). Philosophically, variable and model selection have similar objectives, but the variables are usually selected for a pre-specified model structure, of which a very common example is multiple linear regression (Heinze et al. 2018), whereas model selection involves the selection of the model structure and its parameters (Heinze et al. 2017). However, it should again be emphasised that the wrapper approaches for variable selection demand simultaneous model parameter identification (Kohavi & John 1997). In the variable selection for a predictive model, the objective is to maximise *predictive performance*, whereas in the variable

selection for explanatory models, the main emphasis is on testing the causal hypotheses and the system's theoretical considerations (Shmueli 2010).

When choosing the subset for a prediction or explanatory model, selection can be conducted manually or automatically by algorithmic means. In manual selection, the domain knowledge of the system is used for selection or as an initial guess for selection. There are obvious advantages in manual selection, such as the fact that an input variable in a model is included because of true causality (Heinze et al. 2018), provided that the domain knowledge is sufficiently precise. However, whether this dependency provided by the knowledge can be identified from the data is often another question, because the relation can be biased or non-observable because of other sources of variation, including measurement noise and collinearities. This issue, among others, could make the subset selection based on the hypothesis testing problematic, and thus support the use of domain knowledge in selection (Burnham & Anderson 2002; Harrell 2015).

The algorithmic approach can be considered a good alternative for variable selection if the problem to be modelled is very complex (Breiman 2001), the domain knowledge is not properly addressed, and a large number of variable candidates is available. The variable selection algorithms can be further divided into *wrappers*, *filters*, and *embedded methods* (Guyon & Elisseef 2003), which can be described as follows:

– In wrapper algorithms, the search is carried out based on the objective function that considers the performance of the model itself (Kohavi & John 1997). The wrappers are computationally intensive, because they require several model parameter identifications and performance evaluations to construct the final solution.
– Filter approaches are computationally more efficient, because the variables are selected based on a specified ranking criterion. Several criteria exist, but the correlation and the mutual information between the input and the output variables can be considered influential (Guyon & Elisseef 2003).
– Embedded methods include shrinkage or regularisation operators such as Lasso (Tibshirani 1996), among many others. In these techniques, the model parameters are reduced towards zero, making the corresponding variables meaningless in the model. As stated earlier in Section 3.3.1., the Bayesian regularisation technique is similar to Lasso, but in this study, it is only used in the context of neural networks.

Generally, as can be summarised based on the thoughts of Breiman (2001), it is good not to stick to a single solution but to experiment with several techniques in constructing the solution (Breiman 2001). In this study, four main techniques are compared for model selection, i.e. manual selection with domain knowledge, exhaustive search, deterministic forward selection and its variates (3 versions), and genetic algorithms. These algorithms are chosen because it is characteristic of hot metal desulphurisation data that it contains a limited number of observations but a relatively large number of predictors, and is thus substantially exposed to the selection of overfitted or coincidental models (Baumann 2005; Harrell 2015). The next sections describe the implemented variable selection algorithms.

3.5.1 Manual selection

In the literature, manual selection is often referred to as variable subset selection using domain knowledge (Guyon & Elisseef 2003; Heinze et al. 2018). For example, Burnham and Anderson (2002), Guyon and Elisseef (2003), and Heinze et al. (2018) suggest the use of domain knowledge as a preferable strategy for subset selection, because domain knowledge contains information that the data itself does not easily reveal.

Domain knowledge can be used either in generic model or mechanistic model selection. However, the formulation of manual selection in the case of mechanistic models is not as straightforward as it is for generic models. In mechanistic models, manual selection can be seen as merely corresponding to the derivation of the model, although the model structure is selected simultaneously. In generic models, manual selection refers to the selection of a meaningful subset of variables to be used in a generic model structure (Heinze et al. 2018). However, the need to include a variable in a model is debated. It is known to influence the output, even though this does not appear significant. For example, Burnham and Anderson (2002) prefer the inclusion of models that make sense physically, and the exclusion of models of which estimated parameters do not make sense at all, even though they appear significant in hypothesis tests (Burnham & Anderson 2002).

With reference to the above guidelines, domain knowledge is of great importance in this study, both in model selection and the evaluation of model performance. The extraction of domain knowledge is mainly conducted with literature reviews that create the baseline for reasoning. Such studies are performed in Publications I, II, and V.

3.5.2 Optimal search strategies for variable selection

As stated earlier, variable selection can be treated as an optimisation problem, where the objective is to find the subset of variables that minimises the given objective function. The optimal search strategies are algorithms that basically evaluate all the possible subsets. This approach is also referred to as best subset selection (Hastie et al. 2009). However, the computational complexity of the optimal search is exponential, because the required number of model evaluations is 2^n–1, where n is the number of variable candidates. Because of the computational complexity, this strategy is considered only for linear models constructed from small datasets. The variant of best subset selection is the branch-and-bound algorithm (Guyon & Elisseef 2003), which is not considered here. As the variable selection problem is discontinuous, the optimal search strategies are the only ones that can guarantee the globally best model. In this study, this strategy is used as a comparison, but only for Dataset 1 (primary desulphurisation), because it is the only dataset for which the strategy is computationally feasible.

3.5.3 Deterministic algorithms for variable selection

Deterministic variable selection algorithms are strategies that solve the problem of variable selection with sequential elimination or additive operations (Guyon & Elisseef 2003). In the statistical literature, strategies have often been referred to as stepwise regression techniques (Harrell 2015). The use of stepwise regression with hypothesis testing has been criticised, because it results in nested models, of which the R^2 values are biased highly, and the p-values are too small due to an underestimation of standard errors and confidence intervals that are too narrow (Harrell 2015). Instead of stepwise regression, Harrell (2015) recommends using the full model for statistical inferences. An alternative view of this is the one proposed by Burnham and Anderson (2002), who state that the full model inference may be problematic especially in the case of a limited number of observations and large number of irrelevant variable candidates. According to the authors, this setting leads to a situation in which the bias in the estimates deteriorates the inference (Burnham & Anderson 2002). In this study, the selection of a proper subset is favoured over the full model inference, because the number of observations is relatively limited.

Forward Selection 1

The forward selection algorithm belongs to a family of deterministic model selection methods. The forward selection algorithm is initialised with an empty variable vector, and the variables are added to the model in a greedy manner, i.e. such that the variable that most increases the model prediction performance during the search is included. The algorithm has also been referred to in the literature as sequential forward selection (SFS) or the hill-climbing algorithm (Kohavi & John 1997; Guyon & Elisseef 2003) As was empirically proved by Kohavi and John (1997) in the context of classifiers, the main problematic of forward selection is that the algorithm often tends to stick in the local optimum in the variable space. The steps of the forward selection can be given as follows:

1. Add each feature candidate x_i to the trial model from a set $X_k = \{1, 2 \ldots M–k\}$.
2. Evaluate the objective function value for each of the trial models.
3. Select the best model along the set of $M–k$ trial models.
4. If the objective function value is improved, update the current model containing the k number of features.

It can be shown that worst-case complexity is the forward selection of order $O(n^2)$, where n is the number of available variable candidates corresponding to the case, where all the variable candidates existing in the data are selected in the model. This is often a more practical choice than the exhaustive search, of which the complexity is of order $O(2^n)$, because $2^n–1$ subset combinations are theoretically available. The algorithm does not consider the possibility that eliminating a variable from a model will improve prediction performance (Kohavi & John 1997; Guyon & Elisseef 2003). For this, floating search extensions of the sequential selection algorithms have been proposed (Pudil et al. 1994). Compared to the backward elimination strategy, which starts from a full set of variables, forward selection tends to select more parsimonious models but does not in theory evaluate the interacting variable combinations as efficiently as the backward elimination (Kohavi & John 1997). However, as the datasets in this study are rather small, and the number of observations is limited, one of the primary objectives is to keep models parsimonious, which is why forward selection is preferred to the backward elimination.

Forward Selection 2

Filters in variable selection are often used to reduce the complexity of deterministic wrappers. In this study, the tested filter approach combines forward selection and variable ranking. The algorithm proceeds as follows:

1. Sort the variables based on a ranking criterion into descending order
2. Add variables sequentially to the model
3. If the objective function value is decreased, the variable is added to the model and otherwise discarded.

The filter approach is designed to provide a fast alternative for model selection, because the complexity of the approach is only linear for linear regression. However, the performance of the filter greatly depends on the selected ranking criteria (Guyon & Elisseef 2003).

Forward Selection 3

In this study, the Forward Selection 3 algorithm is used for the explanatory analysis of the datasets and domain knowledge. The algorithm follows similar principles to the stepwise forward selection given in Heinze et al. (2018), for example. The modified algorithm proceeds as follows:

1. sort the values with respect to their significance (in t-test terms) in decreasing order;
2. initialise the model with the most significant single predictor;
3. add the variables sequentially to the trial model in the given order and estimate the model after each inclusion;
4. if the addition of a variable results in a statistically significant p-value ($\leq \alpha$), leave it to the model;
5. if any of the variables already included in the model get an insignificant p-value, eliminate them sequentially from the model;
6. reconsider the previously eliminated variables;
7. repeat until all variables have been included in the model at least once, along with some of the eliminated variables.

The computational details of the used hypothesis tests are given in Section 3.7.1. The reasoning behind the algorithm is that it favours significant single variables, but also considers their explanatory power if used with other variables. In addition,

the elimination step allows all variables that appear insignificant compared to the others to be discarded, whereas the reconsideration step makes it possible to include a variable that could be significant in terms of the new trial model, even though it is considered insignificant in previous iterations. By doing this, the algorithm guarantees a convergence to a subset in which all the estimated p-values are below the chosen risk level; the full model is chosen by setting $\alpha = 1$. This is necessary, because some of the variables may seem insignificant individually but may be useful as part of a multivariable model, and vice versa (Guyon & Elisseef 2003). However, with reference to the Freedman paradox (Freedman & Freedman 1983), which states that the input variables independent of the output variable can pass the hypothesis test, the solution provided by this algorithm is only used as a supportive solution for domain knowledge.

3.5.4 Stochastic algorithms for variable selection

Unlike deterministic algorithms, stochastic variable selection techniques rely on random operators to modify the solution vectors. Metaheuristic algorithms for variable selection are a family of methods that rely on the heuristic evaluation of solution candidates. The term metaheuristic refers to the fact that the solutions provided by these algorithms are not necessarily optimal, but otherwise of high quality. The stochastic metaheuristic algorithms that have been used successfully for variable selection include genetic algorithms, tabu search, simulated annealing, and particle swarm optimisation (Baumann 2003; Guyon & Elisseef 2003; Wang 2007), and many more. In this study, a genetic algorithm is chosen as the search engine to solve the variable selection problems, in which it is proved to perform relatively well, both in regression and classification problems (Siedlecki & Sklansky 1993; Leardi et al. 2002; Huang et al. 2007; Sorsa et al. 2013; Oreski & Oreski 2014). The genetic algorithm is considered the wrapper search engine, because it uses the system model to evaluate the solution candidates (Kohavi & John 1997; Guyon & Elisseef 2003).

When a genetic algorithm is used for variable selection, each of the solution candidates, i.e. the individuals in a population, corresponds to a binary-encoded variable vector. In this vector, if a gene corresponds to 1, the variable is selected and otherwise not (Siedlecki & Sklansky 1993). Consequently, the algorithm uses of binary crossover and mutation operators, which differ somewhat from the operations defined for real-coded genetic algorithms. Such a genetic algorithm is used in Publications III and IV. However, in the latter, the binary-encoded genetic

algorithm is responsible for selecting the subset of variables, whereas the neural network structure is selected using integer-based encoding, which results in a hybrid overall encoding of the chromosomes.

However, the selection operators are the same as in the real-coded version. The real-coded genetic algorithm with hybrid operators is used in this study for the parameter identification of a dynamic model, which is why the selection operators are reported within it. The following sections introduce the binary-encoded operators used in this study.

Fitness function

As the regression model selection aims to minimise the prediction error, the fitness function of the genetic algorithm is given as the inverse of the objective function. In this study, the genetic algorithms use the inverse of the internal validation error defined for N repetitions (or cross-validation error) as the fitness measure. To avoid the selection of collinear variables, Publications III and IV also use the VIF index, which is estimated based on the collinearities between input variables. In Publication III, the maximum VIF between input variable pairs is used, whereas in Publication IV, the VIF is defined as the largest element in the inverse of the correlation matrix. However, it was found that the problem was also solved well without constraining the objective function.

Crossover

In a binary-encoded crossover, the offspring of the selected parents is formed such that segments of two individuals, i.e. the parent individuals, are swapped with each other. The swapped segment of the chromosome is defined either with a single or two randomly chosen crossover spots (Goldberg 1989). Crossover between individuals is regulated by crossover probability p_C. Typically, a high crossover probability (i.e. $p_C = 0.8–0.9$) is used. In practical implementation, crossover occurs if a continuous random number in the range {0, 1} is smaller than the specified crossover probability p_C (Goldberg 1989).

Mutation

The mutation operator is implemented to increase the diversity of a population by making stochastic changes to individuals. In mutation, a random bit, several bits,

or a segment of bits in a chromosome are inverted (Goldberg 1989). The mutation of individuals is controlled by the mutation probability. The criteria for mutation to occur are the same as for crossover. The mutation operator can be implemented in various ways, but the following implementations are used in this study:

1. Uniform dot mutation – A single bit of an individual is inverted with a probability of p_M.
2. Segment mutation – A whole segment of an individual is inverted after the specified cut-off point with a probability of $p_M \cdot (1 - p_{M,D})$, where $p_{M,D}$ is the probability of dot mutation. A similar mutation strategy for a whole bit string is illustrated in Gharahbagh and Abolghasemi (2008), for example.

In this study, the mutation probability evolves deterministically within generations, according to the scheduling rule proposed by Bäck and Schultz (1996). The idea behind the control of mutation probability is to target the high mutation rates at the algorithm's first generations, i.e. the global search phase, where the region of the optimal (or a very good) solution exists. In later generations, the mutation probability decreases to reduce the stochastic changes in the population during the proceeding of generations. A similar reasoning for controlling the mutation effect is proposed in the context of real-coded genetic algorithms by Michalewicz (1992).

3.6 Model parameter identification

The estimation of model parameters is often referred to as model training or model parameter identification, and is conducted via optimisation by minimising or maximising a suitable objective function in a continuous space. Model training can be expressed as a form of optimisation problem, in which the squared prediction residual is usually minimised. The objective functions used in this study are introduced within the model derivation. A suitable solution strategy is determined by the model's characteristics. As has been seen, in multivariable linear regression or other linear models, the parameters can be estimated analytically with the pseudoinversion given on page 45. The requirement for this is that the parameter identification problem can be expressed in a closed-form solution, which demands a linear model structure (Harrell 2015). In non-linear and dynamic models, a suitable iterative procedure is needed to determine the parameters.

The iterative model parameter identification or training strategies – in other words, the optimisation algorithms – can be categorised in gradient-based and gradient-free optimisation. Gradient-based optimisation algorithms obviously

require the computation of the gradient of the objective function to search for the optimum. However, the main problem of the gradient-based search in the context of parameter identification is that the functions are usually multimodal (i.e. have multiple local optima), and gradient-based methods therefore tend to get stuck. However, some algorithms deal well with the function valleys. For example, the Levenberg-Marquardt algorithm uses an adaptive damping parameter (λ) (Hagan & Menhaj 1994), which both helps the algorithm to jump over the function valleys and to converge with the bottom of the valley. The multimodality itself arises from measurement errors (or noise), both in the variables to be predicted and explanatory variables, and in determining the problem. To tackle this, gradient-free search strategies are usually suggested (Katare et al. 2004; Nyarko & Scitovski 2004; Khalik et al. 2007). In this study, neural networks are trained using analytical and gradient-based optimisation, whereas parameterised models are trained using gradient-free strategies, mainly genetic algorithms. More details on the training strategies are given in Publications I, IV, and Section 5.

3.6.1 Real-coded genetic algorithm in parameter identification

Compared to a binary-coded genetic algorithm, its real-coded invariant provides a more reliable and precise alternative for a parameter identification algorithm. Whereas operators of the binary-coded algorithm are defined in the discrete space, operators in the real-coded algorithm define them in the continuous space, which allows an increased precision of solutions near the minima. The real-coded genetic algorithm therefore does not suffer from the precision problems commonly attributed to the binary-encoded population when used in a continuous space, also referred to as Hamming's cliff (Deep et al. 2009). Otherwise, the basic idea of the real-coded genetic algorithm is similar to that of the binary-coded version, but the implementation of the mating operators differs.

Selection

The genetic algorithm used in this study implements two different selection strategies, i.e. tournament selection (TS) and roulette wheel selection (RWS), as well as the combination of both. The tournament selection for k individuals is implemented as follows (Goldberg et al. 1989):

1. generate a random permutation vector of size n_{pop};

2. sort the population according to the permutation vector;
3. define a scale of k individuals that is moved across the population;
4. select the winner of the tournament as the fittest in the scale of k individuals;
5. repeat until the pre-specified number of individuals is selected.

The roulette wheel selection is implemented as follows:

1. calculate the cumulative probability distribution based on the fitness likelihoods of the population;
2. draw a random number between 0 and 1;
3. the index which corresponds to the drawn random value in the cumulative probability distribution is selected.

Arumugam et al. (2005) offer convincing results on the effect of hybrid selection strategy on the convergence of the algorithm in the scheduling problem, which encourages the use of a hybrid selection strategy. Consequently, in this study the real-coded genetic algorithm selects 50% of the individuals with tournament selection and 50% with roulette wheel selection (Arumugam et al. 2005).

Crossover

For the crossover operator, the hybrid approach proposed by Arumugam et al. (2005) is used. The hybrid crossover combines two different crossover operators and treats a predefined part of the population with a specified operator. This study uses a hybrid of the arithmetic convex crossovers. The arithmetic crossover is defined as (Arumugam et al. 2005)

$$\gamma_{OC,1} = \alpha\gamma_{P,1} + (1-\alpha)\gamma_{P,2}, \tag{18}$$

and

$$\gamma_{OC,2} = \alpha\gamma_{P,2} + (1-\alpha)\gamma_{P,1}, \tag{19}$$

where α is a uniformly distributed random number between -0.25 and 1.25. It is common that α is defined between 0 and 1. However, the expansion of the range allows the exploration of a wider region. It should be noted that an excessively wide range for α would probably result in the divergence of the population from the minima, because the changes made in the individuals would be large. As an alternative crossover method, the average convex crossover is used. The average convex crossover is like an arithmetic crossover, expect that the crossover

parameter is set as $\alpha = 0.5$ (Arumugam et al. 2005). The individuals are treated with the crossover operators so that approximately 50% of the individuals are treated with the arithmetic crossover, and the rest with the average convex crossover.

Mutation

Mutation is implemented for the real-coded genetic algorithm with two strategies, the Mäkinen-Periaux-Toivanen mutation (MPT) (Mäkinen et al. 1998) and non-uniform mutation, proposed originally by Michalewicz (1992). The Mäkinen-Periaux-Toivanen mutation is described in more detail in Publication IV, where it is used to mutate the real-coded part of the individuals. The non-uniform mutation is defined by Michalewicz (1992) as

$$\gamma'_{OC} = \begin{cases} \gamma_{OC} + f(i,\ u - \gamma_{OC}), & if\ r < 0.5 \\ \gamma_{OC} - f(i,\ \gamma_{OC} - l), & if\ r \geq 0.5 \end{cases}, \tag{20}$$

where γ'_{OC} is a mutated individual, γ_{OC} is an individual, i is the current generation, u and l correspond to the upper and lower bounds of the mutated individual, and r is a randomly drawn number $r = \{0, 1\}$. f is a function that defines the mutation range as a function of the current generation. The function f is defined as (Michalewicz 1992)

$$f(i, dx) = dx\left(1 - r^{\left(1-\frac{t}{T}\right)^b}\right), \tag{21}$$

where dx is the mutation function. The algorithm for parameterisation is implemented so that approximately 50% of the mutated individuals are treated with either MPT or non-uniform mutation. The reasoning behind such an implementation is that non-uniform mutation shifts to a local search in the latter generations, of which dependency is controlled with parameter b. In this study, $b = 2$. However, MPT mutation does not depend on generation, and it thus allows the algorithm to use both globally and locally efficient mutation operators, even in latter generations.

Elitism

Stochastic optimisation strategies tend to suffer from a loss of information due to an unpredictable event during the search. This issue can be overcome with the elitism operator, which allows a monotonic decrease of the objective function value. The elitism operator is implemented such that the globally best individual is preserved in the population during the iterations, preventing the loss of information.

Population shifting

Population shifting towards the elite is proposed as the convergence acceleration method by Wong et al. (1997). The original idea in shifting relies on the assumption that the elite of the population is correctly directed in relation to the optima. The approach is extended by Arumugam et al. (2004). In their study, population shifting is carried out as

$$\gamma'_{OC} = \gamma_{OC,gb} + r\left(\gamma_{OC,gb} - \gamma_{OC}\right), \tag{22}$$

where $\gamma_{OC,gb}$ is the elite of the population, and r is a random number defined between $r = \{-1\ 1\}$. Shifting towards the elite is not conducted for each individual, but approximately 10% of them, i.e. the probability for shifting is $p_S = 0.1$.

3.7 Model performance assessment and validation

Evaluating model performance is a crucial step in model selection, because the step gathers evidence on the model's applicability. Model validation techniques differ concerning whether the model is used for explanation or prediction. For a predictive model, generalisability is an important factor, because it is defined as model performance in independent test data (Hastie et al. 2009). In validating an explanatory model, it is of interest to consider whether the model explains the stated hypotheses, and how well it fits the collected observations. These can be answered by diagnosing residuals and goodness-of-fit tests. Standard hypothesis tests are also often used in explanatory model validation. It is important to point out that the goodness of an explanatory model is not a guarantee of its good predictive performance (Shmueli 2010).

3.7.1 Hypothesis testing for linear models

In the explanatory analysis of the data, multivariable linear regression models are used as a basis for hypothesis testing. The hypothesis tests determine whether the variable of interest has an observable and reliable effect on the output variable. It is well known that hypothesis tests are limited to the assumptions considering the distributions of the input and output variables, regression coefficients, and the prediction residuals. The formulation for the hypothesis considering the significance of the regression coefficients is (Harrell 2015)

H_0: the estimated effect of the variable is not significant, i.e. b_j does not differ significantly from zero, and

H_α: the estimated effect of the variable is significant, i.e. b_j differs significantly from zero.

The hypothesis testing for the regression coefficient is conducted using a *t*-test. The *t*-test statistic for a regression coefficient of an input variable *j* can be calculated by equation (Hastie et al. 2009)

$$t_j = \frac{\hat{b}_j - b_j}{\mathrm{SE}(\hat{b}_j)}, \tag{23}$$

where $\hat{b}_j$ is the least-squares estimate of the regression coefficients, b_j is the expected regression coefficient, and $\mathrm{SE}(\hat{b}_j)$ is the standard error associated with the regression coefficient $\hat{b}_j$. The computation of the standard error can be conducted with a variance-covariance matrix, which is given as (Hastie et al. 2009)

$$C = \sigma^2 (X^T X)^{-1}, \tag{24}$$

where C is the variance covariance matrix, and σ^2 is the mean squared error estimate for a given degrees of freedom. The standard errors are given by the square root of the diagonal elements of C. σ^2 is thus defined as (Hastie et al. 2009)

$$\sigma^2 = \frac{SSE}{n-k-1} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-k-1}, \tag{25}$$

where SSE is the sum of the squared error, and k is the number of input variables in the model. Hypothesis tests can be problematic if the model residuals have a non-constant variance, i.e. the error terms are heteroscedastic. To reduce the effect of heteroscedasticity, James et al. (2013) suggest a shrinkage of Y with logistic ($\ln(Y)$) or with root transform ($\sqrt{Y}$), which belong to a family of Box-Cox transformations (Box & Cox 1964). Indeed, Box and Cox (1964) propose a logistic transform ($\lambda = 0$) as the primary assumption in modelling chemical rate law dependencies (Box & Cox 1964). To test the validity of the selected Box-Cox transformation, the simple Anderson-Darling test is used to test the normality of variables. The test rejects in a null hypothesis if data does not belong to normal distribution with a 5% significance level. The test is also used to test the normality of the prediction errors in Publication IV. Another issue concerns the correlation of error terms. If the error terms are correlated with the predictors, the estimated standard errors may underestimate the true standard error (James et al. 2013), which results in an increased significance in t-test terms. Similar problems have been outlined by Harrell (2015) in considering the stepwise techniques for variable selection based on hypothesis tests.

3.7.2 Bootstrapping

Another option for determining the distributions of the coefficient estimates is the bootstrapping method. Bootstrapping allows the testing of coefficients without any assumptions of the distribution of the residual or the data in general, and is thus tolerant to heteroscedasticity. In addition, bootstrapping can also be used to define the confidence intervals and the shape of the parameter distributions for non-linear models. For example, Ghosh et al. (2011) use the quantile method and the bias-corrected and accelerated method to determine the distributions of the Richards growth model, the parameters of which are identified with a genetic algorithm (Ghosh et al. 2011), whereas Loibel et al. (2006) use a similar procedure, relying, however, on the Box-Cox transformation of the data to obtain the logistic form of the growth model. The computation of the bootstrapped distribution of a random variable, or in this case for a regression coefficient, can be presented with the following generalised steps (Efron & Tibshirani 1993):

1. draw n samples from the data with n rows with replacement;
2. compute the estimate for $\hat{b}^*_{\mathrm{BS}}$ for the drawn sample;
3. repeat steps 1 and 2 N times.

The SE for the estimate $\hat{b}_j$ can then be defined by computing the standard deviation over the sample distribution. The confidence intervals can be approximated, for example, by making use of the quantile method, in which the confidence intervals are defined simply with the qth quantiles of the observed distribution, although more sophisticated methods also exist (Efron & Tibshirani 1993). The significance of the regression coefficient can be evaluated with the bootstrap p-value. According to Davison and Hinkley (1997), the bootstrap p-value can be computed as the probability of bootstrap sample observations differing from the test-statistic (Davidson & Hinkley 1997). Keeping this and the proposed hypotheses in mind, the equal-tail bootstrapped p-value can be approximated as (MacKinnon 2009)

$$p_{\mathrm{BS}} = 2 \min \left(\frac{\sum_{i=1}^{N} I[\hat{b}_{\mathrm{BS}}^{*} > \bar{b}]}{N}, \frac{\sum_{i=1}^{N} I[\hat{b}_{\mathrm{BS}}^{*} \leq \bar{b}\,]}{N} \right), \tag{26}$$

where p_{BS} is the bootstrap p-value, N is the number of repetitions, I is the logical function returning 1 for a satisfied condition, and $\bar{b}$ is the test-statistic value, which in this case is 0, because in H_0 it is proposed that $\hat{b}_j = 0$. It is arguable how many repetitions are needed to generate reliable bootstrap estimates. For example, Efron and Tibshirani (1993) suggest 200 repetitions for standard error estimation, and Davidson and Hinkley (1997) propose that the number of bootstrap samples should be at least 100 and up to 1,000. Harrell (2015) suggests using as many replicates as needed for the convergence of the distribution of the estimates. However, it should be kept in mind that if bootstrapping is used in the stepwise variable selection stage for inferences, the number of resamples increases the computational complexity of the algorithm by a factor of N_{BS}. It should be noted that bootstrapping can also be used for prediction error estimation (Harrell 2015), but this case is not considered in this study.

3.7.3 Estimation of model performance with cross-validation

Cross-validation is a popular technique to estimate the prediction error of a model (Hastie et al. 2009). The general idea of cross-validation is simple. First, the data is divided into training and validation sets. The training set is used for model selection, and the validation set is used to estimate the prediction error. The division is carried out by sampling without replacement (Hastie et al. 2009). Cross-validation makes no assumptions concerning the linearity of the model parameters

or normality of error distribution, for example (Shao 1993). However, as the validation data is sampled without replacement, the coefficient estimates are less precise than in the case of estimates drawn from the full data (Harrell 2015). This is because the training error tends to be less than the true error, meaning that the training error is too optimistic an estimate of model performance (Hastie et al. 2009). A very common error metric for model selection is the sum of the squared error (SSE) and the mean squared error (MSE). There are many variants of cross-validation, for example, hold-out cross-validation, *k*-fold-cross-validation, leave-one-out (LOO) cross-validation ($k = 1$), and repeated random subsampling cross-validation, i.e. Monte Carlo cross-validation or leave-multiple-out (LMO) cross-validation (Baumann 2003). The key difference between the methods is that in *k*-fold-cross-validation, each of the observations is used in training and testing at least once (Hastie et al. 2009), whereas in repeated random subsampling, this is not obviously guaranteed. For unbalanced data, each of these can be carried out with intentional balancing between class labels, which is referred to as stratified cross-validation (Kohavi 1995). The other computational details for these are given, for example, in Hastie et al. (2017), Baumann (2003), and Kohavi (1995). Of these, this study uses the hold-out method and repeated random subsampling and their stratified variants. In Baumann's study (2003), leave-one-out cross-validation is outperformed by leave-multiple-out (LMO) cross-validation, provided the number of data splits is adequate. If the cross-validation procedure constitutes multiple data splits, the error estimate is drawn as the mean of the splits. To exemplify this, if the SSE is used as the error metric, the cross-validation error is given as

$$\mathrm{SSE} = \frac{1}{k}\sum_{i=1}^{k}(y_i - \hat{y}_i)^2\,, \tag{27}$$

where k is now the number of splits. Cross-validation also allows the use of other metrics such as the coefficient of determination (R^2), mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE). If random subsampling is used, Baumann (2003) recommends the use of multiple splits for the evaluation, for example, $k = 2N$ splits, where N is the number of data rows in the training data. The adequate number still depends on the dataset size. This is because the training and validation errors are not independent of the data subset, especially for small and sparse datasets (Baumann 2003). Considering the practicality of the model selection, the computational complexity limits the

reasonable number of splits, and it is intended to find a trade-off between reliability and computational complexity. If cross-validation is to be used as the objective function in the model selection, a double cross-validation procedure is recommended (Baumann & Baumann 2014). In model selection, double cross-validation proceeds as follows:

1. divide the data into training and external validation datasets;
2. divide the training set into training and internal validation sets;
3. estimate the model performance based on N splits in the internal validation loop;
4. select the candidate model with the best performance in the internal validation loop;
5. assess the prediction performance of the model with the external validation set.

According to Baumann and Baumann (2014), model performance cannot be assessed only with the internal validation dataset, because it is affected by model bias. Model selection bias can be avoided by using the external validation set (Baumann & Baumann 2014). Publications III and IV use the double cross-validation procedure as the default, with a single data split in the outer loop. In Publication III, $4N$ repetitions in the internal validation loop are used, whereas in Publication IV, the number of splits repeated is reflected in the uncertainty of the mean. Double cross-validation is a rarely referenced term, but the procedure is widely used in model selection. However, the internal validation set is often referred to as the *validation set*, and the external validation set is referred to as the *test set* (Hastie et al. 2009). In model training (or model parameter identification), control of model variance is often based on the validation set error; the *validation error estimate* is computed within iterations, and the training is stopped as soon as the validation error increases. In the context of neural networks, this is sometimes referred to as the early stopping method. Hagan et al. (1997) propose that Bayesian regularisation provides a somewhat equivalent result to early stopping, without the need for data splits (Hagan et al. 1997). In data splitting, the size of the data subsets needs to be considered carefully, because the extraction of the validation data subsets steals information from the training set, and too small a validation may result in the test set having an over-optimistic model performance. Typically, 50–70% are used for training, and 15–25% are the validation sets. However, the proper subset sizes depend at least on the data's signal-to-noise ratio, its functional form, the complexity of the model being fitted to the data, and the number of available observations (Hastie et al. 2009).

3.7.4 Information criteria

As outlined in Section 3.2., several information criteria are available for the selection of models from the candidate models. Of these, Mallows's C_p and small-sample corrected AIC (AIC_C) were used. These are given as (Mallows 2000; Hurvich & Tsai 1989; Burnham & Anderson 2002)

$$C_P = \frac{SSE}{\hat{\sigma}^2} - n + 2p, \tag{28}$$

and

$$\mathrm{AIC_C} = \mathrm{AIC} + \frac{2(p+1)(p+2)}{n-p-2}, \tag{29}$$

where $\hat{\sigma}^2$ is estimated with the mean squared error for the full model, p is the number of model parameters, and n is the number of observations. For a model of which parameters are estimated by means of least squares, the AIC is defined as (Burnhan and Anderson 2002)

$$\mathrm{AIC} = n \ln \hat{\sigma}^2 + 2p. \tag{30}$$

For small datasets, Burnham and Anderson (2002), Hurvich and Tsai (1989), and Hurvich and Tsai (1995), for example, recommend using the AIC_C instead of C_P and AIC for small datasets, because this adds an additional term for the criterion, penalising from the complexity, thus trading some model variance to bias, yielding as more parsimonious models. According to Burnham and Anderson (2002), the interpretation of AIC_C cannot be taken as an absolute value, but it must be compared to other model candidates in the set. For this purpose, the authors suggest the following metric (Burnham & Anderson 2002):

$$\Delta_i = \mathrm{AIC}_{C,i} - \mathrm{AIC}_{C,\min}, \tag{31}$$

where Δ_i is the difference in AIC_C values between the model i and the smallest AIC_C of all candidate models ($AIC_{C,min}$). The authors also suggest using 2 as the threshold for significant model candidates (Burnham & Anderson 2002). However, determining the best model from the set of model candidates is not a trivial task. As outlined in Section 3, the selection of the best model can be guaranteed only

with an optimal search. In this case, an exhaustive search was used. The procedure for the search was as follows:

1. evaluate the AIC_C for all the possible models;
2. find the model with the minimum AIC_C;
3. calculate the Δ_i for all the models;
4. select the final model candidates as the ones with $\Delta_i < 2$.

4 Datasets and pre-processing

In this study, two different datasets are available to model the hot metal desulphurisation process. The considered datasets are gathered from SSAB Raahe. In addition, a dataset used for the predictive modelling of sulphide capacity and the sulphur partition ratio is extracted from the literature. The characteristics and pre-processing of data is outlined in the following sections.

4.1 Primary desulphurisation dataset (Publications I and III)

The first dataset is used to study primary hot metal desulphurisation. The data is mainly used in Publications I and III. The first pre-processing step considers the outlier detection and treatment. Outlier detection methods are used to exclude both contextual and global outliers from the data. For global outlier detection, the assumption of normal distribution is used as an exclusion criterion, in which if a datapoint is more than ±3σ from the basic distribution, it is excluded from the data. Consequently, 6 treatments are removed. For example, blast furnace cold runs are excluded, as they are both contextual and global outliers with respect to the normal function of the process. In addition, cold runs are rather rare, and the aim of this study is only to model a characteristic behavior of the process. After pre-processing, the dataset constitutes 39 observations and more than 20 variable candidates. The full datasets constitute the following measurements:

1. composition of the metal before and after the treatment;
2. composition of the slag before and after the treatment;
3. temperature of the hot metal before and after the treatment;
4. particle size distributions of the applied reagents;
5. injection parameters.

The average hot metal composition at the primary hot metal desulphurisation site before the treatment is C = 4.5 wt-%, Si = 0.45 wt-%, S = 0.045 wt-%, and Mn = 0.172 wt-%, whereas the average temperature is around 1,623 K (1,350 °C). The average mass of the hot metal is around 80 t. The composition of the slag before the injection is on average CaO = 39.28 wt-%, SiO_2 = 38.82 wt-%, Na_2O = 13.66 wt-%, Al_2O_3 = 5.78 wt-%, and MgO = 1.63 wt-%. After the injection, the CaO content can be as high as 78–85 wt-%, ignoring the metal droplets entrapped in the slag phase.

The particle size distributions of the reagents are defined on a volume basis with the laser diffraction method. The corresponding values for the measured distributions used in the identification of the models are given in Publications I and III. The values for the characteristic particle sizes (d_{10}, d_{25}, d_{50}, d_{80}, d_{90}) are extracted from Vuolio (2019), except for reagent E, whose characteristic particle size distribution is determined later as an average of several measurements. Reagent D is a commercially produced reagent, and the others are produced onsite. The sampling of the reagents onsite is carried out batchwise, by taking several samples of a single reagent batch of 10–12 tonnes. The sampling of reagent E is similar to reagents A–D. During the manufacture of reagents A, B, and E, varying amounts of $CaCO_3$ wt-% are mixed in the CaO. The particle size distribution of the $CaCO_3$ measured and its effect are mathematically extracted from the distribution. However, the extraction of $CaCO_3$ from the reagent mix is found to be negligible for the characteristic particle size distribution values, because it changes the values to an order of 0.5–5 μm. In addition, the dataset is enriched with variable construction by adding computational variables to the set. Depending on the model's characteristics, variables are constructed statically or dynamically. Dynamic variable construction is used in the case of parameterised models, of which the sulphur trajectory is solved numerically (Section 5.3.), whereas static construction is used in cases where the analytic solution of the end sulphur content or the static output variables are available.

Prior to modelling, the missing values are treated with regression and mean imputation. The imputation mainly considers the compositions of the slag, which are not available for every treatment. It has been shown that imputation techniques perform better than complete case analysis, which means that only the observations with no missing values in the explanatory data are included in modelling. A complete case analysis may cause significant loss of information and even lead to falsely interpreted results (Stavseth et al. 2019). The following procedure is used for regression imputation:

1. search a functional relation within the data;
2. fit a model to the data;
3. predict the missing values based on the regression model;
4. impute the values to the data.

Mean imputation corresponds simply to imputing mean values to the data. During the analysis of the slag compositions, a functional relation is found with the

temperature and (CaO) wt-%, presented in Figure 3. The corresponding linear regression model used to impute the values is (CaO) wt-% = 0.3878T (◦C) – 501.3.

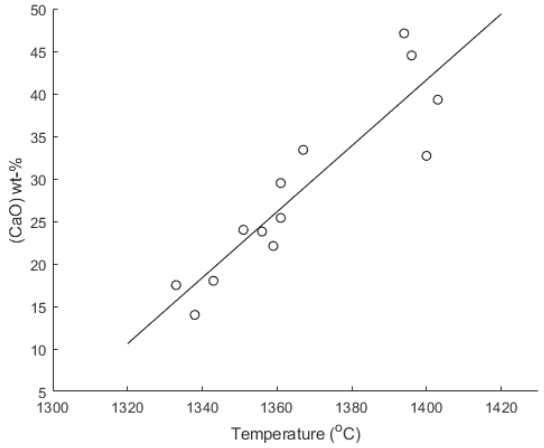

Fig. 3. Relationship of the temperature and CaO content of the slag.

4.2 Secondary desulphurisation dataset (Publication IV)

The dataset considers secondary desulphurisation and is used in the prediction model identification in Publication IV, as well as in Section 5.6. The experimental data is gathered from the secondary hot metal desulphurisation process at SSAB Europe Oy in Raahe, Finland. The dataset is rather similar to that collected from primary desulphurisation with respect to available attributes. However, this dataset constitutes only production data, and it therefore omits the composition of the slag and particle size distributions from the variable candidates. The data that is treated with variable construction consists of 551 treatments and 23 candidate variables. In addition, to test the performance of the model selection algorithm, 10 to 20 columns of variables constituting only white noise $N(0,1)$ are added to the data. However, the experiments carried out with the white noise imputed datasets are not reported in Publication IV, but only in this study.

4.3 Sulphide capacity dataset (extended from Publication II)

As outlined in Publication II, a plethora of sulphide capacity models is available in the literature, but the most referenced models with a simple model structure usually lack data on CaO-SiO_2-Na_2O slag systems. Dataset 3 is therefore collected to cover this slag system and other systems. The referenced studies are Chan and Fruehan (1986), Chan and Fruehan (1989), Fetters and Chipman 1941, Grant et al. (1951), Hatch and Chipman (1949), Inoue and Suito (1982), Taylor and Stobo (1954), Tsao and Katayama (1986), Venkatradi and Bell (1969), Van Niekerk and Dippenaar (1993), and Winkler and Chipman (1946). The referenced studies cover the slag systems with CaO, SiO_2, Al_2O_3, MgO, Na_2O, FeO, and MnO, and the temperature ranges from 1200 to 1700 °C. However, it should be noted that the dataset itself is multivariate and sparse, which could affect the model's reliability and generalisability. However, the objective is not to gather a comprehensive dataset, but a relatively representative one for identification studies. The data consists of sulphur partition ratios determined with the slag-metal-equilibrium technique. In this technique, the slag and metal phases are equilibrated, and the sulphur content in both phases is measured. Based on these measurements, the slag's sulphide capacity is computed. A detailed description of this method can be found in the referenced studies. The observed partition ratio of sulphur and metal between the slag and metal phases can be converted to sulphide capacity with the equilibrium constant derived for slag-gas and slag-metal desulphurisation reactions. The following expression for the sulphur partition ratio can be derived from the details given in Young et al. (1992) as

$$\log_{10} L_S = \log_{10} C_S - \log_{10} a_{[O]}^{H} - \frac{935}{T} + 1.375 + \log_{10} f_{[S]}^{H}, \qquad (32)$$

where $f_{[S]}^{H}$ is the Henrian activity coefficient of sulphur in the metal phase. As seen in the previous equation, the sulphide capacity is a property that seeks to capture the slag's sulphur extraction capacity, independently of the activity of oxygen prevailing in the system. The slag's sulphide capacity determines the thermodynamic driving force for the slag-metal reaction. To predict the sulphide capacity of slag, a description of the activity of oxygen in the equilibrium state is needed (Oeters 1994). The activity of oxygen can be measured or estimated from the expression of the equilibrium constants for the governing oxidation reactions. In this study, the activity of oxygen is assumed to be defined by one of the reactions

$$\mathrm{Fe} + [\mathrm{O}] \rightarrow (\mathrm{FeO}), \tag{33}$$

$$[\mathrm{C}] + [\mathrm{O}] \rightarrow \{\mathrm{CO}\}, \tag{34}$$

and

$$[\mathrm{Si}] + 2[\mathrm{O}] \rightarrow (\mathrm{SiO_2}). \tag{35}$$

On the basis of the reactions above, and by assuming that the $p_{CO} = 1$ and $a^R_{Fe} = 1$, the following expressions can be derived for the equilibrium activity of oxygen based on the Henrian standard state:

$$\log_{10} \mathrm{a}^{\mathrm{H}}_{[\mathrm{O}]} = -\log_{10} K_{\{\mathrm{CO}\}} - \log_{10}\left(f^{\mathrm{H}}_{[\mathrm{C}]}[\%\mathrm{C}]\right), \tag{36}$$

$$\log_{10} \mathrm{a}^{\mathrm{H}}_{[\mathrm{O}]} = -\frac{1}{2}\log_{10} K_{(\mathrm{SiO_2})} - \frac{1}{2}\log_{10}\left(f^{\mathrm{H}}_{[\mathrm{Si}]}[\%\mathrm{Si}]\right) + \frac{1}{2}\log_{10}\left(\gamma^R_{\mathrm{SiO_2}} x_{\mathrm{SiO_2}}\right), \tag{37}$$

and

$$\log_{10} \mathrm{a}^{\mathrm{H}}_{[\mathrm{O}]} = -\log_{10} K_{(\mathrm{FeO})} + \frac{1}{2}\log_{10}\left(\gamma^R_{\mathrm{FeO}} x_{(\mathrm{FeO})}\right), \tag{38}$$

where γ^R_{i} is the Raoultian activity coefficient for component i in the slag phase, x_{SiO_2} is the molar fraction of component i in the slag phase, and $f^{\mathrm{H}}_{[\mathrm{i}]}$ is the Henrian standard state activity coefficient for [i] in the metal phase. The equilibrium constants for the oxidation of carbon and silicon are extracted from Pal and Patil (1986) and Görnerup and Sjöberg (1999) respectively. The equilibrium constant for the oxidation of iron to iron(II)oxide is computed based on the change of Gibb's free energy ($\Delta G°$) reported in Basu et al. (2008). The value on the basis of which the sulphide capacity is calculated is chosen as the minimum value of the oxygen activities. The activity coefficient for (SiO_2) is calculated with the extended regular solution model proposed by Ban-Ya (1993). In this model, the activity coefficient following a Raoultian standard state is given as (Ban-Ya 1993)

$$RT \ln \gamma_i = \sum_j \alpha_{ij} X_j^2 + \sum_j \sum_k (\alpha_{ij} + \alpha_{ik} - \alpha_{jk}) X_j X_k + I^{\circ}_{C.}, \tag{39}$$

where α_{ij} is the interaction energy between two corresponding cations (i and j), X_i is the cation fraction of i, and $I^{\circ}_{C.}$ is the conversion factor of the activity coefficient

between a hypothetical regular solution and a real solution (J/mol). The interaction energies for the slag components are extracted from (Ban-Ya 1993). The activity coefficient for (FeO) is estimated with the empirical model proposed by Basu et al. (2008). Similarly, the activity coefficients for the metal phase are calculated with the WLE formalism for dilute solutions presented in Sigworth and Elliot (1974) by using equation

$$\log_{10} f_i^H = \sum_{j=1}^{n_j} e_i^j [\% j], \tag{40}$$

where e_i^j is the mass-based interaction coefficient between component i and j. The values for the interaction coefficients are extracted from the study of Sigworth and Elliot (1974).

5 Results and discussion

In this section, the results presented in Publications I–V are presented with yet unpublished results considering the parameterised models and the modelling of sulphide capacity. The sections go through the analysis of the dataset, using both generic and parameterised reaction models. In addition, some explanatory modelling results not included in the publications are introduced.

5.1 Explanatory analysis of Dataset 1

The objective of the explanatory analysis of the data is divided into two subsections:

– In the first section, the different particle size distribution parameters and their effects are estimated on a linear and logarithmic model basis.
– In the second subsection, the dependency of different particle size distribution parameters on the rate of desulphurisation is analysed in more detail.

5.1.1 Analysis with the linear model basis

Prior to the model structure selection for prediction models, an explanatory analysis was carried out for Dataset 1. The explanatory analysis was made to support the identification of the presented models, both in terms of model parameter identification and model structure selection. In this analysis, two output variables, namely the molar efficiency of the desulphurisation reagent and the time constant for the overall reaction, were the dependent variables of interest. The definition for the assumed time constant can be found in Publication I, whereas the molar efficiency of the reagent is defined as the molar ratio of the removed sulphur and the injected reagent. A similar metric to describe the efficiency of desulphurisation reagents can be found in the study of Lindström and Sichen (2015), although the authors use the inverse of the metric, along with many others not referenced here.

For explanatory purposes, the predictive properties of the models are not accounted for in this section. It should be noted that the models identified in this section can only consider the magnitude and sign of the effects of individual variables in an explanatory manner. A more detailed analysis of the mechanisms, as well as their contributions, demands more sophisticated model structures, examples of which Publication I and Sections 5.2. and 5.3. in this thesis provide.

In the explanatory model selection, manual selection, an exhaustive search with information theoretic criteria, and Forward Selection 3 are used. Referring to the guidelines given by various authors (e.g. Burnham & Anderson 2002; Heinze et al. 2018), manual selection is considered the selection basis prior to applying any search strategies.

A comprehensive description of the basis and execution of the manual selection is discussed in detail in Publications I and V. In principle, the manual selection of variables is conducted using the literature survey, considering the process phenomena to guide the selection. The literature survey and simulations are elaborated in detail in Publications I and V. With reference to Publications I and V, the relevant attributes are particle size distribution, injection gas rate, the injected quantity of gas-forming compounds, the mass flowrate, the mass of the hot metal, and the Henrian standard state activity of sulphur. To analyse the significance of regression coefficients, the *t*-test with normality assumption is used, and the level of significance is set as $\alpha = 0.05$. The analysis is conducted for both datasets with and without the slag components, using MLR as the model basis.

In the search of the explanatory model, *Forward Selection 3* is used. The convergence graph for *Forward Selection 3* used in model selection for explaining molar efficiency is given in Figure 4. The figure shows that the selection method sequentially improves model performance in terms of root mean squared error and the adjusted coefficient of determination. The figure on the right-hand side presents Mallows's C_p and the small sample corrected AIC_C values for the models considered during the search.

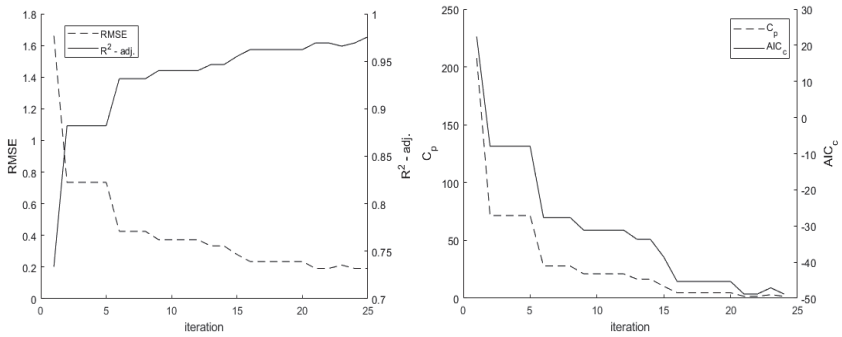

Fig. 4. Convergence of Forward Selection 3 to select a model for the stoichiometric efficiency of the reagent for the entire dataset, with imputation of the slag components without a transform of the input and output variables.

As Figure 4 shows, stepwise selection tends to converge towards smaller values for both metrics, i.e. to models with a small error but relatively small complexity. To evaluate the reliability of the stepwise search, the procedure using exhaustive search and information criteria introduced previously is applied for comparison. The results of the exhaustive search and information criteria-based analysis are given in Table 2. The table shows that the particle size distribution parameters d_{80}, Q_{tot}, $m_{\text{Fe,}}$ and $\dot{m}_r$ are the variables of interest related to the transitory contact reaction. In addition, some of the metal and slag phase components appear in the set of models with a similar performance.

Table 2. Selected sets of variables with exhaustive search and setting $\Delta_i < 2$ for n_{eff} (%) and k_{tot}.

y	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
n_{eff} (%)	$\dot{m}_r$	Q_{tot}*	d_{80}	m_{Fe}	[Si]	$a^{H}_{[S]}$	(MnO)	-	-
	$\dot{m}_r$	Q_{tot}*	d_{80}	m_{Fe}	[Si]	$a^{H}_{[S]}$	(MgO)	(MnO)	-
	$\dot{m}_r$	Q_{tot}*	d_{80}	m_{Fe}	[C]	[Si]	$a^{H}_{[S]}$	(MgO)	(MnO)
	$\dot{m}_r$	Q_{tot}*	d_{80}	m_{Fe}	[Si]	$a^{H}_{[S]}$	(S)	(MgO)	(MnO)
k_{tot} (1/min)	$\dot{m}_r$	Q_{tot}*	d_{80}	m_{Fe}	-	-	-	-	-
	$\dot{m}_r$	Q_{tot}*	d_{80}	m_{Fe}	[C]	-	-	-	-
	$\dot{m}_r$	Q_{tot}*	d_{80}	m_{Fe}	[P]	-	-	-	-
	$\dot{m}_r$	Q_{tot}*	d_{80}	m_{Fe}	[V]	-	-	-	-
	$\dot{m}_r$	Q_{tot}*	d_{80}	m_{Fe}	[Cr]	-	-	-	-
	$\dot{m}_r$	Q_{tot}*	d_{80}	m_{Fe}	$a^{H}_{[S]}$	-	-	-	-
	$\dot{m}_r$	Q_{tot}*	d_{80}	m_{Fe}	[P]	$a^{H}_{[S]}$	-	-	-
	$\dot{m}_r$	Q_{tot}*	d_{80}	m_{Fe}	[P]	(MnO)	-	-	-
	$\dot{m}_r$	Q_{tot}*	d_{80}	m_{Fe}	[Ti]	$a^{H}_{[S]}$	-	-	-
	$\dot{m}_r$	Q_{tot}*	d_{80}	m_{Fe}	[V]	$a^{H}_{[S]}$	-	-	-
	$\dot{m}_r$	Q_{tot}*	d_{80}	m_{Fe}	[Cr]	$a^{H}_{[S]}$	-	-	-
	$\dot{m}_r$	Q_{tot}*	d_{80}	m_{Fe}	[Cr]	MnO	-	-	-
	$\dot{m}_r$	Q_{tot}*	d_{80}	m_{Fe}	[Mn]	$a^{H}_{[S]}$	-	-	-
	$\dot{m}_r$	Q_{tot}*	d_{80}	m_{Fe}	$a^{H}_{[S]}$	B_2	-	-	-

Notes: *) Q_{tot} (m^3/s) is the total volume of gas in the conditions at the lance tip, accounting for both the injection gas and the gas-forming compounds (assuming 100% decomposition).

The selection results for Forward selection 3 for both the considered output variables are given in Table 3 for the non-transformed case and in Table 4 for the data with logistic transform**.** As the table shows, all the explanatory variables can be considered significant in *t*-test and *p*-value terms, because the associated *p*-value is well below the chosen risk level. It can also be seen that the explanatory performance of all the selected models is relatively good, because the R^2 – adj. varies from 0.85 to 0.98, and the RMSE value is relatively small. The variable subsets differ slightly depending on the output variable to be explained, but are still in high agreement, for example, with the results of Chiang et al. (1990), Coudure and Irons (1994), Lindström and Sichen (2015), and Lindström et al. (2015). In addition, the logistic transform of the input and output data seems to eliminate the compositional variables related either to hot metal or slag composition from the selected subset. The reasoning behind this could be either that the compositional variables lack an observable effect on the output variables as the covariance between the reagent and injection properties is relatively strong. A less likely

explanation would be that the compositional variables lack any effect on the reagent efficiency or the rate of reaction measured in terms of the time constant, because it is known, for example, that the activity of silicon may control the oxygen potential in the system and cause the formation of $CaO{\cdot}SiO_2$ (Oeters 1994). Interestingly, the temperature effect seems not to be significant. However, this is probably related to the fact that the considered volumetric gas flowrate depends on the temperature and that the variance of the temperature is relatively small. However, it is also plausible that as the transitory reaction is controlled by the mass transfer rate, which depends only moderately on the temperature, the temperature seems to lack an observable dependency.

As can be interpreted from the regression coefficient signs and magnitudes, an increase in the reagent and gas flowrates increases the observed rate of desulphurisation, whereas decreasing the particle size d_{80} increases the rate. As indicated in Publication I, these can be attributed to the general properties of mass transfer controlled reactions. For example, a coarser particle size distribution results in a slower reaction rate, because the particle size is strongly related to the solid interfacial area available for mass transfer. Interestingly, the reagent efficiency seems higher with slow particle flowrates ($\dot{m}_r$), which can be further associated with the fact that with lower sulphur content, the reaction rate decreases due to the decrease in the driving force. Increasing the time constant therefore does not result in a significant improvement in reagent efficiency, referring to the expression R_j. Another explanation could be the transition from a coupled to a non-coupled flow scheme, which is further discussed in Irons (1988). A more detailed analysis considering the magnitudes of the effects of the selected variables in the case of the rate of transitory reaction can be found in Publication I.

To summarise, the results of the explanatory analysis determined with three different methods arrive at basically the same conclusion, because the results provided by the algorithmic approach and the domain knowledge are in high agreement. Publication III also presents that the genetic algorithm that uses repeated cross-validation as the objective function produces similar results, although a more intensive tuning of the computational parameters is needed. However, it should be noted that the data of this set does not represent regular production, but an experimental setting, because the main objective is to identify the effects of particle size distribution and other injection parameters on the rate and efficiency of the process. However, this analysis provides at least a relatively strong indicator of the explanatory power of these factors, and based on this, the mentioned factors could induce a substantial amount of variance in the process if

the reagent properties and injection parameters are not controlled appropriately. However, it should be noted that the data and linear models suffer moderately from heteroscedasticity, which directly affects the reliability of the hypothesis tests. The domain knowledge can therefore be considered a very effective model selection method, even though the algorithmic approaches produce quite similar results. In addition, as the linear estimates may also suffer from heteroscedasticity, it would be beneficial to consider a non-linear modelling scheme.

Table 3. The selected models for the molar efficiency of the reagent and the time constant of desulphurisation, using Forward Selection 3 for non-transformed data.

y	Variable	Estimate	SE	*t*-value	*p*-value	$R^2_{\text{adj.}}$	RMSE
n_{eff} (%)	b_0	0.82	1.75	0.47	0.64	0.96	0.54
	$a^{\text{H}}_{[\text{S}]}$	33.36	1.68	19.82	$6.68\cdot10^{-20}$		
	d_{80}	–0.01	$1.84\cdot10^{-3}$	–4.42	$1.00\cdot10^{-4}$		
	$\dot{m}_r$	–0.04	$8.30\cdot10^{-3}$	–5.12	$1.29\cdot10^{-5}$		
	Q_{tot}	0.09	$2.33\cdot10^{-2}$	3.72	$7.39\cdot10^{-4}$		
	m_{Fe}	$5.28\cdot10^{-5}$	$1.67\cdot10^{-5}$	3.16	$3.39\cdot10^{-3}$		
n_{eff} (%)*	b_0	–2.34	1.66	–1.41	0.17	0.98	0.48
	$a^{\text{H}}_{[\text{S}]}$	35.14	1.62	21.65	$7.22\cdot10^{-20}$		
	d_{80}	–0.01	0.00	–3.37	$2.10\cdot10^{-3}$		
	$\dot{m}_r$	–0.05	0.01	–6.91	$1.13\cdot10^{-7}$		
	Q_{tot}	0.09	0.02	4.98	$2.47\cdot10^{-5}$		
	(MnO)	2.89	0.65	4.47	$1.04\cdot10^{-4}$		
	m_{Fe}	$5.43\cdot10^{-5}$	$1.38\cdot10^{-5}$	3.94	$4.50\cdot10^{-4}$		
	[Si]	2.56	1.05	2.43	0.02		
	(MgO)	0.26	0.12	2.11	0.04		
k_{tot} (1/min)	b_0	0.14	0.06	2.15	0.04	0.85	0.02
	d_{80}	$-5.59\cdot10^{-4}$	$6.30\cdot10^{-5}$	–8.87	$2.29\cdot10^{-10}$		
	Q_{tot}	0.01	$8.68\cdot10^{-4}$	6.12	$5.97\text{E}\cdot10^{-7}$		
	m_{Fe}	$-1.48\cdot10^{-6}$	$6.03\cdot10^{-7}$	–2.46	0.02		
	$\dot{m}_r$	$1.60\cdot10^{-3}$	$2.97\cdot10^{-4}$	5.39	$5.40\cdot10^{-6}$		
k_{tot} (1/min)*	b_0	0.14	0.06	2.15	0.04	0.85	0.02
	d_{80}	$-5.59\cdot10^{-4}$	$6.30\cdot10^{-5}$	–8.87	$2.29\cdot10^{-10}$		
	Q_{tot}	$5.31\cdot10^{-3}$	$8.68\cdot10^{-4}$	6.12	$5.97\cdot10^{-7}$		
	m_{Fe}	$-1.48\cdot10^{-6}$	$6.03\cdot10^{-7}$	–2.46	0.02		
	$\dot{m}_r$	$1.60\cdot10^{-3}$	$2.97\cdot10^{-4}$	5.39	$5.40\cdot10^{-6}$		

Notes: *) With the slag components included in the dataset.

Table 4. The selected models for the molar efficiency of the reagent and the time constant of desulphurisation, using Forward Selection 3 for transformed data (λ = 0).

ln (*y*)	ln(Variable)	Estimate	SE	*t*-value	*p*-value	$R^2_{\text{adj.}}$	RMSE
n_{eff} *	b_0	–0.47	2.96	–0.16	0.87	0.95	0.10
	$a^{\text{H}}_{[\text{S}]}$	0.89	0.05	19.61	9.39^{-20}		
	d_{80}	–0.22	0.05	–5.00	$1.81 \cdot 10^{-5}$		
	$\dot{m}_r$	–0.35	0.13	–2.74	0.01		
	Q_{tot}	0.34	0.07	4.83	$3.05 \cdot 10^{-5}$		
	m_{Fe}	0.53	0.24	2.20	0.03		
k_{tot}*	b_0	3.19	3.93	0.81	0.4	0.81	0.14
	d_{80}	–0.49	0.06	–8.43	$7.58 \cdot 10^{-10}$		
	Q_{tot}	0.55	0.09	6.26	$4.04 \cdot 10^{-7}$		
	m_{Fe}	–0.71	0.32	–2.21	0.03		
	$\dot{m}_r$	0.87	0.18	4.93	$2.09 \cdot 10^{-5}$		

Notes: *) With the slag components included in the dataset.

5.1.2 Analysis of the rate with different particle size distribution parameters, using surface area approximation

Single particle models

As Coudure and Irons state (1994), there is often a question of which distribution parameter should be used in modelling mass transfer controlled reactions, in which solid particles are used for extraction (Coudure & Irons 1994). In Publication I's results, only the identification results for the distribution parameter d_{80} are reported. This is mainly because it gives the best results in terms of goodness of fit measures, and the values are most consistent with the values predicted by surface-area approximation. However, various other characteristic values for particle size distributions exist. The values can be calculated from either the raw values of the particle size distribution or by using a fitted mathematical distribution. For example, a commonly used measure for mean diameter in mass transfer processes is the Sauter mean diameter, commonly denoted as d_{32}. The Sauter mean diameter can be calculated based on the number of the particle size distribution as follows (Pacek et al. 1998):

$$d_{32} = \frac{\sum_{i=1}^{k} n_i d_i^3}{\sum_{i=1}^{k} n_i d_i^2}, \tag{41}$$

where n_i is the number of particles in channel i of diameter d. Due to the characteristics of particle size distribution, some intrinsic problems arise in using it to predict the rate of hot metal desulphurisation. The major problem is that in practical desulphurisation reagents, the measured particle size distribution contains an enormous number of very small particles ($d < 1$ μm) assumed to float without a reaction to the surface of the metal bath (Lee & Morita 2004). However, the *Sauter mean diameter*, if calculated for the whole distribution, suggests that the interfacial area for the mass transfer is very high, because the d_{32} usually varies from 3–7 μm, even though it is assumed a relatively large number of particles do not participate in the reactions at all. Assuming that the residence times of the particle are as long as suggested – Oeters (1994), for example, proposes a residence time of 75 s (Oeters 1994), and Ma et al. (2017) suggests 28.6 s – at full scale, the extraction capacity of the overall distribution will be sufficiently large if all the particles are assumed to come into contact the melt. Coudure and Irons (1994) use a value of 0.83 sec at lab scale, and as referenced in Section 2, find that the number of particles that come into contact is in the order of 20–30% for CaC_2. However, Coudure and Irons (1994) use their own metric as the mean, denoted as the equivalent diameter for a diffusion-controlled process (d_{ka}) given by (Coudure & Irons 1994)

$$d_{ka} = \left(\sum_{i=1}^{k} \frac{Vol_i(\%)}{100 d_{p,i}^2} \right)^{-2}, \tag{42}$$

where d_{ka} is the equivalent diameter for the diffusion controlled process, $Vol_i(\%)$ is the volumetric amount of size class i in volume-based particle size distribution, and $d_{p,i}$ is the diameter of size class i. However, with reference to the original publication, the exponent should be -0.5 instead of -2, i.e. the equation comes in the form of

$$d_{ka} = \left(\sum_{i=1}^{k} \frac{Vol_i(\%)}{100 d_{p,i}^2} \right)^{-0.5}, \tag{43}$$

because the equivalence is expressed in terms of the surface-area approximation as (Coudure and Irons 1994)

$$\frac{12\dot{m}_r D_{[S]} L_p}{\rho_r u_p V} \sum_{i=1}^{k} \frac{Vol_i(\%)}{100 d_{p,i}^2} = \frac{12\dot{m}_r D_{[S]} L_p}{\rho_r u_p V d_{ka}^2}, \quad (44)$$

where L_p is the immersion depth of the lance, and u_p is the ascent velocity of the particles in the melt. Nevertheless, the dependency of the particle size and rate constant is found to be significantly smaller than the squared dependency (Coudure & Irons 1994). The calculated particle size distribution parameters for the studied reagents are given in Table 5.

Like the applied particle size distribution parameter, the assumed residence time of the particles affects the relation. The residence time greatly depends on the assumptions drawn from the system. For example, Oeters (1994) suggests a residence time of 39 s for the boundary-layer diffusion-controlled case, and 78 s for the solid-state diffusion-controlled case. However, in the study of Oeters (1994), the mean particle size used in the calculations is presumably quite large, although it is not reported in detail. In the study of Visuri et al. (2019), the measured particle size distribution is used as the identification basis, and a residence time of 2 s is obtained (Visuri et al. 2019). This value resembles what can be computed based on the procedure proposed by Seshadri et al. (1997), who uses the empirical correlation proposed by Sahai and Guthrie (1982) to determine the plume velocity by equation

$$U_{plume} = 4.5 \frac{\left(Q_{N_2}^{0.33} h_{Fe}^{0.25}\right)}{R_{ladle}^{0.33}}, \quad (45)$$

where U_{plume} is the plume velocity, R_l is the radius of the ladle, and Q_{N_2} is the flowrate of the injection gas (N_2) in the conditions corresponding to the lance tip. The gas is assumed to heat up to 80% of the bath temperature. The force balance calculations reveal that the terminal velocities of the particles are very near to velocity of the continuous phase surrounding them. Assuming that the particles reach their terminal velocities very quickly once injected yields a residence time of $1.93 \approx 2$ s for an average treatment. However, this assumption ignores the possibility of recirculating the particles in the melt. Seshadri et al. (1997) propose that the mass transfer coefficient can be estimated with the Sherwood number and the correlation derived from Kolmogoroff's theory of local isotropy. The particle size is used as the characteristic to compute the mass transfer coefficient, based on the Sherwood number (Seshadri et al. 1997). On average, this yields a mass transfer

coefficient of $1.23 \cdot 10^{-4}$ m/s, taken as a mean of the entire particle size distribution for Reagent A. The metal-side mass transfer coefficients are presented for each of the characteristic reagent parameters in Table 5. Of these, the values corresponding to d_{80} are the closest to those proposed by Oeters (1994), who suggests using a residence time as high as 39 s to 78 s, yielding a best fit for a mass transfer coefficient of $6.3 \cdot 10^{-5}$ m/s.

Table 5. Mass transfer coefficient computed with different characteristic particle size distribution values.

Reagent	d_{80}*	d_{ka}**	d_{32}***	β (m/s) (d_{80})	β (m/s) (d_{ka})	β (m/s) (d_{32})
A	135.00	2.73	6.37	$4.56 \cdot 10^{-5}$	0.0023	$9.66 \cdot 10^{-4}$
B	223.60	2.88	6.30	$2.75 \cdot 10^{-5}$	0.0021	$9.77 \cdot 10^{-4}$
C	233.60	2.94	6.80	$2.63 \cdot 10^{-5}$	0.0021	$9.05 \cdot 10^{-4}$
D	73.30	2.08	3.92	$8.39 \cdot 10^{-5}$	0.0030	0.0016
E	69.90	2.55	5.38	$8.81 \cdot 10^{-5}$	0.0024	0.0011

Note: *) Extracted from Publication I. **) Differs from Publication I due to erroneous interpretation of the equation found in Coudure and Irons (1994). β = mass-transfer coefficient. ***) Calculated with equation 42.

When using either d_{32} or the d_{ka} as the identification basis for Dataset 1, it can be seen that the direction of the dependency is similar to that suggested by Coudure and Irons (1994). For example, when using d_{80} and ln-transformation, the corresponding dependency of the particle size is –0.22. For smaller particle sizes, the estimated dependency is obviously slightly larger. However, the parameter itself is only a descriptive relation of the dependency.

To estimate the reaction rate without the corresponding parameter, an estimate of the efficient surface area is needed. As previously stated, the surface area estimated from the nominal particle size distribution plausibly over-estimates this value. For this purpose, *an effective Sauter mean diameter* ($d_{32,\text{eff.}}$) is proposed. The effectiveness is deduced by using the prior information provided by the particle size distribution and the contact probability, based on the *expected value of the particle size distribution* coming into contact with the melt. In the distribution, the expected volume for size class i is stated as the conditional expected value as

$$Vol(\%)_{i,c.} = E\big(Vol(\%)_{i,c.} \,\big|\, Q_{\text{tot,ac.}}, u_c\big), \tag{46}$$

where $Vol(\%)_{i,c.}$ is the volumetric amount of particles in contact with size class i, $E(Vol(\%)_{i,c.} \mid Q_{tot}, u_c)$ is the expected value for volume fraction of particles that come into contact with the melt, $Q_{tot,ac.}$ is the volume of gas that is formed in the

decomposition of the auxiliary compounds, and u_c is the critical velocity calculated from the critical Weber number. To estimate the expected value, the probabilistic law for the volume of particles that come into contact is needed. For this purpose, a logistic risk function is used, in which the probability is that *particles come into contact* given the *volumetric amount of gas generated* due to the decomposition of the gas-forming agents and the critical penetration velocity of the particles. The probabilistic function can be given as a continuous sigmoid function as

$$P(\phi = 1 \mid Q_{\mathrm{tot,ac.}}, u_c) = \frac{1}{1 + \exp\left(-\left(b_0 + b_1 Q_{tot,ac.} + b_2 u_c\right)\right)}. \tag{47}$$

The empirical fitting parameters [b_0, b_1, b_2] define the direction and magnitude of the effect on the conditional contact probability. A more detailed description of the properties of the logistic regression and the sigmoid function is given in Harrell (2015), for example. The expected volume for a particle size distribution in contact with the metal is therefore

$$E\left(Vol(\%)_{i,c.} \left| Q_{\mathrm{tot,ac.}}, u_c\right.\right) = \sum_{i=1}^{k} Vol_i(\%) f(Q_{\mathrm{tot,ac.}}, u_c), \tag{48}$$

where $Vol_i(\%)$ is the volume of size class i in the volume-based particle size distribution, and k is the number of size classes. The volume-based distribution is then converted into a number-based distribution, based on which the *effective Sauter mean diameter is estimated*. The time constant for the surface-area approximation is then given as (Oeters 1994)

$$k_{\mathrm{tot}} = \beta_{tot} \frac{6}{d_{32,eff.}} \frac{\dot{m}_r}{\rho_{\mathrm{CaO}}} \frac{\rho_{\mathrm{Fe}}}{m_{\mathrm{Fe}}} t_{res}. \tag{49}$$

where $d_{32,\mathrm{eff.}}$ is the *effective Sauter mean diameter,* β_{tot} is the metal side mass transfer coefficient, and t_{res} is the residence time of the particles. It should be noted that it is expected that $d_{32,\mathrm{eff.}} > d_{32}$. It should also be noted that a somewhat similar approach is used in Publication I, but the expected value for contact is based on the estimation of the weighted time constant for the surface area approximation, and assuming that the particle velocities are random, following a normal distribution.

The parameter identification for the probabilistic risk function parameters is carried out with the real-coded genetic algorithm, using the RMSE as the objective function. In the identification, the mass transfer coefficient and the residence time are set as $\beta_{tot} = 5.5 \cdot 10^{-5}$ and $t_{res} = 12$ s respectively. Since the objective of this study is not to demonstrate the predictive performance but carry out a model-based estimation of particle size distribution, the identification is not validated exhaustively. The best performing parameters for a few repetitions are found to be $b = [b_0, b_1, b_2] = [1.39, 18.36, -0.39]$. With these parameters, the figures of merit are $R^2 = 0.93$ and $MAE_{ev} = 8.8$ ppm for the external validation set respectively. The corresponding objective function value is RMSE = 0.0017 for the training set. The resulting probabilistic functions are presented in Figure 5. The figure shows that according to the probability risk function, estimated based on the surface-area approximation, the probability of direct metal-particle contact increases withi the size of the particles, which is due to the decrease in the critical penetration velocity. Similarly, adding $CaCO_3$ to the reagent increases the probability of contact, which, according to Irons (1989), is because the decomposition of $CaCO_3$ causes the eruption of the carrier gas bubbles that entrap the reagent, and thus the injection of auxiliary compounds increases *the probability of direct reagent-metal contact*. With reference to this, the behaviour of the likelihood function is logical, and the contact probability may partly explain the covariance between the injection of $CaCO_3$ and the rate constant.

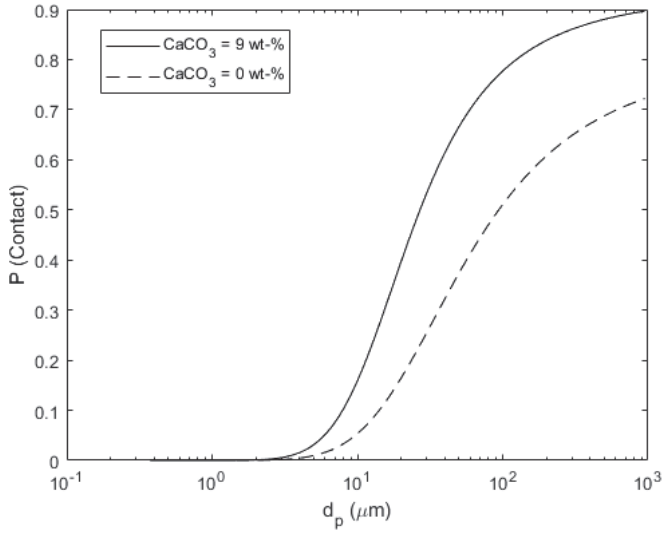

Fig. 5. Probability functions for particle-metal contact.

It should be noted that this approach suffers from two limitations. The first is that the surface area approximation assumes that the extraction capacity of the particles is of infinite magnitude. This means that as it ignores the solid-state diffusion-controlled phase, the rates for the smaller particle sizes are highly over-estimated (Oeters 1994), which can be seen in Figure 6. The form of the estimated particle size distribution indicates that the small particles do not come into contact with the metal at all, which results in the estimated mean surface area for mass transfer being significantly larger than the corresponding value measured for the powder prior to injection. However, it should also be noted that the values for the probability functions depend on the setting of the residence time, because the error concerning the limited extraction capacity is in fact a function of residence time, again due to solid-state diffusion control (Oeters 1994). For a more detailed estimate, a numerical solution and the expression of the solid-state diffusion control is needed to limit the extraction capacity of the distribution.

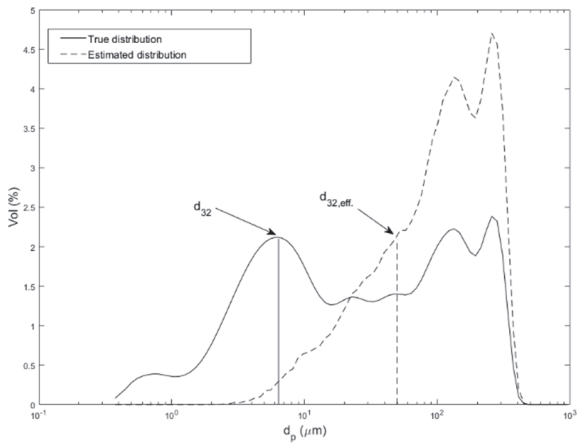

Fig. 6. The estimated particle size distribution in contact with the melt and the measured particle size distribution (true distribution). The estimated distribution has been normalised such that $\sum_{i=1}^{k} \boldsymbol{V}(\%) = \mathbf{100}$.

Models using particle size distribution

According to Publication V, the only model using the full particle size distribution is presented in the study of Visuri et al. (2019). The time constant surface area approximation that uses the expected value for the contacted volume fraction can be given as the sum of the time constant, defined for individual size classes:

$$k_{\mathrm{tot}} = \sum_{i=1}^{k} \beta_{tot,i} \frac{6}{d_{p,i.}} \frac{\dot{m}_r}{\rho_{\mathrm{CaO}}} \frac{\rho_{\mathrm{Fe}}}{m_{\mathrm{Fe}}} t_{res} \frac{E(m\ (\%) \mid Q_{\mathrm{tot,ac.}}, u_c)}{100}. \tag{50}$$

In this approach, it is assumed that mass transfer is controlled by the transport on the metal side, and the mass transfer coefficient is therefore approximated as discussed earlier. Hence, in this case, the number of parameters left to be optimised is four, one being the residence time and the rest arguments of the probabilistic function. The parameters are optimised with the real-coded genetic algorithm. To study the distribution of the residence time and the parameters of the probabilistic expression for the particle size distribution, 1,000 bootstrap samples are drawn from the data, based on which the parameters are estimated. The distributions of the parameters are given in Figure 7, which shows that the estimated parameters

follow an almost normal distribution. However, the parameter associated with the gas flow associated with the decomposition of the particles is biased to the right and follows an almost exponential distribution. The estimated residence time seems to be 9.28 s on average, and the standard deviation of the bootstrap samples is 1.97 s. The 2.5% and 97.5% quantiles of the distribution are 5.61 s and 12.88 s respectively.

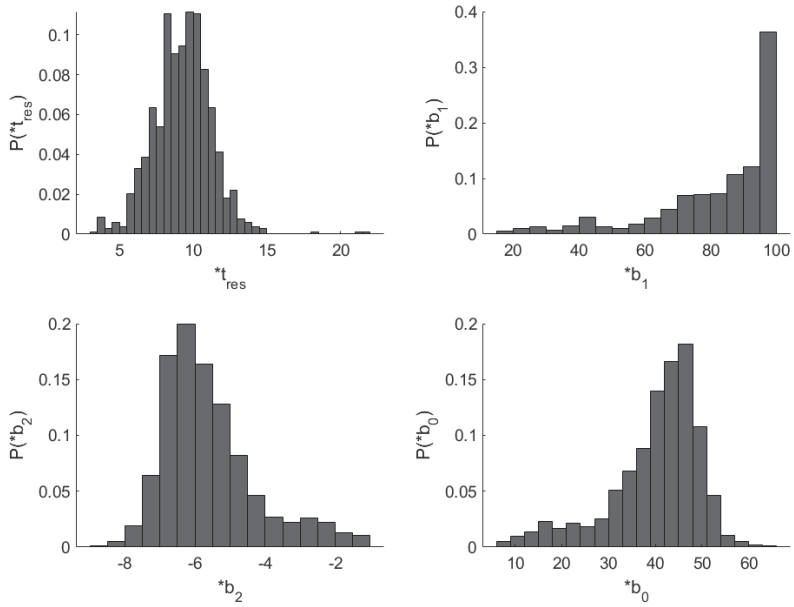

Fig. 7. Distribution of the parameters for 1,000 bootstrap samples.

5.2 A parameterised model for the transitory contact reaction in primary desulphurisation (Publication I)

In Publication I, a parameterised prediction model is presented to predict the end sulphur content based on Dataset 1. The final model structure is partly based on the surface-area approximation presented by Oeters (1994), which provides the initial subset of variables. Further, it is assumed that the desulphurisation reaction can be described well with first-order kinetics, provided the process operates relatively far from the thermodynamic equilibrium. This allows the neglection of the $[S]^{eq}$ by the

concentration difference. It is often suggested in the literature that the reaction rate of the transitory reaction is much higher than the rates R_{ii} and R_{iii}, and their exclusion from the rate expression may therefore be justified. The other assumptions considering the derivation of this model are given in Publication I. The detailed reasoning behind the model structure is discussed in detail in Publication I.

As said, in Publication I, the model structure and the predictor variables are identified with manual and forward selection, supported by regression-based hypothesis testing using a multiple linear regression as the model basis. As seen in the previous section, the data-driven reasoning and explanatory analysis with variable selection algorithm provides a very similar result to manual identification. Consequently, the variables chosen in the model are d_{80} particle size, Q_{tot}, $\dot{m}_r$ and m_{Fe}, because it appears to be the most parsimonious model explaining the variance in the overall time constant. The used objective function is derived such that the integral form of the first-order desulphurisation expression approaches the sulphur content at the end of the treatment when integrated from t_0 to t_{end}. In addition, it was also found that $[\mathrm{S}]_{\mathrm{t,pred}} = 0$ for all treatments is a local minimum for the non-constrained cost-function, which as a solution is practically infeasible. The details of the objective function are presented in Publication I, and are therefore not repeated here.

In the original publication, the model is trained with the genetic algorithm. The prediction performance of the parameterised model is evaluated with an external dataset, i.e. with a completely independent dataset. As the dataset that is used for the model parameter identification is small, and the used reagents discretise the variable space, the random subsampling method and the simple hold out are discarded as a validation technique. Instead, the data was split into training and validation sets using a following stratified hold-out procedure:

1. sort the data with respect to used reagent classes in ascending order;
2. draw class-wise samples randomly without replacement from the data;
3. continue until the specified number of samples is taken.

The number of selected samples for validation was 15, i.e. 37.5% of the full dataset. As the number of reagents in the data is 5, the above procedure ensures that at least three samples present each reagent. Table 6 presents the quantitative figures of merit and the suggested model parameters. The measured and predicted sulphur values at the end of the treatment for both sets are presented in Figure 8. The table and the figure show that the model is capable of explaining the variance in the

sulphur content at the end of the treatment with sufficient accuracy. The quantitative figures of merit are R^2 = 0.91 and MAE = 11.3 ppm for training, and R^2 = 0.90 and MAE = 10.9 ppm for the external validation set selected with the stratified hold out. The simple model structure and the consistency of the external validation predictions with the predictions on the training support the applicability of the model structure.

The suggested model parameters are presented in Table 6. The direction of the parameters shows that increasing the particle size decreases the rate of reaction, whereas the addition of gas-forming compounds to the reagent increases the rate of desulphurisation. The parameter values are consistent with those identified in the previous section, but only semi-quantitatively, because the model structure is non-linear. The parameter values are also directionally similar to the findings of Chiang et al. (1990) in the case of CaC_2 and Coudure and Irons (1994), who have a similar model structure. Thus, the domain knowledge supports the identification results. However, it should be noted that the parameters do not necessarily generalise well to other processes, because the modelling scheme makes no assumptions considering the process equipment, for example, which may induce variance to the rate of desulphurisation, especially when considering the parameters attributed to the effect of particle size distribution and the average volumetric amount of gas in the system. In addition, it should be noted that without the information concerning the variance of the particle size distributions of the lime-based reagents during the treatment, the effect of this attribute on the overall rate obviously cannot be identified.

Table 6. Model parameters and the figures of merit computed for training and external validation datasets (constructed based on the data given in Publication I).

Value	b_0	b_1	b_2	b_3	b_4
Model parameters (GA)	–1.05	0.50	0.48	0.91	1.03
R^2- training	0.91				
R^2 - external validation	0.90				
MAE - training (ppm)	11.3				
MAE - external validation (ppm)	10.9				

During the analysis, it was also found that the parameterisation of the model significantly improved model performance. Indeed, the model is more accurate

than detailed mechanistic models such as the one presented in the study of Visuri et al. (2019). However, the parameterised model is only suitable for process control purposes, because it does not provide a detailed insight into the flows prevailing in the ladle or rate controlling mechanisms from the perspective of individual particles, for example. However, this allows a fast computation of the estimation of the end sulphur content, but it provides no further information. However, the model structure serves its use in predictive purposes very well.

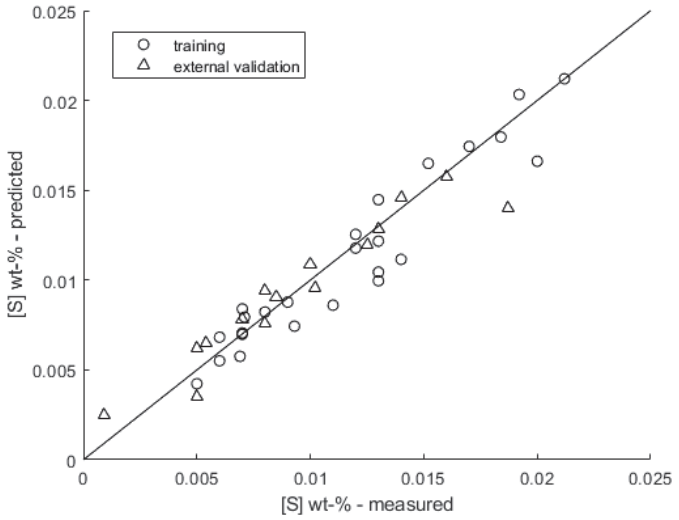

Fig. 8. Measured and predicted sulphur content at the end of the treatment.

5.3 Parameterised model for transitory and permanent contact reactions

As seen in the previous section, the variance in the rate and the end content of sulphur can be explained well by the kinetics of the transitory reaction that have been described by the parameterised surface area approximation. As Oeters (1994) points out, surface area approximation assumes that the extraction capacity of the particles is of infinite magnitude, and the expression may therefore over-estimate the contribution of the transitory contact reaction. Furthermore, the model proposed earlier conducts the identification by ignoring the physical meanings of the parameters, considering only dimensionless parameters, of which comparison with

the domain knowledge is not straightforward. Considering the reasoning above, the model selection and parameter identification is conducted by taking account of the slag-metal reaction and the finite extraction capacity of the overall particle size distribution. The overall rate is formulated as a sum of the considered reaction rates. The model structure selection is presented in the following subsections.

5.3.1 Kinetic treatment of the transitory reaction

In the models of Deo and Boom (1993), Seshadri et al. (1997), and Visuri et al. (2019), the sulphur extraction potential of the individual particles is constrained by introducing the microkinetic efficiency of the particles. In addition, to account for the fraction of particles that remain entrapped in the carrier gas bubbles, the rate expression is constrained by weighting the feed rate of the particles with the term $1 - f_p$ in the macrokinetic term (Deo & Boom 1993). With these considerations, the overall rate of the transitory contact reaction can be expressed as:

$$R_i = -\left(1 - f_{\mathrm{p}}\right) L_{S,\mathrm{CaO}} \frac{\dot{m}_r \rho_m}{\rho_{\mathrm{CaO}} m_m} \left[1 - \exp\left(-\frac{6}{d_p} \beta_m t_{res} \frac{1}{L_{S,\mathrm{CaO}}}\right)\right] [\%\mathrm{S}], \qquad (51)$$

where f_p is the fraction of non-contacted particles, $L_{S,\mathrm{CaO}}$ is the partition ratio of sulphur assuming that the sulphur is in equilibrium with the solid CaO, and t_{res} is the residence time of the particles. It should be noted that the overall expression is similar to that reported by Rastogi et al. (1994), who employ the expression to study the relative contributions of different rate mechanisms. For each iteration, the residence time is chosen from $t_{\mathrm{res}} = \min(t, t_{\mathrm{res}})$. As discussed in Publications I and V, the residence time of the particles, the fraction of non-contacted particles, and the mass transfer coefficient for the metal phase cannot be estimated accurately without formulating the governing equations for the momentum and mass transfer, which makes the model overly complex for online use. Instead, these parameters are optimised as in Publication I, but for a dynamic case.

As Coudure and Irons (1994) and Chiang et al. (1990) state, the fraction of non-contacted particles is not independent of the particle size, injection gas flowrate, and the injection of gas-forming compounds. The study of Lindström et al. (2014) recognises the effect of gas-forming compounds on the increased probability that the particles will be scattered, although in a two-phase system in which the contact probability with the reagent and metal is significantly higher

(Lindström et al. 2014). The dependency of particle size can be further associated with the critical Weber number and thus the critical penetration velocity (u_c), which by definition shows that decreasing the particle size increases the velocity, which must be overcome for the particles to wet the metal phase (Nakano & Ito 2016). The critical velocity is the velocity of an object required to overcome the surface forces in the situation where the particle meets the metal, which, in the case of small particles, leads to the entrapment of particles in carrier gas bubbles, which is often referred to as contact control in the literature (Chiang et al. 1990; Irons 1989). More detailed analyses of this velocity and the critical Weber number in the context of hot metal desulphurisation are given in Oeters (1994) and Nakano and Ito (2016). However, the evaluation of the single particle size class penetration probability is not straightforward. As described by Farias and Irons (1985), the penetration behaviour of particle jets also depends on the flow scheme prevailing in the submerged lance. The authors divide the schemes into coupled and non-coupled flow. In a coupled flow situation, the particles travel with a velocity close to that of the superficial gas velocity, whereas in a non-coupled flow, the difference between the superficial gas velocity and particle velocity is greater (Farias & Irons 1985). As it can be assumed that the flow characteristics are somewhat stochastic in nature, the fraction of non-contacted particles is described by a parameterised expression. For this purpose, the following functional relation was proposed to describe these effects:

$$f = b_0 \ln(Q_{tot}) + b_1 \ln(u_C), \tag{52}$$

where u_c is the critical penetration velocity calculated from the correlation given in Nakano and Ito (2016). As argued by both authors and Irons (1989), the natural constraints for the suggested parameters would be $b_0 < 0$ and $b_1 > 0$, because the injection of gas-forming additives increases the probability of direct particle-metal contact (Irons 1989), whereas a larger particle size decreases the critical penetration velocity, increasing the probability of contact (Nakano & Ito 2016). In the lance tip, the volumetric gas rate is the sum of the injection gas rate and the amount of gas generated by the decomposition of carbonates, assuming a 100% conversion to CO_2. The total gas volume is computed such that it is expected to be heated to 80% of the bath temperature (p = 1 atm + $\rho_{Fe}gh$; $T = 0.8T_m$), which is on average just above the decomposition temperature of the calcium carbonate. This assumption agrees with Irons (1989), who proposes that the decomposition of carbonates occurs instantly when the particles end up in the melt. In addition, Yousuf et al.

(2012) observe that the decomposition rate of particles larger than 250 μm is remarkably slower than that of smaller particles in the same system conditions. This behaviour may be associated with the heat transfer-related control of decomposition inside the particles due to the endothermic nature of the carbonate decomposition (Yousuf et al. 2012).

Similarly, as in the case of non-contacted particles, the overall mass transfer coefficient is computationally extremely complex to estimate, because it depends, for example, on the characteristic length of diffusion and the flow conditions in the ladle, which are rather inaccurately quantified in the context of hot metal desulphurisation (Oeters 1994). Similar reasoning supporting the estimation of the mass transfer coefficient from the data is provided by Brooks et al. (2009) in the context of emulsified droplets. The mass transfer coefficient therefore remains to be identified.

Treatment of particle size distribution

To estimate the particle size specific rate constant, two approaches are used. In the first approach, a single characteristic value is used as the identification basis. This comes with an assumption that this single attribute describes the differences between the distributions consistently and adequately. The attributes used are d_{80} (the particle size below 80% of the particle volume), d_{ka}, d_{32} (the Sauter mean diameter), and d_{V} (the volumetric mean diameter). However, as the single particle models only reveal the averaged effect of particle size distribution, they are not studied further in this case.

The second approach includes the use of the Rosin-Rammler-Sperling distribution, whose applications originate in mineral processing (Bayat 2015). In this model, the cumulative mass fraction of particles is expressed as

$$W_p = 1 - \exp(-c_1 \, d_p^{c_2}), \tag{53}$$

where W_p is the cumulative mass fraction of a particle with a diameter d, and c_1 and c_2 are experimental fitting parameters, which are optimised by means of least squares with the distribution expression and characteristic particle sizes d_{10}, d_{25}, d_{50}, d_{80}, and d_{90}. The overall time constant can be expressed as a sum of particle size class specific time constants, as in Visuri et al. (2019) and Publication I. It should be noted that the use of RRS distribution in the computation decreases the computational load of the model parameter identification, because time integration

is carried out numerically instead of using measured particle size distribution. The characteristic particle size distributions and the corresponding RRS distribution parameters for each of the studied reagents are given in Table 7, and the functional forms are exemplified in Figure 9. With reference to Publication I, Reagents A, B, and E contain 9, 5, and 0 wt-% of $CaCO_3$, which affects the total volumetric rate of gaseous compounds in the system.

Table 7. Characteristic particle size distributions of the studied reagents with corresponding RRS distribution parameters.

Reagent	d_{90}	d_{80}	d_{50}	d_{25}	d_{10}	c_1	c_2
A	240.0	135.0	26.5	6.2	2.8	0.06	0.69
B	482.5	223.6	33.8	6.4	3.0	0.06	0.61
C	410.1	233.6	43.7	7.2	3.2	0.06	0.62
D	97.9	73.3	24.2	4.1	1.2	0.09	0.67
E	117.1	69.9	10.4	4.4	2.2	0.07	0.76

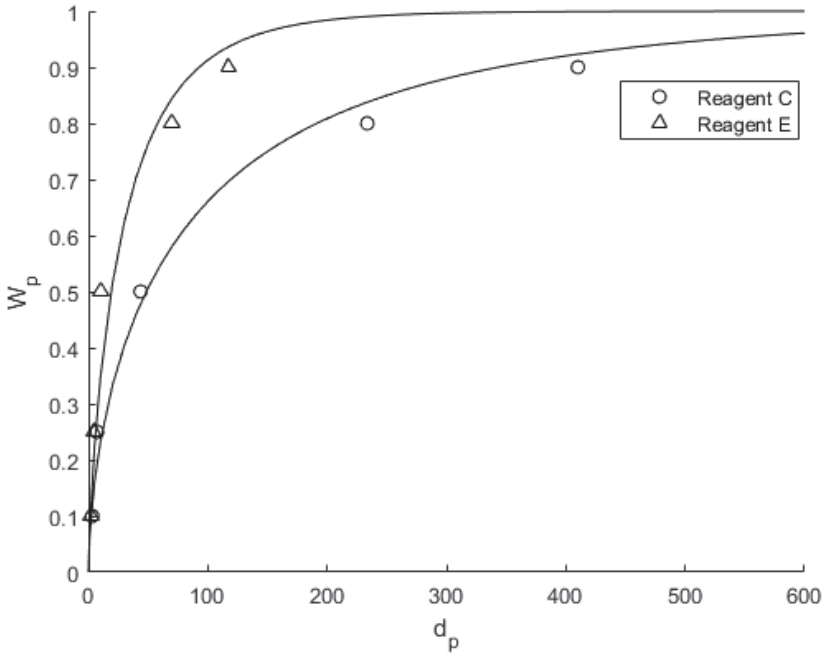

Fig. 9. RRS distributions fitted to the data for Reagents C and E. Cumulative mass fraction of a specific particle size is presented on the *y*-axis.

5.3.2 Kinetic treatment of the slag-metal reaction

According to Seshadri et al. (1997), the expression for the rate of slag-metal can be given as follows:

$$R_{ii} = -\beta_{tot}\frac{A}{V_m}\left([\%\mathrm{S}] - \frac{(\%\mathrm{S})}{L_\mathrm{S}}\right), \tag{54}$$

where β_{tot} is now the overall mass transfer coefficient. In a steady flow field, the rate of the control step of the slag-metal desulphurisation reaction can be either the mass transfer in the slag or the metal phase. To describe the sulphide capacity of the slag phase, two different approaches are used, i.e. the MLR model selected based on Dataset 3 and the model proposed by Zhang et al. (2013). The reasoning behind this is given in Publication II and Section 5.7.2. As during the injection, the

slag phase is obviously not in the steady flow state, and the mass transfer correlations assuming steady flow conditions are unsuitable. However, to determine these attributes accurately, an approximation of the flow field in the ladle is needed. Instead, the semi-empirical correlation proposed by Riboud and Olette (1982) is used to describe the overall mass transfer coefficient for the slag-metal reaction.

As Riboud and Olette (1982) derive the approach by assuming that the gas ascending in the metal phase is heated to the melt temperature when it reaches the slag-metal interface, the sum of the volumetric rate of gas formed by the decomposition of carbonates and the injection gas is assumed to correspond to Q_{tot} (p = 1 atm; $T = T_m$). Riboud and Olette (1982) suggest a value τ = 500 $m^{0.5}$ for practical applications, i.e. in the desulphurisation of steel in the converter and ladle with bottom plug stirring (Riboud & Olette 1982). However, the value proposed by the authors has been estimated for a permanent contact reaction with bottom stirring and for mainly liquid slags, which is why in this work, the τ is omitted from the estimation of the data. In addition, Chiang et al. (1990) have empirically proved that the rate of desulphurisation for a "dry" slag is significantly slower than for liquid slags (Chiang et al. 1990). As discussed in Publication II and by Schrama et al. (2020), the solid fraction of the HMD slag is around 20–100%, depending on the injection stage, composition, and temperature, which can be assumed to make the slag quite stiff, and as Chiang et al. (1990) state, dry. Consequently, it can be assumed that the magnitude of τ may be relatively low compared to the original value proposed by Riboud and Olette (1982).

5.3.3 Numerical solution and the parameterisation strategy

As the system state, and especially the sulphide capacity of the slag and the residence time of the particles, evolve dynamically, an analytical solution is unsuitable for the objective function. Instead, a numerical approximation obtained with an implicit Euler method is used. The implicit Euler method is suitable for stiff differential equations, which usually contain more than one-time constants. In the implicit Euler method, the sulphur content at the next time instant can be given as follows (Kreyszig 2006):

$$[\%S]_{t+h} = [\%S]_t + f(t + h, [\%S]_{t+h}). \quad (55)$$

As both sides of the equation now contain the term $[\%S]_{t+h}$, it needs to be solved either analytically or numerically. In this study, the Newton-Rhapson method is used. As seen in the description of the model, the total number of parameters is five, i.e. the parameter vector to be optimised is $b = [b_0\ b_1\ \beta_{tot}\ t_{\mathrm{res}}\ \tau]$. It should be noted that the convergence is tested by systematically making each of the design variables constant one by one (i.e. the penetration ratio, mass transfer coefficient, and residence time), following an experimental design matrix. With this procedure, it is found that the parameter combination presented above gives the lowest error for the full data. It has been suggested in the literature, for example, in Oeters (1994), that the suitable value for the residence time is a maximum of 78 s, which suggests constraining the residence time in the search. Similarly, as the reagent is practically free from sulphur, the driving force of the reaction supports desulphurisation, and therefore a negative rate expression. The time constant for the expression of the rate of transitory reaction is therefore constrained as $k_i \geq 0$, which in practice, partly rules out the non-feasible solutions connected with the empirical relation considering the particle-metal contact. With these considerations, the objective function used for the identification is

$$\min \mathrm{RMSE} = \min \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left([\mathrm{S}]_{t,meas.} - [\mathrm{S}]_{t,pred}\right)^2}\ , \tag{56}$$

which is subject to

$$t_{res} \leq t_{res,max}, \tag{57}$$

and

$$k_i \geq 0, \tag{58}$$

where RMSE is the root-mean squared error, and $t_{res,max}$ is the maximum residence time. Because the model used is dynamic, the function to be optimised is assumed to be multimodal, which makes the use of gradient-based methods non-reliable for a global search. The real-coded genetic algorithm is therefore used as the solving strategy. The algorithm hyperparameters were set to $n_{\mathrm{pop}} = 50$, $max_{gen} = 100$ and $p_C = 0.9$. The mutation probability was set to evolve deterministically in line with Bäck and Schultz (1996). The schematic illustration of the parameter identification flowchart is presented in Figure 10.

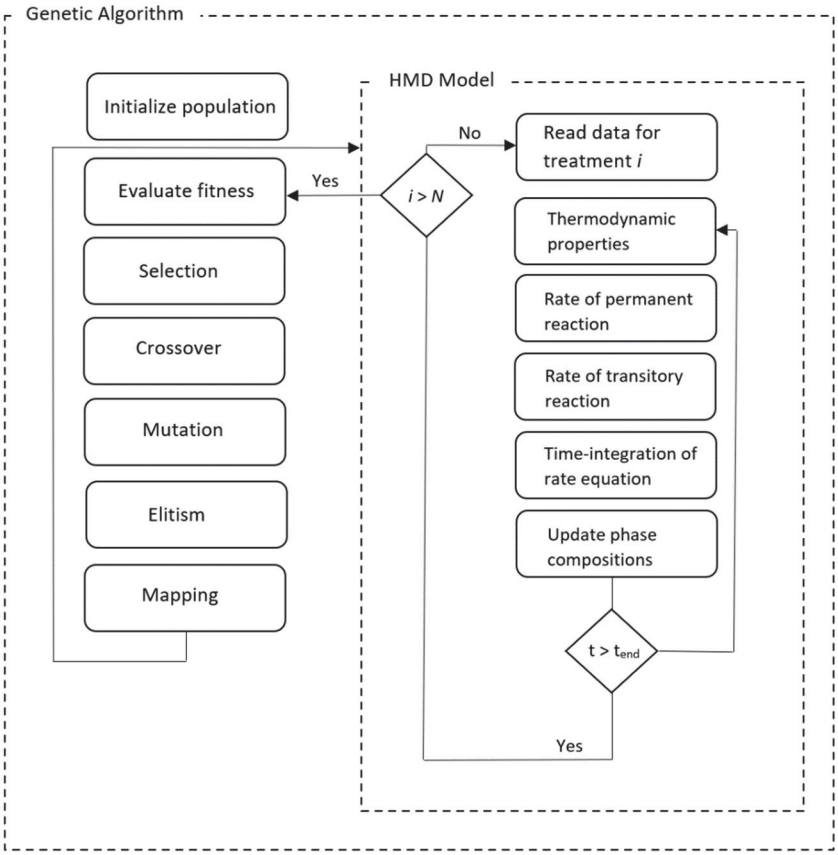

Fig. 10. Flowchart of the parameter identification for the proposed model. The figure is simplified to illustrate the treatment of a single individual in the genetic population.

5.3.4 Results of the identification and analysis of the system

To identify the model parameters, the genetic algorithm and the stratified data split are used. In the first phase, the model performance and parameter identification procedures are compared for a static data split, and the variability of the parameter values is studied with the corresponding objective function values. In the second step, the identification is conducted with repeated stratified cross-validation to study the variability of the parameters and the model predictions for different data

splits. The best parameters obtained in the first identification step are given in Table 8**.** The values for the mass transfer coefficient agree well with the literature, especially with the values proposed by Oeters (1994). The figures of merit for both cases are approximately R^2 = 0.87–0.88 and MAE = 13–14 ppm for the training, and R^2 = 0.91–0.93 and MAE = 9.24–9.57 ppm for the external validation dataset. The model predictions obtained with the given parameters are presented in Figure 11.

Table 8. The best parameters obtained with the identification in the first step for the data set 1.

C_S	b_0	b_1	β_{tot} (m/s)	t_{res} (s)	τ ($m^{-0.5}$)
Zhang et al. (2013)	–0.22	0.09	$6.1 \cdot 10^{-5}$	7.9	93.6
This work	–0.25	0.06	$1.7 \cdot 10^{-5}$	21.3	128.9

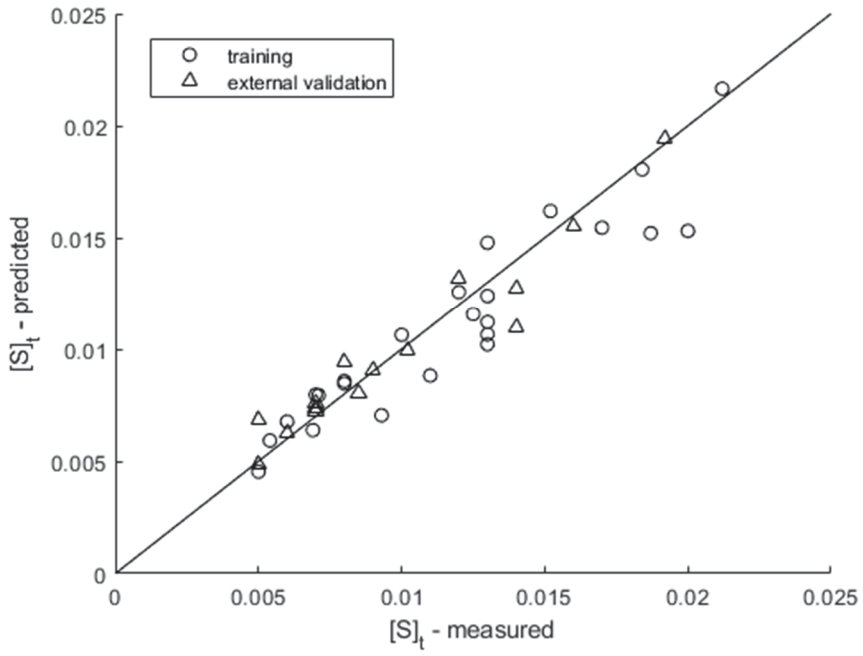

Fig. 11. Measured and predicted end sulphur content using RRS distribution with 50 size classes.

To characterise the distributions of the suitable parameter values for a single piece of data, a repetitive test is undertaken. The dataset is divided into training and external validation sets with the procedure described in Section 5.2. In this phase, the identification is repeated 50 times. It is found that for each of the tests, the algorithm converges to the same objective function value, RMSE = 0.0017 wt-%. However, as Figure 12 illustrates, the parameter values vary between repetitions. This behaviour can be explained with the multimodality of the objective function, even for a single data split. The multimodality partly arises from the uncertainties discussed earlier that consider the phenomenological behaviour of the system.

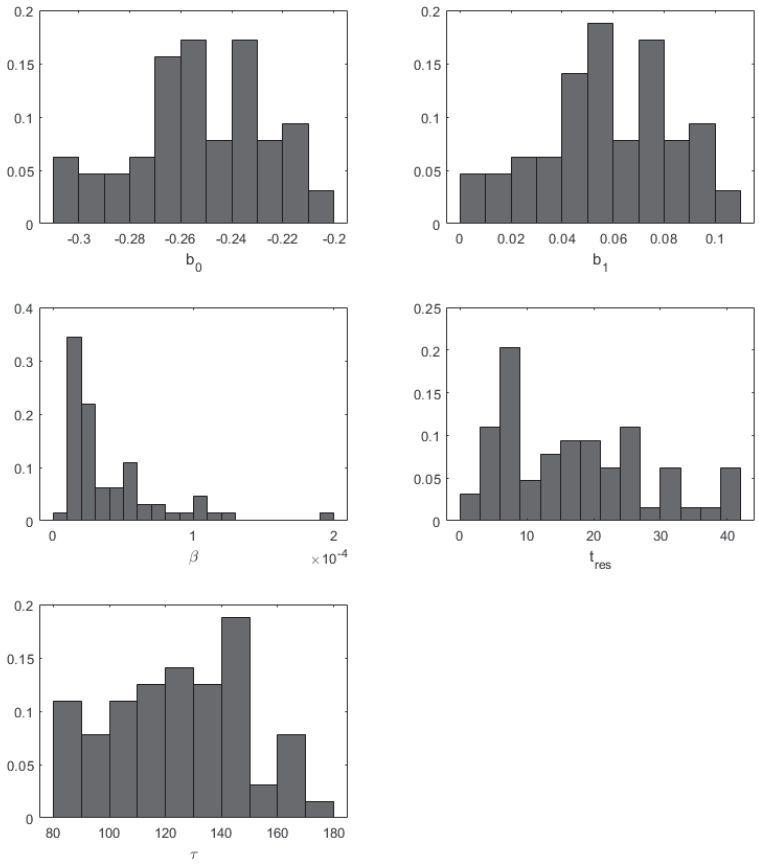

Fig. 12. Histograms of the parameter occurrences in repetitive testing for a static split. The identification is conducted using the sulphide capacity model of Zhang et al. (2013).

As can be seen, the values identified for the residence time agree with the literature, but the variance in the parameter values and the multimodality of the objective function still indicate that the uncertainty is relatively large. In Figure 13, the objective function values are illustrated with respect to the discretised parameter space. Instead of a projection onto two parameter spaces, the objective function values are illustrated as circles, and two address the issues concerning system identification. In the figure, the size and colour of the circle illustrate the objective function values. The objective function values show that there are multiple equally good solutions with respect to the error function, and no explicit conclusions can

be drawn from the parameters' absolute values. This strongly agrees with the findings of Oeters (1994), who states that the best fit for the residence time greatly depends on the assumptions made concerning the system. However, as seen in the figure, all the residence times that provide a physically relevant residence time for the particles agree with the literature, for example, with the studies of Oeters (1994) and Ma et al. (2017).

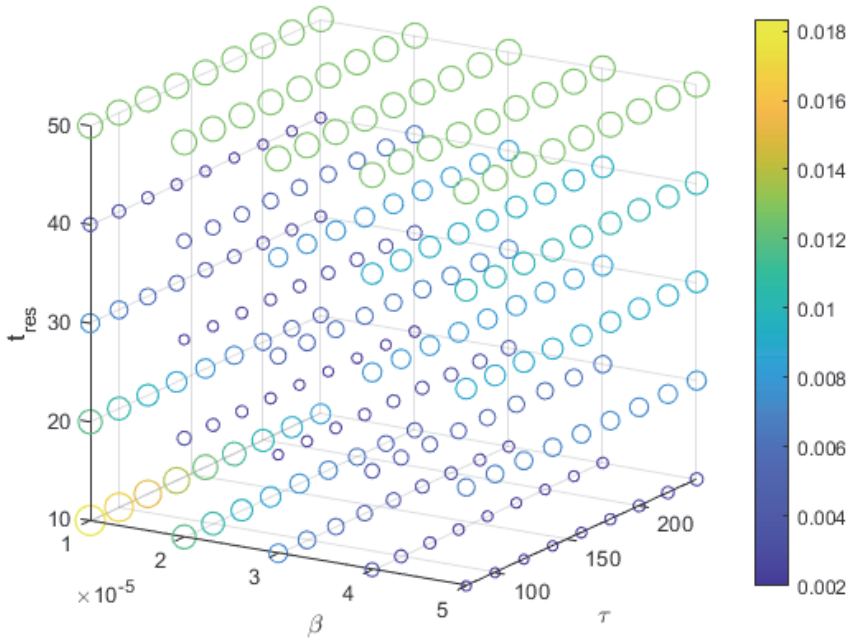

Fig. 13. Objective function values as a function of the physically meaningful parameters. The objective function is constructed by applying the sulphide capacity model of Zhang et al. (2013).

As was seen in Section 5.1.2., if the residence time and mass transfer are assumed to be constant, the degree of freedom in the system decreases, which provides a solution that can be interpreted more precisely but is not necessarily physically feasible. Yet with fewer parameters, the degrees of freedom drop in the system, allowing a more explicit solution. Another reasoning behind the multimodality of the objective function could be because the effect of slag cannot be explicitly identified in the data, which may be partly because the covariance of the slag is

simply not observable in the data, because the covariance between the reagent and rate of desulphurisation is simply higher. To solve this issue, a more detailed dataset is needed. For example, a dataset without a reagent injection and in which the slag is of equal composition when there is an injection would probably indicate the total mass transfer coefficient for the permanent contact reaction of a stiff slag. Another source of variance is found to consider the used sulphide capacity model for the slag phase. This issue is discussed in more detail in Publication II and Section 5.7. However, with the current knowledge and data, no unambiguous solution is available. More detailed reasoning concerning these issues can be found in Publication V.

In the second step, model parameter identification is conducted for repeated stratified data splits to study the effect of data split on the variability of parameter values and predictive indicators. The splits are repeated 100 times for both sulphide capacity models. In Figure 14, the histograms of the R^2 values for the external validation set are presented in the case of the model of Zhang et al. (2013). The distribution of R^2 values for the sulphide capacity model derived in this study is very similar, so for the sake of clarity, it is not presented here. The remaining repetition statistics and parameter values for both cases are presented in Table 9. The table shows that as the prediction indicators are quite stable, it is reasonable to expect a prediction error of 11–13 ppm for a dataset similar to this data. For both models, the mean and median values of R^2 for the test set are 0.89, and the MAE is around 12 ppm. However, a slight instability in the parameter values can be found, and in some cases, the parameter values are clearly overfitted, resulting in an external validation error of 17–20 ppm, even though the training error is around 10 ppm.

The table shows the standard deviation of the parameters between the split repetitions, especially for the residence time and the τ. It is therefore obvious that more data would be beneficial for making a more accurate estimate of the prediction error and improving the model's stability, i.e. to decrease the standard error of the parameters. An obvious solution for model simplification and increasing stability would be the neglection of the R_{ii} (slag-metal reaction) from the rate expression, because the original data for the slag composition variables is relatively sparse, and the number of imputed points is therefore quite large.

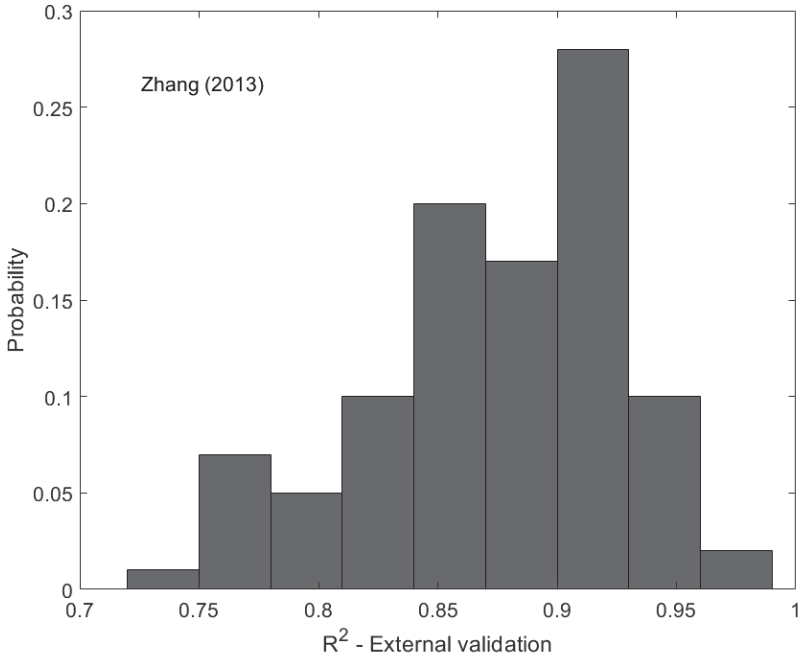

Fig. 14. Histogram of coefficient of determination (R^2) values for the external validation set. The values are identified using the sulphide capacity model of Zhang et al. (2013).

Table 9. Repetition test statistics for stratified split test.

Study	b_0	b_1	b_2	b_3	b_4	Metric	Mean	Best	Worst
This study									
mean	–0.26	0.05	$3.5 \cdot 10^{-5}$	21.57	118.51	R^2 - t.	0.90	0.96	0.84
std	0.03	0.03	$3.3 \cdot 10^{-5}$	13.62	31.60	MAE- t.	11.17	7.99	15.46
						R^2 - ev.	0.89	0.95	0.74
						MAE - ev.	12.58	8.39	19.92
Zhang et al. (2013)									
mean	–0.248	0.065	$3.3 \cdot 10^{-5}$	24.12	88.21	R^2 - t.	0.89	0.95	0.80
std	0.031	0.029	$2.1 \cdot 10^{-5}$	12.20	41.29	MAE- t.	11.47	7.93	14.67
						R^2- ev.	0.88	0.97	0.74
						MAE - ev.	13.82	8.29	20.52

The estimated mass fractions of particles that come into contact with the metal phase are given in Figure 15. The figure shows that the fractions of particles that come into contact with the metal phase depend on the particle size and the volumetric rate of gas in the system, accounting for the decomposition of the carbonates and injection gas. According to this approach, the solid flowrate increases the fraction of particles in contact with the melt, only if carbonates are mixed with the reagent. The dependency of the carbonate injection is most prominent between 0 to 5 wt-%, because of the assumed logistic dependency. However, a comparison of the figures with the values obtained with surface area approximation shows that the non-contacted mass fraction approaches 0 when the particle size approaches 0. With limited extraction capacity, it is estimated that 25–40 % of the 10 µm particles get in contact with the melt. However, the dependency is not as strong as in surface area approximation, because the surface area approximation proposes that the expected volume fraction is nearly 0 for particles smaller than 10 µm, even with shorter residence times.

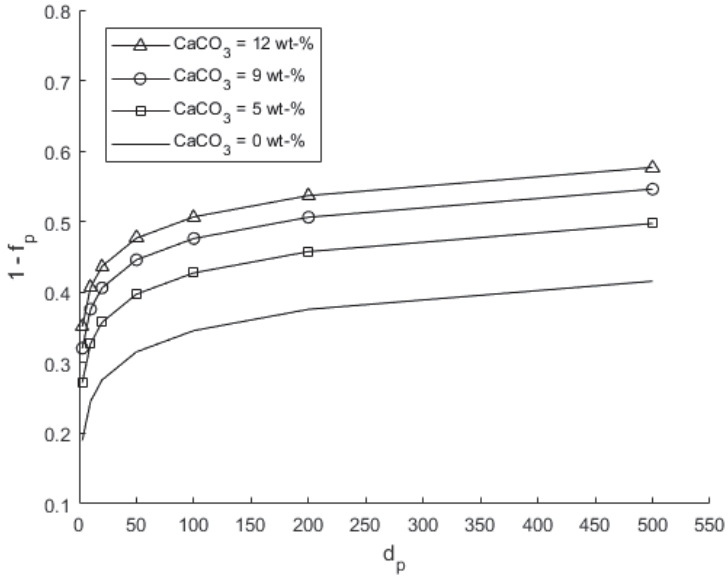

Fig. 15. Estimated fraction of particles in contact with the metal phase ($1 - f_p$) as a function of particle size and carbonate content in the reagent in the conditions assumed in the lance tip. The solid flowrate was assumed constant (m_r = 100 kg/min). The particle size distribution is estimated using the RRS distribution with N = 50 size classes.

Simulation study

To analyse the behaviour of the process and the dependencies, a simulation study is undertaken. The simulations are conducted by systematically changing the initial conditions by interpolating between the extreme values. The identified parameters for the simulation correspond to those given in Table 8. The simulated sulphur trajectory with different particle size distributions is shown in Figure 16. The figure shows that by decreasing the particle size distribution, the overall desulphurisation rate increases. This can mainly be attributed to the increased solid surface area, which results in an increased rate of reaction and thus a higher utilisation ratio of individual particles.

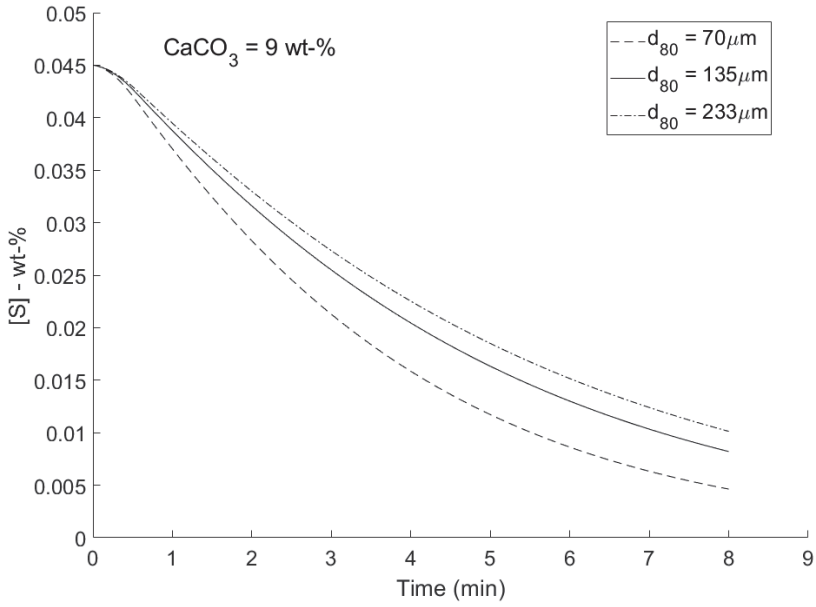

Fig. 16. Simulated sulphur trajectory with different particle size distributions. The sulphide capacity of the slag is estimated using the correlation proposed by Zhang et al. (2013).

The simulated sulphur trajectory with varying amounts of carbonate injected within the primary reagent and constant other system properties are shown in Figure 17. The sulphur trajectory shows that the injection of gas-forming compounds increases the rate of reaction. In the model, this is attributed to an increased fraction

of particles in contact with the metal phase, which is in line with the findings of Irons (1989) and Lindström et al. (2015) that consider the effect of decomposition of gas-forming additives on the reaction mechanism.

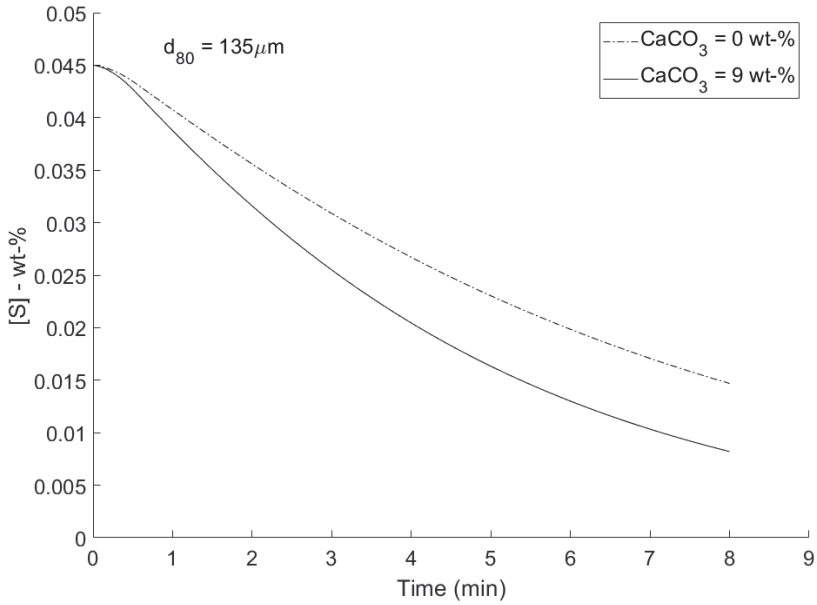

Fig. 17. Simulated sulphur trajectory with different $CaCO_3$ wt-% in the reagent. The sulphide capacity of the slag is estimated using the correlation proposed by Zhang et al. (2013).

A further analysis of the simulations reveals that the effect of particle size and the flowrate of gas-forming compounds on the reagent efficiency interact. This is because the injection of gas-forming compounds increases the fraction of particles coming into contact with the melt, and the reagents without gas-forming compounds therefore benefit the finer gradation more. Decreasing the particle size d_{80} from 233 µm to 70 µm, and assuming the shape of the distribution remains the same, the increase in the reagent efficiency is around 1%, whereas when the $CaCO_3$ content of the reagent is 9 wt-%, the increase is 0.9%. It should be noted that in this case, such an increase in the reagent efficiency corresponds to a 50–60 ppm drop in the end sulphur content under these conditions.

Figure 18 shows the effect of Na_2O content of the slag on the endpoint of sulphur is not as remarkable as the other variables. This is mainly because with such high injection rates, the end compositions of both cases are practically the same, resulting in sulphide capacity of a similar magnitude. The difference between the sulphur content is around 10 ppm for a simulated 8-minute treatment. A difference as small as this is quite difficult to verify, because the random of the measurement for the sulphur content of the metal phase is in the order of several ppms. It is therefore obvious that the effect of slag composition on the endpoint of sulphur cannot be reliably estimated from the current data because of their non-constant reagent properties and strong effects on the overall rate. With a different dataset, the estimation of the effect of slag would be more reliable, meaning that the rate of the transitory reaction and the permanent contact reaction could be differentiated.

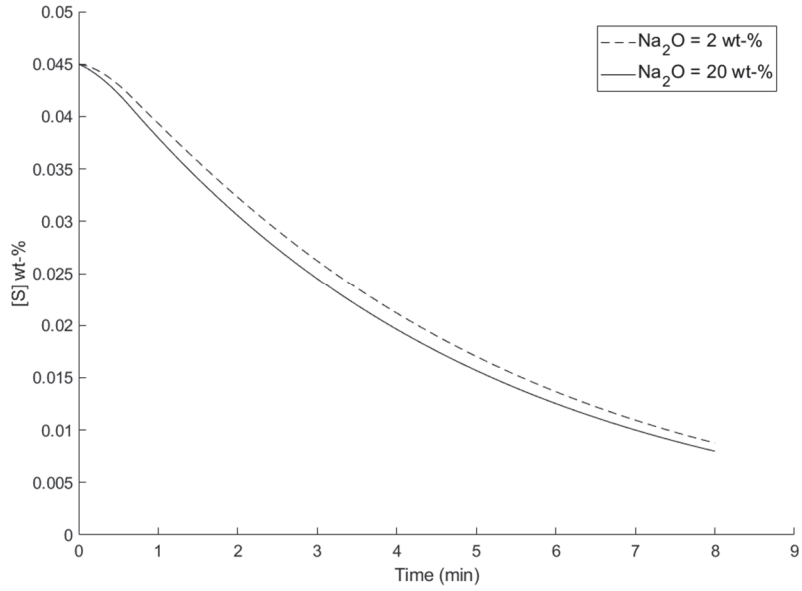

Fig. 18. Simulated sulphur trajectory with varying amounts of Na_2O in the initial slag. The sulphide capacity of the slag is estimated using the correlation proposed by Zhang et al. (2013).

5.4 Data-driven model selection for the transitory contact reaction (Publication III)

As indicated in Publication I and the previous sections, variable and model structure selection usually requires external knowledge of the system to derive meaningful relations. The discussed situations are two special cases of manual selection and model parameter identification. As seen in the previous section, the use of expert knowledge in model structure selection may sometimes allow a more comprehensive explanatory analysis of system behaviour, but model structure selection needs to be conducted carefully and with a solid argumentative basis, and a wide knowledge base is therefore needed. In this section, the results of the automatic identification are compared with manual model identification strategies. As stated in Section 3, the use of model identification algorithms requires a rather generic model structure, like the one presented in Publication III. In the referenced article, a model with a generic structure is identified solely based on the experimental data using a binary-encoded genetic algorithm, a description of which was provided earlier in this study. The applied dataset is the same as used in the parameterisation and validation of the parameterised reaction models presented in the previous sections.

The results of the variable selection algorithm in Publication III agree with the results in Publication I, because the proposed algorithm repeatedly selects the same variables as manual selection based on phenomena-based reasoning. In addition, the results are very similar to the results of an exhaustive search. The other deterministic algorithms (Forward Selection 1 and Forward Selection 2) are also compared to the results of the genetic algorithm. Similarly, as in Publication III, the objective function is based on $4N$ cross-validation, and the fractions of data for internal and external validation are the same. The results of the genetic algorithm are presented only for $n_{pop} = 100$, and the dependency of the search results on the algorithm parameters is more comprehensively reported in Publication III. For example, increasing the number of individuals in the search increases the probability of finding the best subset (d_{80}, Q_{tot}, $\dot{m}_r$ and m_{Fe}) for this data. However, increasing the number of individuals comes with the increased computational cost. The results of the repetitive tests are presented in Figure 19. The figure shows that the results of Forward Selection 1 and the genetic algorithm are very similar. However, Forward Selection 2 tends to select a larger number of variables and consequently some assumedly redundant variables for the model. In addition, the hit rate of the most important variables is ~100% for the genetic algorithm and

Forward Selection 1, whereas Forward Selection 3 tends to omit the particle size distribution (x_5) from the subset. In principle, the results are comparable to exhaustive search.

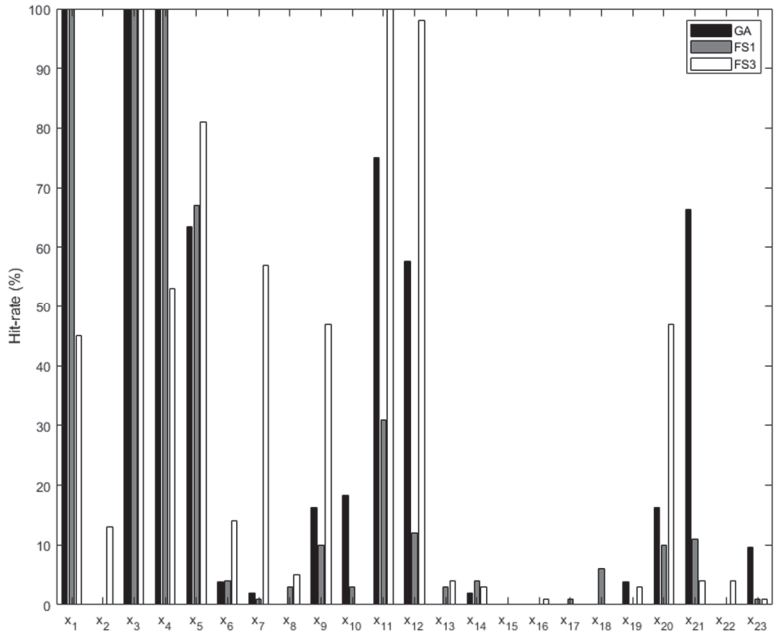

Fig. 19. Selected variables with different methods, using repeated random subsampling as the objective function. GA = Genetic Algorithm, FS1 = Forward Selection 1 and FS3 = Forward Selection 3. The data for the Genetic Algorithm is given in Publication III.

5.5 Identification of a neural network model for carbide-based hot metal desulphurisation (Publication IV)

Publication IV introduces a model selection scheme completely independent of the expert knowledge. In this case, the reagent used is calcium carbide. As the previous sections show, several uncertainties prevail in the hot metal desulphurisation system, and the process is non-linear in nature. It is also expected that the uncertainties, especially in the thermodynamic driving force, make the assumptions concerning the system unsuitable, because there is no information on the

composition of the slag phase. A black box approach is therefore introduced. In this study, a neural network model is identified, using a genetic algorithm with hybrid encoding. In this approach, the variable space is described as binary-encoded vectors, whereas the neural network architecture is assumed to be a fully connected network, whose structure can be expressed with a single integer. To decrease the computational complexity of model selection, the ELM architecture is used as the model basis. After the selection phase, the final model is trained using the Bayesian regularisation algorithm. In the original Publication IV, it was found that a simple genetic algorithm using either a single or two-point crossover could be used to select well-generalising neural network models for carbide-based hot metal desulphurisation. However, the selection results were found to depend on the applied objective function and the size of the population. The number of internal validation splits was especially found to be an important parameter to consider.

Figure 20 presents an exemplified training and internal validation error (mean squared error averaged over 50 data splits) for the end population. It should be noted that the model error is only a function of network complexity, because the convergence criterion is set as a homogeneous population. The figure clearly shows that the use of cross-validation error as the selection criterion instead of training error provides a less optimistic estimate of model prediction performance, and the network that minimises the internal validation error should therefore be chosen instead of the one that minimises the training error. A model with a good prediction performance could thus be repeatedly identified for a static external validation set. The model's prediction results for an external validation set of 100 treatments is presented in Figure 21.

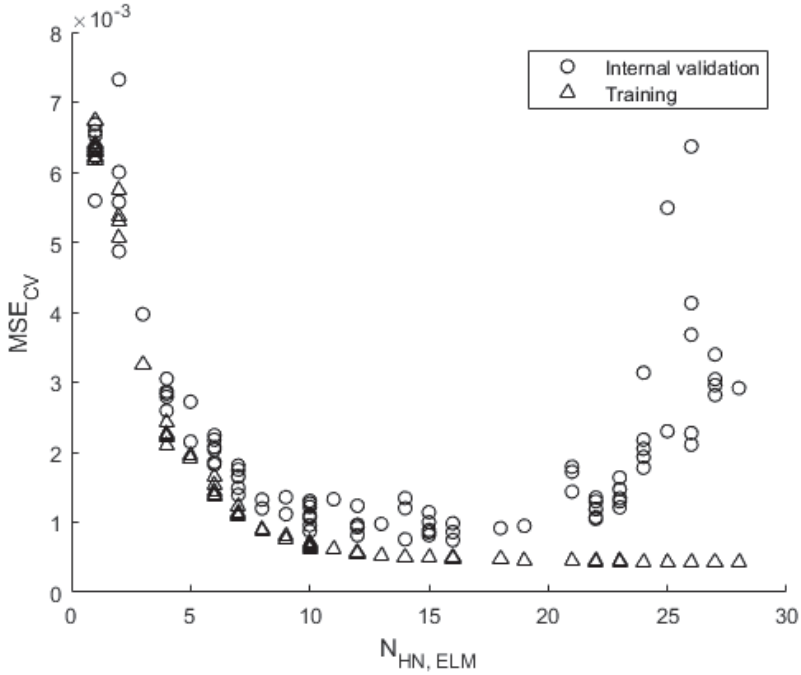

Fig. 20. Exemplified training and internal validation errors as a function of hidden neurons for the end population.

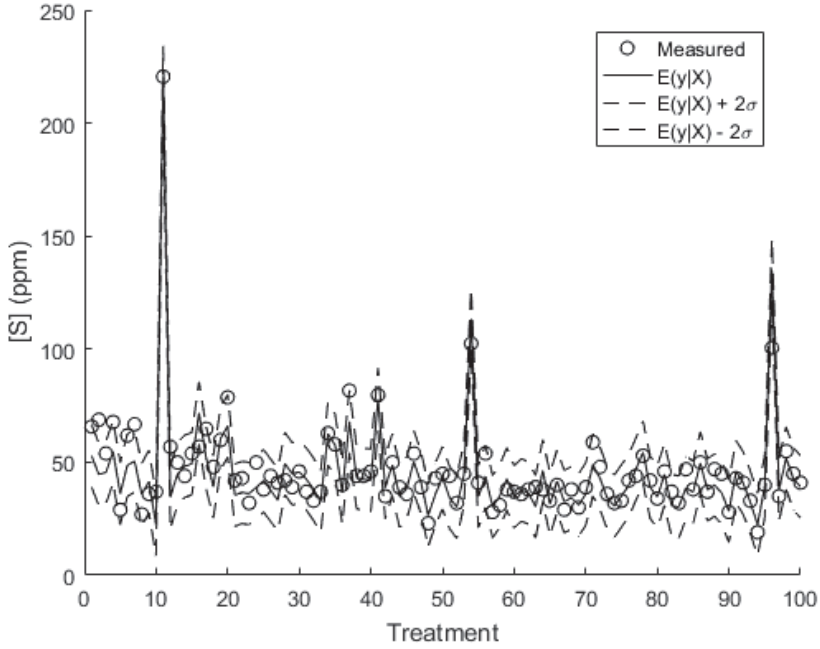

Fig. 21. Predicted end sulphur content at the end of injection. (Reprinted, with permission, from Publication IV © 2020 Elsevier).

Prior to using the algorithm to identify hot metal desulphurisation, it is tested with several open datasets found in the UCI Machine-Learning Repository with a varying number of variable candidates. The results of these tests are beyond the scope of this work, so they are not reported here. However, it is found that the genetic algorithm can select well-performing models, the dimensions of which are significantly smaller than the dimensions of the full model.

In this study, the performance of the algorithm is also compared to Forward Selection 1 and Forward Selection 2, as in the previous section. Prior to selection, the input data is scaled between 0 and 1. The general procedure for comparison is similar to Publication IV and is illustrated in Figure 22**.** When using the deterministic algorithms, the number of hidden neurons is optimised with the grid search from 1 to 30. The performance of the algorithms is compared by repeating the search 50 times with a different external validation data split for each repetition – tournament selection and roulette wheel selection (50/50 selection scheme). The applied GA uses a population of 100 individuals, p_C = 0.9, and a hybrid of the

performance of each model is not found to depend on the algorithm used, because the average R^2 for the external validation set of 100 treatments was 0.83–0.86 and MAE = 5.13–5.20 ppm for each case. However, the selected input variables vary. The equal performance can be explained by the use of the regularisation algorithm in model training, which shrinks the coefficients of the redundant variables close to zero. Thus, each of the studied algorithms performs well in a dimensional reduction that improves the convergence of the Bayesian regularisation training algorithm. A more detailed description of the behaviour of the GA in the variable selection problem is discussed in Publication IV.

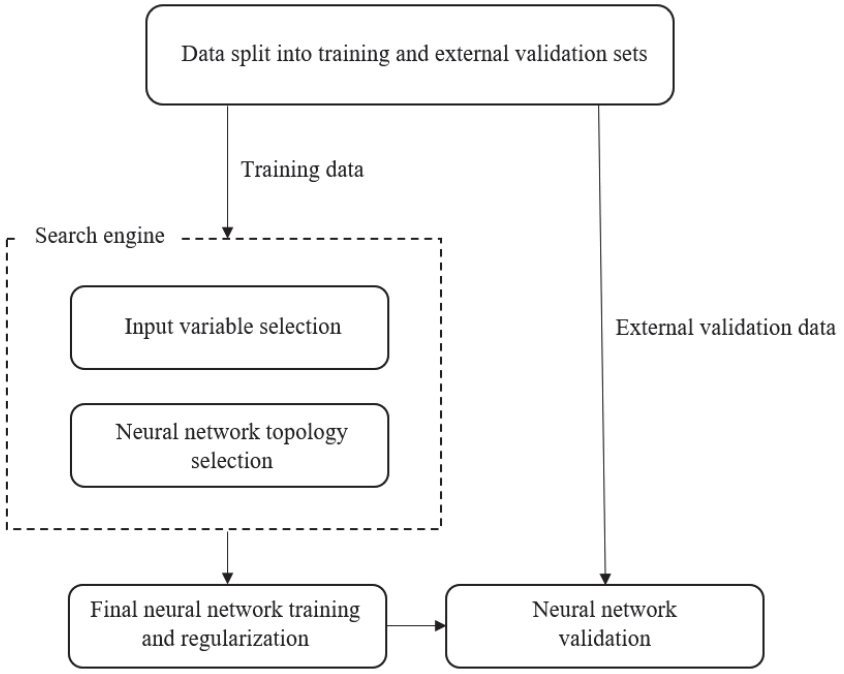

Fig. 22. Procedure for model selection (Reprinted [adapted], with permission, from Publication IV © 2020 Elsevier).

5.6 Kinetic studies on re-sulphurisation of hot metal with experimentally oriented modelling (Publication II)

Resulphurisation is an important factor to consider in industry. Publication II presents a study of the resulphurisation of hot metal. The study finds that the resulphurisation of hot metal can occur, provided the sulphur content of the top slag is sufficiently high. In other words, a thermodynamic potential for resulphurisation and an existing phase contact are needed. The addition of Na_2O to the slag increases the sulphur partition ratio between the slag and metal, regulating the rate of re-sulphurisation. However, the heterogeneous slag complicates the analysis. Assuming the sulphur content in the top slag follows a first-order reaction rate, the change in sulphur content in the slag phase can be written as

$$\frac{\mathrm{d}(\%\mathrm{S})}{\mathrm{d}t} = -k_{\mathrm{tot}}\left((\%\mathrm{S}) - (\%\mathrm{S})_{\mathrm{eq}}\right), \tag{59}$$

where k_{tot} is the rate constant for the re-sulphurisation reaction via a permanent phase contact, (%S) is the sulphur content in the slag phase, and $(\%S)_{eq}$ is the equilibrium sulphur content in the slag phase. The identified rate constants are presented in Table 10. It is also found that the sulphide capacity models presented in the literature do not cover lime-based slags containing Na_2O. To this end, a meta-analysis of the experimental data collected from previous studies is conducted, and the existing models are compared to the collected data. In the analysis, it is found that the Na_2O content significantly increases the sulphide capacity of the slag. However, it is obvious that the sulphide capacity concept is not well defined for heterogeneous slags, although these slags are common in real-life applications. The study reported in Publication II is extended in Section 5.7. to other slag systems.

Table 10. Calculated rate constants for re-sulphurisation, compared to the data of Liu et al. (2015) (Reprinted under CC BY 4.0 license from Publication II © 2019 Authors).

Series	k (1/s)	k' (m/s)	$k'A$ (m^3/s)	$k'A$ (cm^3/s)	$k'A$ (cm^3/s)*	k' (m/s)*
1	0.0045	$2.0 \cdot 10^{-4}$	$3.0 \cdot 10^{-7}$	0.30	0.07	$2.0 \cdot 10^{-5}$
2	0.0044	$1.9 \cdot 10^{-4}$	$2.9 \cdot 10^{-7}$	0.29		
3	0.0033	$1.4 \cdot 10^{-4}$	$2.2 \cdot 10^{-7}$	0.22		
4	0.0022	$9.5 \cdot 10^{-5}$	$1.4 \cdot 10^{-7}$	0.14		

Note: *) Value extracted from the study of Liu et al. (2015).

5.7 Data-driven modelling of the sulphide capacity of the slag phase based on slag-metal equilibrium experiments (extended from Publication II)

In this section, the estimation and explanation of the dependency of slag composition on sulphide capacity based on the slag-metal equilibrium experiments described in Dataset 3 is conducted using meta-analysis. The section first estimates a model based on the data, and then compares the model to the existing models found in the literature. It should be noted that Publication II conducts a similar comparative study for only the CaO-SiO_2-Na_2O system, whereas this section also accounts for the other slag components.

As discussed in Publication II, a relatively large number of sulphide capacity models is available in the literature. The models can be categorised as theoretical models (semi-white box), semi-empirical models (grey box), and data-driven black box models. A more detailed categorisation is given in Publication II**,** so it is not repeated here. However, the general advantage of semi-empirical and data-driven models is that they are simple to implement within a prediction model, whereas theoretical models often require a more rigorous analysis of the model structure during implementation. This comparative study therefore considers only the models that are relatively easy to implement and addresses some of their limitations. The focus is on the most applicable, yet most referenced models found in the literature.

5.7.1 Model selection for the sulphide capacity

The first step of the study considers the identification of two new sulphide capacity models, whose model structure is based either on the neural network or multiple linear regression, using the logistic transformation $\ln(x+1)$ for input and output data. The explanatory variable candidates are selected as the molar fractions of the slag components, and the temperature of the system is expressed in Kelvins. The output variable is the estimated sulphide capacity, which is calculated as described in Section 4.3.

Model selection is divided into data pre-treatment, variable selection, and final model training steps. Prior to variable selection, the data is split into training and external validation sets such that 20% is used for external validation. The training set is further divided into training and internal validation sets. Variable selection is conducted using Forward Selection 1, and the sum of the squared error for the

internal validation set over 8*N* data splits is used as the objective function to yield a realistic estimate of the internal validation error. After selection, the model parameters are estimated using the Moore-Penrose inversion for the training data. The final model performance is then evaluated, using the external validation set.

The selection is repeated several times for different training and external validation splits, and the results are found to be consistent between splits. As the number of input variable candidates is relatively small, the double cross-validation is not used. It is found that when using the MLR model as the basis for selection, Forward Selection 1 tends to select all the candidate variables except the P_2O_5 for the model. However, this is not considered an issue, because the objective function is relatively pessimistic, being based on around 3,500 data splits in the internal validation step. As Figure 23 shows, forward selection converges to a set of 9 variables, indicating a rather monotonic objective function with respect to the variable space. However, as the figure shows, the last step results in a rather small improvement, and two elimination steps are therefore performed to reduce the variance of the model, yielding a total of 7 input variables, including molar fractions of CaO, SiO_2, Na_2O, Al_2O_3, and MgO.

The input variables selected with the MLR basis and the data division is used for the training and external validation of the network. In training, the number of hidden neurons is kept as small as possible to reduce the variance of the model. The number of hidden neurons is therefore set to 3, and the number of hidden layers to 1. The neural network model is trained with the Levenberg-Marquardt algorithm. Prior to training, the input data is treated with standardisation ($\mu = 0$, $\sigma^2 = 1$), in which for each column the mean is subtracted, and the result is divided by the standard deviation.

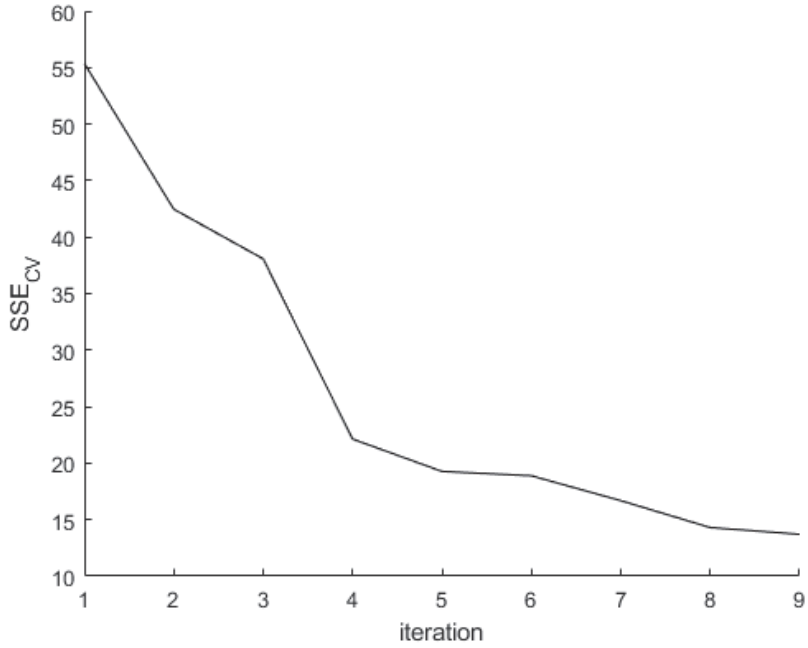

Fig. 23. Convergence of Forward Selection 1 for the sulphide capacity data.

5.7.2 Analysis of the considered models

The prediction performance indicators for the models selected based on Dataset 3 are found to be R^2 = 0.89 and MAE = 0.24 (-) for the MLR model, and R^2 = 0.94 and MAE = 0.17 (-) for the ANN model, determined for the external validation dataset. The figures of merit for the models and their comparison are given in Table 11.

Table 11. Figures of merit for the compared models.

Model	R^2-full	MAE-full	R^2-t.	MAE-t.	R^2 – ev.	MAE-ev.
This study - MLR	0.89	0.24	0.89	0.25	0.89	0.23
This study - ANN	0.94	0.16	0.94	0.16	0.95	0.17
Sosinsky and Sommerville (1986)	0.57	1.59				
Young et al. (1992)	0.77	0.59				
Zhang et al. (2013)	0.68	0.61				

This study's model predictions are found to be in reasonable agreement with the predictions carried out with earlier models. The largest deviation was found to be with the model of Sosinsky and Sommerville (1986). This is mainly because the temperature range of Dataset 3 is between 1200 °C and 1700 °C, which is considerably distant from the range of validity in Sosinsky and Somerville (1986), which is between 1400 °C and 1700 °C. Second, as the data based on which the model is identified does not contain any Na_2O, whose high sulphide capacity has been widely discussed in the literature (Chan & Fruehan 1986; Chan & Fruehan 1989; van Niekerk & Dippenaar 1993; Kunisada & Iwai 1993), the model is assumed to underestimate the sulphide capacities for the slag systems containing relatively high amounts of Na_2O. The model of Zhang et al. (2013) is similar to that proposed by Sosinsky and Sommerville (1986), but is fitted to a larger and more recent set of experimental data (Zhang et al. 2013). However, the model of Zhang et al. (2013) does not cover the Na_2O-containing slag systems, which consequently results in an underestimating of the sulphide capacity of the corresponding slag systems. The best agreement is found to be with the model of Young et al. (1992). However, this also suffers from the lack of data from the Na_2O-containing slags. Figure 24 illustrates the predictions made for the full data with the selected MLR model and the model of Young et al. (1992). The figure shows that the predictions are consistent except for the negative deviation from the diagonal of the model of Young et al. (1992). It should be noted that the consistent predictions for systems other than those with Na_2O support the assumptions made in the feature generation step, i.e. in the computation of the equilibrium activity of oxygen and sulphide capacity.

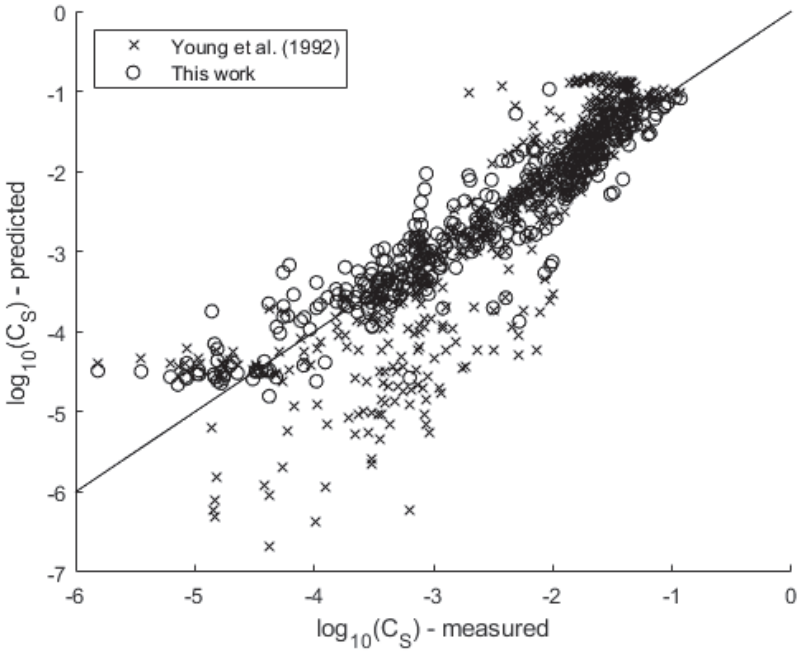

Fig. 24. Comparison of the MLR model with the model of Young et al. (1992).

The residual plots for the compared models presented in Figure 25 show that the residuals are positively biased and correlate with the molar fraction of Na_2O in the slag phase. This basically indicates that the compared models suffer from a lack of fit because of the absence of the effect of Na_2O. This arises from the definition of optical basicity: as there are several slag compositions with the same optical basicity, the dependencies of the optical basicity and the sulphide capacity are assumed to be of a higher order than for linear. The non-linearity problem is addressed by Young et al. (1992), who add the squared terms, which still does not fix the problem that the coefficient estimates deviate towards the slag compositions without Na_2O, resulting in the underestimation.

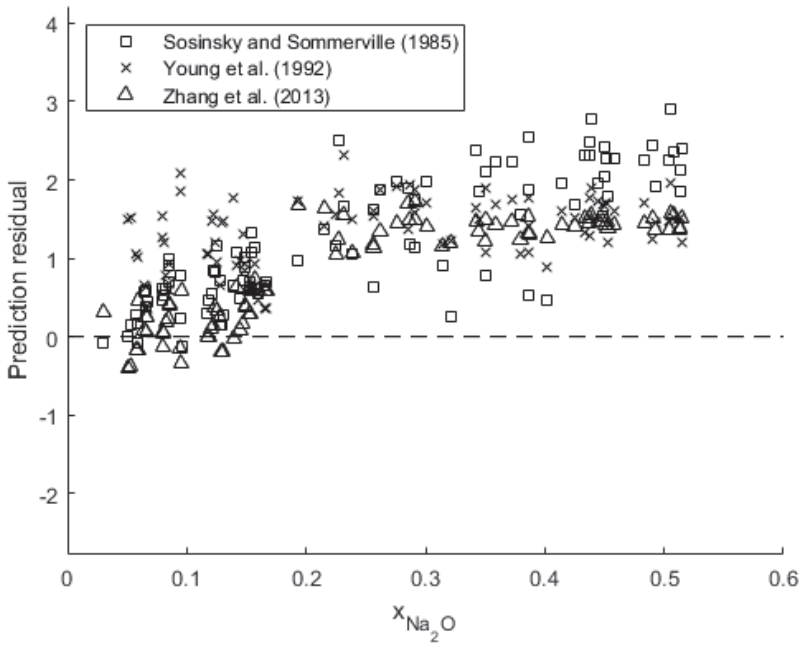

Fig. 25. The residual plots with the analysed models with respect to molar fraction Na_2O. The *x*-axis is marked with the dashed line.

In the second model selection step, the hypothesis concerning the effect of slag components on the sulphide capacity is tested with the linear model basis. The testing is conducted by drawing 100,000 bootstrap samples from the training data and taking the standard deviation of the estimated coefficients as the standard error. Further, the confidence interval is determined with the quantile method, and it is found that all the coefficients differ significantly from 0, and hence the bootstrap *p*-values are practically infinitely close to 0. The estimated distributions for the standardised regression coefficients of CaO, Na_2O, SiO_2, and Al_2O_3 are presented in Figure 26. The figure shows that the empirical distributions of the estimates are nearly normal, and the full data estimate falls near the mean of the distribution. The standard error of the estimates shows that the observed coefficient values differ significantly from zero, which can be regarded as sufficient evidence for rejecting H_0. The expected value of the coefficient estimates of Na_2O and CaO, shows that the ratio ($\approx 1/3$) for the sulphide capacity is very near the Na_2O equivalent value (0.3), determined for the sulphur partition ratio reported by Van Niekerk and

Dippenaar (1993) and Inoue and Suito (1982). In practice, this means that when adding the same amount of Na_2O and CaO, the magnitude of the positive effect is 3 times larger for Na_2O under the compositional ranges considered in this study. In addition, there seem to be no systematic errors in the variable construction stage, because the estimates determined for the partition ratio and sulphide capacity agree. The effects of SiO_2 and Al_2O_3 seem to be the opposite of that of Na_2O and CaO. The direction of the effects agrees with the model proposed by Young et al. (1992), who similarly estimated the negative effect of SiO_2 and Al_2O_3 on sulphide capacity. With reference to the bootstrap statistics, all the estimated effects seem to be reliable, and the probability of making a false interpretation based on chance given this data is very small.

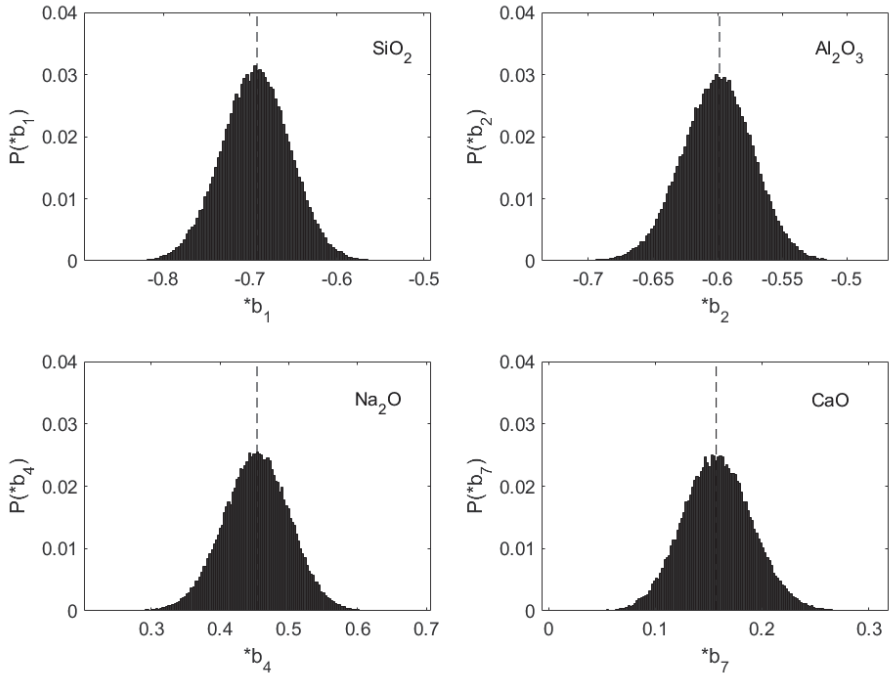

Fig. 26. Empirical bootstrap distributions of the linear model estimates of the main slag components to sulphide capacity.

With reference to the poor extrapolation properties of the neural network, in practice, the linear model is more applicable in predictive use. This is because the composition of the hot metal desulphurisation slag is well beyond the confidence

region of the neural network trained in this data, mainly because the CaO contents in the slag are excessively high, especially during the end stage of injection, and the amount of non-dissolved CaO is relatively high. With reference to the large variance of the network models, the models tend to be unstable outside the region of the training data. In practice, only the interpolation properties of the network can therefore be guaranteed, and the extrapolation with a network should be evaluated carefully (Hagan et al. 1997). In other words, to use the neural network model in predictive inference, more data must be collected from the sulphide capacities or effective sulphide capacities of slags with a high CaO content (> 60%). Unfortunately, this data is not available in the literature, because sulphide capacity was originally formulated for liquid slags, which limits the use of the concept in analysing the sulphur extraction capacity of slags with high solid fractions. This issue has also been outlined in Publication II in the context of $CaO\text{-}SiO_2\text{-}Na_2O$ slag systems. Considering the reasoning above, the linear model is chosen within the permanent contact reaction model presented in Section 5.3.

It should be noted that the KTH model is intentionally omitted from this analysis, because the number of model parameters is very large compared to the degrees of freedom available for the standard error estimates in this data. A detailed description can be found in Nzotta et al. (1998), for example. As is pointed out in the study of Zhang et al. (2015), the number of parameters in the KTH model is more than 60 for an 8-component system, taking the binary and ternary interactions of the cations and the temperature into account. Similarly, Publication II points out that the number of parameters in a ternary system is eight, but also that it would be beneficial to re-estimate the interaction parameters as new data is provided, because it is obvious that the parameter estimates are biased if the considered model structure is very far from the best available model. Further, it would be beneficial to consider whether the parameters that consider the interactions of the cations can be estimated from the data, and not to select the model structure based on the assumed interactions, but on the observed ones. However, the latter approach is not typical, because nearly every model considers the higher order interactions before considering the identification on a linear model basis. As has been seen for this data, the linear model structure is arguably sufficient, and more stable for out-of-sample predictions referring to smaller variance in practice. Indeed, model selection is also conducted such that the data includes the higher order interactions. This results in a set of 55 candidate variables, which makes the use of $8N$ repetitions in the inner loop impractically slow. In addition, model performance is not found to be significantly improved, but it is found to be increasingly unstable within the

increase in complexity. Instability is observed in inconsistent training and external validation errors, because the external validation error is found to be systematically several orders larger than the training error, which basically indicates an overly complex model.

Concerning simulation model selection, the presented procedure can be seen as a three-step model construction. In the first step, the sulphide capacity values are generated from the phase compositions and the system conditions, involving multiple variable constructions for both the input and output variable spaces. In the second step, a model for predicting sulphide capacity is selected that best fits this data, whereas in the third step, the selected model is used in parallel with a mechanistic model to provide the boundary condition for the expression of the *rate of a permanent contact reaction*.

5.8 Further work

In this study, different mathematical modelling strategies are employed to study the hot metal desulphurisation process. The models derived for lime-based hot metal desulphurisation allow the identification of process causalities and a quantitative description of the effect of available variables on process efficiency. To extend practical usability, the models can be coupled with economic variables such as costs of reagent materials and in-depth cost optimisation. In addition, modelling strategies can be used as the basis for the data-driven identification of the still non-quantifiable phenomena. As the results show, the effect of top slag on the rate and efficiency can be described only semi-quantitatively, because Dataset 1 does not explicitly reveal the effect. To achieve this, a more comprehensive dataset describing the covariance of the slag and the metal phase compositions would have to be collected, including the information on the variables that relate to the rate of the transitory reaction. In addition, a dynamically evolving dataset would plausibly allow a more detailed analysis of the individual rate mechanisms.

Concerning process modelling in general, this study offers guidelines for model selection in the context of hot metal desulphurisation. As seen throughout the study, other processes such as the basic oxygen furnace and the blast furnace have been well covered in the previous literature. Publication V and this study show that data-driven techniques in general benefit from more data. However, the characteristics of high-temperature processes and the workload needed for analysis procedures often restrict this aim, and thus the use of data-driven techniques. For example, it is laborious to collect datasets describing the particle size distributions

of lime-based reagents, which may limit the use of suggested online models. However, a pre-trained model can be applied to find the best practices and systematically control the relevant process variables if used online.

6 Conclusions

This study's objective was to study hot metal desulphurisation with mathematical and experimental modelling, emphasising the data-driven identification of both parameterised phenomena-based models and a selection of generic prediction models. In the literature study, it was found that the most common mathematical models proposed for the process fell into the category of mechanistic and CFD models. Data-driven models are a minority in the literature, and the use of systematic modelling frameworks for process identification have therefore not been previously studied.

An explanatory analysis of the primary hot metal desulphurisation data revealed that the estimated rate parameters, i.e. the residence time and the mass transfer coefficient, greatly depended on the assumptions, as well as the applied particle size distribution parameters. However, the effect of particle size distribution was found to be similar, regardless of the applied distribution parameter. Indeed, the information concerning particle size distribution was found to explain a major part of the variance and other properties of the reagent, namely the amount of injected auxiliary compounds. However, the role of injection parameters – the loading of the reagent, for example – was not observed or studied.

The predictive performance of the models was found to be sufficient in all cases. However, as measured particle size distribution is not readily available in the full-scale process, and there are some fundamental uncertainties concerning this attribute's use in the full-scale process, the full-scale predictive performance of the models remains to be exhaustively validated.

The contribution of the slag phase to hot metal desulphurisation was twofold. On one hand, for these cases, the injection parameters and reagent properties provided sufficient explanatory and predictive power. The contribution of the slag phase was estimated to be 10–20% to the overall rate if the extraction capacity of the reagent was not assumed to be of infinite magnitude. If surface area approximation was used as the description for the transitory contact reaction, the contribution of the permanent phase contact could be neglected. The suitable model parameter values were not found to be unambiguous. Indeed, the relative contributions and the valid parameter ranges greatly depended on the applied descriptions of the individual rate mechanisms.

Due to the contribution of the slag phase, it was concluded by using a meta-analysis of previous studies that the used description of the sulphide capacity was important. The importance of the description was twofold. On one hand, a realistic

description provided a sufficiently realistic boundary condition for the rate of the permanent contact reaction, but it might also help to estimate the possibility of the re-sulphurisation of hot metal in a production situation, and thus assist in the selection of slag modifiers to avoid the phenomenon. The effect of Na_2O on these was found to be important. As the increase in the Na_2O content of the slag phase increases its sulphide capacity, it can be postulated that it increases the thermodynamic driving force for desulphurisation but regulates the driving force of the inverse reaction. As the addition of Na_2O increases the equilibrium liquid phase fraction of the slag phase, it is assumed that slag modification contributes to rate mechanisms with an increased mass transfer rate.

In the case of all the studied datasets, it was found that automated model selection strategies provided results that were consistent with domain knowledge, provided that the algorithm hyperparameters were carefully chosen. Genetic algorithms were found to be suitable for the identification of process models, both in variable selection and model training. Indeed, in variable selection for Dataset 1, the results of the genetic algorithm were found to be comparable with the results of the exhaustive search. Deterministic forward variable selection strategies could be said to perform as well as metaheuristic algorithms, especially if the dimensionality of the best available model was significantly smaller than the original dimension, and if the objective function was monotonic. However, the importance of choosing a proper objective function for model selection cannot be exaggerated. It is therefore recommended that these strategies be used concurrently with domain knowledge (if available), not exclusively.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723.

Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, *46*(3), 175–185.

Anderson, T. W., & Darling, D. A. (1952). Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *The Annals of Mathematical Statistics*, 193–212.

Arumugam, M. S., & Rao, M. V. C. (2004). Novel hybrid approaches for real coded genetic algorithm to compute the optimal control of a single stage hybrid manufacturing systems. *International Journal of Computational Intelligence*, *1*(3), 189–206.

Arumugam, M. S., Rao, M. V. C., & Palaniappan, R. (2005). New hybrid genetic operators for real coded genetic algorithm to compute optimal control of a class of hybrid systems. *Applied Soft Computing*, *6*(1), 38–52.

Ban-Ya, S. (1993). Mathematical expression of slag-metal reactions in steelmaking process by quadratic formalism based on the regular solution model. *ISIJ International*, *33*(1), 2–11.

Barbosa, B. H., Aguirre, L. A., Martinez, C. B., & Braga, A. P. (2010). Black and gray-box identification of a hydraulic pumping system. *IEEE Transactions on Control Systems Technology*, *19*(2), 398–406.

Basu, S., Lahiri, A. K., & Seetharaman, S. (2008). Activity of iron oxide in steelmaking slag. *Metallurgical and Materials Transactions B*, *39*(3), 447–456.

Baumann, K. (2003). Cross-validation as the objective function for variable-selection techniques. *TrAC Trends in Analytical Chemistry*, *22*(6), 395–406.

Baumann, K. (2005). Chance correlation in variable subset regression: influence of the objective function, the selection mechanism, and ensemble averaging. *QSAR & Combinatorial Science*, *24*(9), 1033–1046.

Baumann, D. , & Baumann, K. (2014). Reliable estimation of prediction errors for QSAR models under model uncertainty using double cross-validation. *Journal of Cheminformatics*, *6*(1), 47.

Bayat, H., Rastgo, M., Zadeh, M. M., & Vereecken, H. (2015). Particle size distribution models, their characteristics and fitting capability. *Journal of Hydrology*, *529*, 872–889.

Bhattacharya, T., Nag, S., & Lenka, S. N. (2004). Analysis of DS reagent consumption using multivariate statistical modeling. *Tata Search*, *1*, 215–223.

Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, *26*(2), 211–243.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32.

Breiman, L. (2001). Statistical modeling: The two cultures. *Statistical Science*, *16*(3), 199–215.

Brooks, G. A., Rhamdhani, M. A., Coley, K. S., & Pan, Y. (2009). Transient kinetics of slag metal reactions. *Metallurgical and Materials Transactions B*, *40*(3), 353–362.

Burnham, D.R., & Anderson, K.P., (2002). *Model selection and multimodel inference: a practical information-theoretic approach*. New York: Springer-Verlag.

Bäck, T., & Schütz, M. (1996). Intelligent mutation rate control in canonical genetic algorithms. In *International Symposium on Methodologies for Intelligent Systems* (pp. 158–167). Berlin, Heidelberg: Springer.

Chan, A. H., & Fruehan, R. J. (1986). The sulfur partition ratio and the sulfide capacity of Na_2O-SiO_2 slags at 1200° C. *Metallurgical Transactions B*, *17*(3), 491–496.

Chan, A. H., & Fruehan, R. J. (1989). The sulfur partition ratio with Fe-C SAT melts and the sulfide capacity of CaO-SiO_2-Na_2O-(Ai_2O_3) slags. *Metallurgical Transactions B*, *20*(1), 71–76.

Chatfield, C. (1995). Model uncertainty, data mining and statistical inference. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *158*(3), 419–444.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785–794). Retrieved from: https://dl.acm.org/doi/abs/10.1145/2939672.2939785.

Chiang, L. K., & Irons, G. (1990). Kinetic Studies of Calcium Carbide Hot Metal Desulphurization by Powder Injection. *Iron Steelmaker*, *17*(1), 35–52.

Choi, J. Y., Kim, D. J., & Lee, H. G. (2001). Reaction kinetics of desulfurization of molten pig iron using CaO–SiO2–Al2O3–Na2O slag systems. *ISIJ International*, *41*(3), 216–224.

Clift, R., Grace, J. R., & Weber, M. E. (1978). *Bubbles, Drops, and Particle*s. New York: Academic Press.

Coudure, J., & Irons, G. A. (1994). The Effect of Calcium Carbide Particle Size Distribution on the Kinetics of Hot Metal Desulphurization. *ISIJ International*, *34*(2), 155–163.

Dan, B. Chen, K., Xiong, L., Rong Z., & Yi, J., (2008). Research on multi-BP NN-based control model for molten iron desulfurization. In *Proceedings of the 7th World Congress on Intelligent Control and Automation*, (pp. 6133–6137). China, Chonqing: IEEE.

Datta, A., Hareesh, M., Kalra, P. K., Deo, B., & Boom, R. (1994). Adaptive neural net (ANN) models for desulphurization of hot metal and steel. *Steel Research*, *65*(11), 466–471.

Davison, A. C. & Hinkley, D. V. (1997). *Bootstrap Methods and their Application* (No. 1). Cambridge University Press.

Deep, K., Singh, K. P., Kansal, M. L., & Mohan, C. (2009). A real coded genetic algorithm for solving integer and mixed integer optimization problems. *Applied Mathematics and Computation*, *212*(2), 505–518.

Deng, J., & Oeters, F. (1990). Mass transfer of sulfur from liquid iron into lime-saturated CaO-Al2O3-MgO-SiO2 slags. *Steel Research*, *61*(10), 438–448.

Deo, B., & Boom, R. (1993). *Fundamentals of Steelmaking Metallurgy*. Prentice-Hall.

Deo, B., Datta, A., Kukreja, B., Rastogi, R., & Deb, K. (1994). Optimization of back propagation algorithm and GAS-assisted ANN models for hot metal desulphurization. *Steel Research*, *65*(12), 528–533.

Dudzic, M., & Zhang, Y. (2004). On-line industrial implementation of process monitoring/control applications using multivariate statistical technologies: Challenges and opportunities. In *Dynamics and Control of Process Systems* (pp. 269–279). Massachusetts: USA, Elsevier (IFAC Publications).

Farias, L. R., & Irons, G. A. (1985). A unified approach to bubbling-jetting phenomena in powder injection into iron and steel. *Metallurgical Transactions B*, *16*(2), 211–225.

Farias, L. R., & Irons, G. A. (1986). A multi-phase model for plumes in powder injection refining processes. *Metallurgical Transactions B*, *17*(1), 77–85.

Feng, K., Xu, A. J., He, D. F., & Yang, L. (2019). Case-based reasoning method based on mechanistic model correction for predicting endpoint sulphur content of molten iron in KR desulphurization. *Ironmaking & Steelmaking*, 1–8.

Fetters K. L. & Chipman J., 1941. Equilibria of Liquid Iron and Slags of the System CaO-MgO-FeO-SiO2. *Transactions of the American Institute of Mining Engineers*, *145*, 95–112.

Foresee, F. D. & Hagan, M. T. (1997, June). Gauss-Newton approximation to Bayesian learning. In *Proceedings of International Conference on Neural Networks (ICNN'97)* (Vol. 3, pp. 1930–1935). IEEE.

Freedman, D. A., & Freedman, D. A. (1983). A note on screening regression equations. *The American Statistician*, *37*(2), 152–155.

Geladi, P., & Kowalski, B. R. (1986). Partial least-squares regression: a tutorial. *Analytica chimica acta*, *185*, 1–17.

Gharahbagh, A. A., & Abolghasemi, V. (2008). A novel accurate genetic algorithm for multivariable systems. *World Applied Sciences Journal*, *5*(2), 137–142.

Ghiaus, C., Chicinas, A., & Inard, C. (2007). Grey-box identification of air-handling unit elements. *Control Engineering Practice*, *15*(4), 421–433.

Ghosh, H., Iquebal, M. A., & Prajneshu (2011). Bootstrap study of parameter estimates for nonlinear Richards growth model through genetic algorithm. *Journal of Applied Statistics*, *38*(3), 491–500.

Gilmour, S. G. (1996). The interpretation of Mallows's Cp-statistic. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *45*(1), 49–56.

Goldberg, D. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. Boston, MA: Addison-Wesley.

Goldberg, D. E., Korb, B., & Deb, K. (1989). Messy genetic algorithms: Motivation, analysis, and first results. *Complex Systems*, *3*(5), 493–530.

Görnerup, M., & Sjoberg, P. (1999). AOD/CLU process modelling: Optimum mixed reductants addition. *Ironmaking & Steelmaking*, *26*(1), 58.

Grant N. J., Kalling U., & Chipman J. (1951). Effects of Manganese and Its Oxide on Desulphurization by Blast-Furnace Type Slags. *Transactions of the American Institute of Mining Engineers*, 191, 666–671.

Guyon, I. & A. Elisseeff, (2003). An Introduction to Variable and Feature Selection, *Journal of Machine Learning Research*, *3*, 1157–1182.

Guyon, I. & Elisseeff, A. (2006). An introduction to feature extraction. In *Feature Extraction* (pp. 1–25). Berlin, Heidelberg: Springer.

Hagan, M. T., & Menhaj, M. B. (1994). Training feedforward networks with the Marquardt algorithm. *IEEE Transactions on Neural Networks*, *5*(6), 989–993.

Hagan, M. T., Demuth, H. B., & Beale, M. (1997). *Neural Network Design*. PWS Publishing Co.

Han, J., Pei, J., & Kamber, M. (2011). *Data Mining: Concepts and Techniques*. Elsevier.

Harrell, F. E. (2015). *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. Springer.

Hastie, T., Rosset, S., Zhu, J., & Zou, H. (2009). Multi-class adaboost. *Statistics and its Interface*, *2*(3), 349–360.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Science & Business Media.

Hatch G., & Chipman J., 1949. Sulphur Equilibria Between Iron Blast Furnace Slags and Metal. *Transactions of the American Institute of Mining Engineers*, *185*, 274–284.

Haykin, S. (1994). *Neural Networks: A Comprehensive Foundation*. Prentice Hall PTR.

Heinze, G., Wallisch, C., and Dunkler, D. (2018). Variable selection–A review and recommendations for the practicing statistician. *Biometrical Journal*, *60*(3), 431–449.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, *2*(5), 359–366.

Huang, G. B., Zhu, Q. Y., and Siew, C. K. (2006). Extreme learning machine: Theory and applications. *Neurocomputing*, *70*(1–3), 489–501.

Huang, J., Cai, Y., & Xu, X. (2007). A hybrid genetic algorithm for feature selection wrapper based on mutual information. *Pattern Recognition Letters*, *28*(13), 1825–1844.

Hurvich, C. M., & Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika*, *76*(2), 297–307.

Hurvich, C. M., & Tsai, C. L. (1995). Model selection for extended quasi-likelihood models in small samples. *Biometrics*, 1077–1084.

Inoue, R., & Suito, H. (1982). Sulfur Partitions between Carbon-saturated Iron Melt and Na_2O-SiO_2 Slags. *Transactions of the Iron and Steel Institute of Japan*, *22*(7), 514–523.

Irons, G. A. (1988). The Principles of Gas and Powder Injection for Iron Refining. In *Foundry Processes* (pp. 303–331). Boston, MA: Springer.

Irons, G. A. (1989). Role of mixing in powder injection desulphurization processes. *Ironmaking & Steelmaking*, *16*(1), 28–36.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning*, New York: Springer.

Jolliffe, I. T. (1982). A note on the use of principal components in regression. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *31*(3), 300–303.

Katare, S., Bhan, A., Caruthers, J. M., Delgass, W. N., & Venkatasubramanian, V. (2004). A hybrid genetic algorithm for efficient parameter estimation of large kinetic models. *Computers & Chemical Engineering*, *28*(12), 2569–2581.

Khalik, M. A., Sherif, M., Saraya, S., & Areed, F. (2007). Parameter identification problem: Real-coded GA approach. *Applied Mathematics and Computation*, *187*(2), 1495–1501.

Kohavi, Ron., (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection, In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, (pp. 1137–1145), San Francisco, United States: ACM Digital Library.

Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, *97*(1–2), 273–324.

Kreyszig, E. (2006). *Advanced Engineering Mathematics*. Columbus: John Wiley & Sons.

Kunisada, K., & Iwai, H. (1993). Effects of CaO, MnO, MgO and Al_2O_3 on the Sulfide Capacities of Na_2O-SiO_2 Slags. *ISIJ International*, *33*(1), 43–47.

Leardi, R., Seasholtz, M. B., & Pell, R. J. (2002). Variable selection for multivariate calibration using a genetic algorithm: prediction of additive concentrations in polymer films from Fourier transform-infrared spectral data. *Analytica Chimica Acta*, *461*(2), 189–200.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.

Lee, J., & Morita, K. (2004). Dynamic interfacial phenomena between gas, liquid iron and solid CaO during desulphurization. *ISIJ International*, *44*(2), 235–242.

Levy, A. Y., Iwasaki, Y., & Fikes, R. (1997). Automated model selection for simulation based on relevance reasoning. *Artificial Intelligence*, *96*(2), 351–394.

Lin, B., Recke, B., Knudsen, J. K., & Jørgensen, S. B. (2007). A systematic approach for soft sensor development. *Computers & Chemical Engineering*, *31*(5–6), 419–425.

Lindström, D., Nortier, P., & Sichen, D. (2014). Functions of Mg and Mg–CaO Mixtures in Hot Metal Desulphurization. S*teel Research International*, *85*(1), 76–88.

Lindström, D., & Sichen, D. (2015). Kinetic study on desulfurization of hot metal using CaO and CaC 2. *Metallurgical and Materials Transactions B*, *46*(1), 83–92.

Lindström, D., & Sichen, D. (2015). Study on desulfurization abilities of some commonly used desulfurization agents. *steel research international*, *86*(1), 73–83.

Liu, C., Huang, F., Wang, X., & Yang, G. (2015). Study on the kinetic mechanism of steel resulfurization by sulfur-bearing slag. *Metallurgical Research & Technology*, *112*(4), 408.

Loibel, S., do Val, J. B., & Andrade, M. G. (2006). Inference for the Richards growth model using Box and Cox transformation and bootstrap techniques. *Ecological Modelling*, *191*(3–4), 501–512.

Ma, W., Li, H., Cui, Y., Chen, B., Liu, G., & Ji, J. (2017). Optimization of desulphurization process using Lance injection in molten iron. *ISIJ International*, 57(2), 214–219.

Mallows, C. L. (2000). Some comments on Cp. *Technometrics*, *42*(1), 87–94.

MacKinnon, J. G. (2009). Bootstrap hypothesis testing. *Handbook of Computational Econometrics*, *183*, 213.

Michalewicz, Z. (1992). *Genetic Algorithms + Data Structures = Evolution Programs*. Springer Science and Business Media.

Mäkinen, R., Neittaanmäki, P., Périaux, J., & Toivanen, J. (1998). A genetic algorithm for multiobjective design optimization in aerodynamics and electromagnetics. *Computational Fluid Dynamics*, *98*, 418–422.

Nakano, M. & Ito, K. (2016). Three Dimensional Simulation of Lime Particle Penetration into Molten Iron Bath Using Smoothed Particle Hydrodynamics. *ISIJ International*, 56(9), 1537–1542.

Nyarko, E. K., & Scitovski, R. (2004). Solving the parameter identification problem of mathematical models using genetic algorithms. *Applied Mathematics and Computation*, *153*(3), 651–658.

Nzotta, M., Sichen, D., & Seetharaman, S. (1998). Sulphide capacities in some multi component slag systems. *ISIJ International*, *38*(11), 1170–1179.

Oberkampf, W. L., & Trucano, T. G. (2002). Verification and validation in computational fluid dynamics. *Progress in Aerospace Sciences*, *38*(3), 209–272.

Odom, M. D., & Sharda, R. (1990, June). A neural network model for bankruptcy prediction. In *1990 IJCNN International Joint Conference on Neural Networks* (pp. 163–168). IEEE.

Oeters, F. (1994). *Metallurgy of Steelmaking*, Düsseldorf: Verlag Stahlheisen.

Oreski, S., & Oreski, G. (2014). Genetic algorithm-based heuristic for feature selection in credit risk assessment. *Expert Systems with Applications*, *41*(4), 2052–2064.

Pacek, A. W., Man, C. C., & Nienow, A. W. (1998). On the Sauter mean diameter and size distributions in turbulent liquid/liquid dispersions in a stirred vessel. *Chemical Engineering Science*, *53*(11), 2005–2011.

Pal, U. B., & Patil, B. V. (1986). Role of Dissolved Oxygen in Hot Metal Desulphurization. *Ironmaking & Steelmaking*, *13*(6), 294–300.

Pettersson, F., Chakraborti, N., & Saxén, H. (2007). A genetic algorithms based multi-objective neural net applied to noisy blast furnace data. *Applied Soft Computing*, *7*(1), 387–397.

Pudil, P., Novovičová, J., & Kittler, J. (1994). Floating search methods in feature selection. *Pattern Recognition Letters*, *15*(11), 1119–1125.

Quinn, S. L., & Vaculik, V. (2002). Improving the desulfurization process using adaptive multivariate statistical modeling. *AISE Steel Technology (USA)*, *79*(10), 37–41.

Rasmussen, C. E., & Ghahramani, Z. (2001). Occam's razor. In *Advances in Neural Information Processing Systems* (pp. 294–300). Cambridge, MA, USA: MIT Press.

Rastogi, R., Deb, K., Deo, B., & Boom, R. (1994). Genetic adaptive search model of hot metal desulphurization. *Steel Research*, *65*(11), 472–478.

Riboud, P. V., & Olette, M. (1982). Mechanisms of some of the reactions involved in secondary refining. In *Seventh International Conference on Vacuum Metallurgy: Special Meltings and Metallurgical Coatings,* (Vol. 2, pp. 879–889). Tokyo, Japan: Iron and Steel Institute of Japan.

Sahai, Y., & Guthrie, R. I. L. (1982). Hydrodynamics of gas stirred melts: Part I. Gas/liquid coupling. *Metallurgical Transactions B*, *13*(2), 193–202.

Sargent, R. G. (1988, December). A tutorial on validation and verification of simulation models. In *Proceedings of the 20th Conference on Winter Simulation* (pp. 33–39).

Sargent, R. G. (2010). Verification and validation of simulation models. In *Proceedings of the 2010 Winter Simulation Conference* (pp. 166–183). Baltimore, MD, USA: IEEE.

Saxén, H., & Pettersson, F. (2007). Nonlinear prediction of the hot metal silicon content in the blast furnace. *ISIJ International*, *47*(12), 1732–1737.

Schrama, F. N. H., Beunder, E. M., Van den Berg, B., Yang, Y., & Boom, R. (2017). Sulphur removal in ironmaking and oxygen steelmaking. *Ironmaking & Steelmaking*, *44*(5), 333–343.

Schrama, F. N., Ji, F., Hunt, A., Beunder, E. M., Woolf, R., Tuling, A., Warren, P., Sietsma, J., Boom, R., & Yang, Y. (2020). Lowering iron losses during slag removal in hot metal desulphurisation without using fluoride. *Ironmaking & Steelmaking*, 1–9.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464.

Seshadri, V., Da Silva, C. A., Da Silva, I. A., & von Krüger, P. (1997). A kinetic model applied to the molten pig iron desulfurization by injection of lime-based powders. *ISIJ International*, *37*(1), 21–30.

Siedlecki, W., & Sklansky, J. (1993). A note on genetic algorithms for large-scale feature selection. In *Handbook of Pattern Recognition and Computer Vision* (pp. 88–107). New Jersey, US: Word Scientific Publishing Co.

Shao, J. (1993). Linear model selection by cross-validation. *Journal of the American Statistical Association*, *88*(422), 486–494.

Shmueli, G. (2010). To explain or to predict? *Statistical Science*, *25*(3), 289–310.

Sigworth, G. K., & Elliott, J. F. (1974). The thermodynamics of liquid dilute iron alloys. *Metal Science*, *8*(1), 298–310.

Sjöberg, J., Zhang, Q., Ljung, L., Benveniste, A., Deylon, B., Glorennec, P. Y., & Juditsky, A. (1995). *Nonlinear black-box modeling in system identification: A unified overview*. Automatica, 31(12), 1691–1724.

Sohlberg, B., & Jacobsen, E. W. (2008). Grey Box modelling – branches and experiences. *IFAC Proceedings Volumes*, *41*(2), 11415–11420.

Sosinsky, D.J., & Sommerville, I. D. (1986). The composition and temperature dependence of the sulfide capacity of metallurgical slags. *Metallurgical Transactions B*, *17*(2), 331–337.

Sorsa, A., Leiviskä, K., Santa-aho, S., Vippola, M., & Lepistö, T. (2013). An efficient procedure for identifying the prediction model between residual stress and Barkhausen noise. *Journal of Nondestructive Evaluation*, *32*(4), 341–349.

SSAB Europe [Raahe steelworks] (2020). *Showcase Material on Ironmaking and Hot Metal Desulphurization.*

Suesserman, M. F., & Spelman, F. A. (1993). Lumped-parameter model for in vivo cochlear stimulation. *IEEE Transactions on Biomedical Engineering*, *40*(3), 237–245.

Suresh, S., Saraswathi, S., & Sundararajan, N. (2010). Performance enhancement of extreme learning machine for multi-category sparse data classification problems. *Engineering Applications of Artificial Intelligence*, *23*(7), 1149–1157.

Stavseth, M. R., Clausen, T., & Røislien, J. (2019). How handling missing data may impact conclusions: A comparison of six different imputation methods for categorical questionnaire data. *SAGE Open Medicine*, *7*, 2050312118822912.

Szekely, J. (1988). The mathematical modeling revolution in extractive metallurgy. *Metallurgical Transactions B*, *19*(4), 525–540.

Tan, K. C., & Li, Y. (2002). Grey-box model identification via evolutionary computing. *Control Engineering Practice*, *10*(7), 673–684.

Taylor J., & Stobo J. J., 1954. The Sulphur Distribution Reaction between Blast–Furnace Slag Metal. *The Journal of the Iron and Steel Institute*, 180 (4), 360–367.

Te Braake, H. A. B., Van Can, H. J. L., & Verbruggen, H. B. (1998). Semi-mechanistic modeling of chemical processes with neural networks. *Engineering Applications of Artificial Intelligence*, *11*(4), 507–515.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288.

Tong, Z., Qiao, J., & Jiang, X. (2017). Hot Metal Desulfurization Kinetics by $CaO–Al_2O_3–SiO_2–MgO–TiO_2–Na2O$ Slags. *ISIJ International*, *57*(2), 245–253.

Tsao T., & Katayama H. G. (1986). Sulphur Distribution between Liquid Iron and $CaO-MgO-Al_2O_3-SiO_2$ Slags Used for Ladle Refining. *ISIJ International, 26*(8), 717–723.

Van Niekerk, W. H, & Dippenaar, R. J. (1993). Thermodynamic aspects of Na_2O and CaF_2 containing lime-based slags used for the desulphurization of hot-metal. *ISIJ International*, *33*(1), 59–65.

Venkatradi A. S., & Bell H. B. (1969). Sulphur Partition between Slag and Metal in the Iron Blast Furnace. *The Journal of the Iron and Steel Institute, 207*(8), 1110–1113.

Vinoo, D. S., Mazumdar, D., & Gupta, S. S. (2007). Optimisation and prediction model of hot metal desulphurisation reagent consumption. *Ironmaking & Steelmaking*, *34*(6), 471–476.

Visuri, V.-V., Sulasalmi, P., Vuolio, T., Paananen, T., Haas, T., Pfeifer, H., & Fabritius, T. (2019). Mathematical Modelling of the Effect of Reagent Particle Size Distribution on the Efficiency of Hot Metal Desulphurisation, In *Proc. of the 4th European Steel Technology and Application Days*. Düsseldorf, Germany: Stahlinstitut VDEh.

Von Stosch, M., Oliveira, R., Peres, J., & de Azevedo, S. F. (2014). Hybrid semi-parametric modeling in process systems engineering: Past, present and future. *Computers and Chemical Engineering*, *60*, 86–101.

Wang, X., Yang, J., Teng, X., Xia, W., & Jensen, R. (2007). Feature selection based on rough sets and particle swarm optimization. *Pattern Recognition Letters*, *28*(4), 459–471.

Winkler T. B., & Chipman J. (1946). An Equilibrium Study of the Distribution of Phosphorus Between Liquid Iron and Basic Slags. *Transactions of the American Institute of Mining Engineers*, *167*, 111–133.

Wong, K. P., Li, A., & Law, M. Y. (1997). Development of constrained-genetic-algorithm load-flow method. *IEEE Proceedings-Generation, Transmission and Distribution*, *144*(2), 91–9.

Xiong, Q., & Jutan, A. (2002). Grey-box modelling and control of chemical processes. *Chemical Engineering Science*, *57*(6), 1027–1039.

Young, R. W., Duffy, J. A., Hassall, G. J., & Xu, Z. (1992). Use of optical basicity concept for determining phosphorus and sulphur slag-metal partitions. *Ironmaking & Steelmaking*, *19*(3), 201–219.

Zeng, N., Zhang, H., Liu, W., Liang, J., & Alsaadi, F. E. (2017). A switching delayed PSO optimized extreme learning machine for short-term load forecasting. *Neurocomputing*, *240*, 175–182.

Zhang, G. H., Chou, K. C., & Pal, U. (2013). Estimation of sulfide capacities of multicomponent slags using optical basicity. *ISIJ international*, *53*(5), 761–767.

Zhao, Y. (1992). *The role of oxygen in hot metal desulphurization with calcium carbide powder injection*, Doctoral dissertation, McMaster University.

Zhao, Y. F., & Irons, G. A. (1994). Calcium carbide powder injection into hot metal. I: Heat transfer to particles. *Ironmaking & Steelmaking*, *21*(4), 303–308.

Zhao, Y. F., & Irons, G. A. (1994). Calcium carbide powder injection into hot metal. II: Simultaneous desulphurisation and deoxidation. *Ironmaking & Steelmaking*, *21*(4), 309–317.

Original publications

I Vuolio, T., Visuri, V.-V., Tuomikoski, S., Paananen, T., & Fabritius, T. (2018). Data-driven mathematical modeling of the effect of particle size distribution on the transitory reaction kinetics of hot metal desulfurization. *Metallurgical and Materials Transactions B*, *49*(5), 2692–2708. https://doi.org/10.1007/s11663-018-1318-4

II Vuolio, T., Visuri, V.-V., Paananen, T., & Fabritius, T. (2019). Identification of rate, extent and mechanisms of hot metal resulfurization with CaO-SiO_2-Na_2O slag systems. *Metallurgical and Materials Transactions B*, *50*(4), 1791–1807. https://doi.org/10.1007/s11663-019-01600-5

III Vuolio, T., Visuri, V.-V., Sorsa, A., Paananen, T., & Fabritius, T., (2019). Genetic Algorithm based variable selection in prediction of hot metal desulfurization kinetics. *Steel Research International*, *90*(8), 1900090. https://doi.org/10.1002/srin.201900090

IV Vuolio, T., Visuri, V.-V., Sorsa, A., Ollila, S., & Fabritius, T. (2020). Application of a genetic algorithm based model selection algorithm for identification of carbide-based hot metal desulfurization. *Applied Soft Computing Journal*, *92*, 106330. https://doi.org/10.1016/j.asoc.2020.106330

V Visuri, V.-V., Vuolio, T., T. Haas, & Fabritius, T. (2020). A review of modeling hot metal desulfurization. *Steel Research International*, *91*(4), 1900454. https://doi.org/10.1002/srin.201900454

Original publications are not included in the electronical version of the dissertation.

761. Abou Zaki, Nizar (2020) The role of agriculture expansion in water resources depletion in central Iran

762. Gyakwaa, Francis (2020) Application of Raman spectroscopy for the characterisation of synthetic non-metallic inclusions found in Al-killed calcium treated steels

763. Pandya, Abhinay (2020) Demographic inference and affect estimation of microbloggers

764. Eckhardt, Jenni (2020) Mobility as a Service for public-private partnership networks in the rural context

765. Apilo, Olli (2020) Energy efficiency analysis and improvements of MIMO cellular communications

766. Gogoi, Harshita (2020) Development of biosorbents for treatment of industrial effluents and urban runoffs

767. Saavalainen, Paula (2020) Sustainability assessment tool for the design of new chemical processes

768. Ferdinando, Hany (2020) Classification of ultra-short-term ECG samples : studies on events containing violence

769. Leinonen, Marko (2020) Over-the-air measurements, tolerances and multiradio interoperability on 5G mmW radio platform

770. Pakkala, Daniel (2020) On design and architecture of person-centric digital service provisioning : approach, fundamental concepts, principles and prototypes

771. Mustaniemi, Janne (2020) Computer vision methods for mobile imaging and 3D reconstruction

772. Khan, Uzair Akbar (2020) Challenges in using natural peatlands for treatment of mining-influenced water in a cold climate : considerations for arsenic, antimony, nickel, nitrogen, and sulfate removal

773. Kumar, Tanesh (2020) Secure edge services for future smart environments

774. Laitila, Juhani Markus (2021) Effect of forced weld cooling on high-strength low alloy steels to interpass temperature

775. Ramesh Babu, Shashank (2021) The onset of martensite and auto-tempering in low-alloy martensitic steels

776. Kekkonen, Päivi (2021) Several actors, one workplace—Development of collaboration of several actors inside and between the organisations

Book orders:
Virtual book store
http://verkkokauppa.juvenesprint.fi

UNIVERSITY OF OULU P.O. Box 8000 FI-90014 UNIVERSITY OF OULU FINLAND

ACTA UNIVERSITATIS OULUENSIS

ISBN 978-952-62-2835-8 (Paperback)
ISBN 978-952-62-2836-5 (PDF)
ISSN 0355-3213 (Print)
ISSN 1796-2226 (Online)

UNIVERSITY OF OULU