Development of an Artificial Neural Network to Predict Sulphide Capacities of CaO–SiO₂–Al₂O₃–MgO Slag System

Alvin MA,¹⁾ Sina MOSTAGHEL²⁾ and Kinnor CHATTOPADHYAY^{1)*}

1) Department of Materials Science and Engineering, University of Toronto, Toronto, M5S 3E3 Canada.

2) Technology Development Group, HATCH Ltd, 2800 Speakman Drive, Mississauga, ON, L5K 2R7 Canada.

(Received on June 18, 2016; accepted on September 21, 2016; J-STAGE Advance published date: November 30, 2016)

Depletion of the high quality ores around the world has forced ferronickel producers to extract metal values from low-grade ore bodies with significant amounts of impurities. Under this condition, maintaining alloy quality is of utmost importance for the smelters; however still, accessibility of a reliable sulphide capacity model for FeNi refining processes is an issue. Many of the current models, such as those incorporating optical basicity, have proven to be erroneous and unreliable for wide ranges of composition and temperature. These models are typically developed and tested without a proper validation method thus allowing for great correlations which may not fare well with the introduction of new data. Models built from fundamental thermodynamic data perform much better in predicting sulphide capacities but are not only complicated to formulate but also too complicated to be used by operators on a day to day basis as multitude of inputs are needed. Hence, development of a reliable model based on fundamentals, which can also be directly used by plant operators is very much demanded by the industry. In the current study, an artificial neural network (ANN) approach has been used to predict sulphide capacities of slag compositions in the CaO–SiO₂–Al₂O₃–MgO system with an objective to be used in ferronickel refining processes. The resulting models are evaluated on: 1) coefficient of multiple determination (R²), 2) correlation strength (r), 3) root mean square error (RMSE) and 4) computation speed. The ANN based model has shown to be superior in predicting sulphide capacities to current models.

KEY WORDS: ferronickel refining; artificial neural network; FactSage; optical basicity.

1. Introduction

The refining of crude metal is of utmost importance due to stringent quality demands and a deterioration in ore quality. In particular, sulphide capacities which is a measure of the sulphide removing power of the slag, is an important metric which should be optimized in order to obtain a desired final product. Despite the wealth of knowledge in the modeling domain, most of the models have largely become stale. The work presented here will tackle sulphide capacity modeling from a new perspective.

Artificial neural network (ANN) modeling has been used in many different fields. This type of predictive statistics has thrived in the present surge in computing power and as a result, has risen to the forefront of many different industries. Some of its applications include investment forecasting,¹⁾ speech recognition,²⁾ and control systems.³⁾ Its perforation into the metallurgical field has largely been non-existent. Only a few researchers in metallurgy have employed machine learning algorithms much less neural network based ones.^{4–8)}

The research presented here attempts to cross-pollinate

the machine learning technology into chemical process metallurgy. Thus an ANN approach was employed by the authors to explore its potential application in process metallurgy with specific emphasis on desulphurization.

2. Sulphide Capacity

Sulphur in the slag can be described by the following equilibria:

Equation (1) describes the equilibrium between the slag and gas phase in terms of relative partial pressures of both sulphur and oxygen. Comparatively, the sulphur system can also be approached from the metal-slag equilibrium (Eq. (2)). Both systems reveal two components imperative for efficient desulphurization: 1) O²⁻ and thus basicity must be sufficiently high and 2) the [O] content must be low. Many fluxing reagents are typically employed to meet these prerequisites.

To quantify the desulphurization of a melt, Fincham *et al.*⁹⁾ introduced the concept of sulphide capacity. Since its conception, the sulphide capacity has been the focus of

* Corresponding author: E-mail: kinnor.chattopadhyay@utoronto.ca
DOI: <http://dx.doi.org/10.2355/isijinternational.ISIJINT-2016-368>

various modeling attempts to understand desulphurization from a chemical standpoint. Fincham *et al.*'s relationship can be derived from the equilibrium coefficient (Eq. (3)) of the gas-slag reaction as seen in Eq. (1). Fincham *et al.*'s definition of sulphide capacity can then be derived through manipulation (Eq. (4)). The correlation also reveals the need for atmospheric and temperature control for efficient sulphur removal.

$$K_e(T) = \frac{P_{O_2}^{\frac{1}{2}} \times f_{S^{2-}} \times (\%S)_{\text{slag}}}{P_{S_2}^{\frac{1}{2}} \times a_{O^{2-}}} \dots\dots\dots (3)$$

$$C_{S_{\text{slg}}} = \frac{K_e \times a_{O^{2-}}}{f_{S^{2-}}} = (\%S)_{\text{slag}} \sqrt{\frac{P_{O_2}}{P_{S_2}}} \dots\dots\dots (4)$$

3. Sulphide Capacity Modeling

3.1. Empirical Modeling

Sulphide capacity has been modeled through both empirical relationships and through modeling of short range interactions:

Empirically, the most common and extensive models are those developed using optical basicity. First proposed by Duffy and Ingram,¹⁰⁾ optical basicity uses the ratio between the electron donor power of oxides in glass and electron donor power of oxide ions to predict sulphide capacity. This paved the way for significant models by Sommerville *et al.*,¹¹⁾ Young *et al.*,¹²⁾ Taniguchi *et al.*¹³⁾ and most recently Zhang *et al.*¹⁴⁾

Sommerville *et al.*'s¹¹⁾ model uses a combined regression correlation to describe sulphide capacities. The model uses a temperature (1 573 K–1 973 K) correlation at various iso-optical basicities to model sulphide capacity. The unifying expression is shown in Eq. (5). While versatile, Sommerville *et al.*'s relationship was found to be inaccurate at higher sulphide capacities.¹²⁾

$$\text{Log}C_s = \frac{22\,690 - 54\,640\Lambda}{T} + 43.6\Lambda - 25.2 \dots\dots\dots (5)$$

In Young *et al.*'s model¹²⁾ a new regression term, Λ^2 , was introduced and the correlation discretized. Based solely on the assumption that a non-linear relationship dominated the optical basicity relationship, Young *et al.* formulated a new relationship shown in Eq. (6). This model better described sulphide capacity than the previous Sommerville *et al.* model¹¹⁾ but could not effectively handle slags with high FeO due to an underestimation of its molecular optical basicity.^{14,15)}

$$\Lambda < 0.8$$

$$\begin{aligned} \text{Log}C_s = & -13.913 + 42.84\Lambda - 23.82\Lambda^2 \\ & - \left(\frac{11\,710}{T} \right) - 0.02223 \times \text{wt} - \% \text{SiO}_2 - 0.02275 \dots \\ & \times \text{wt} - \% \text{Al}_2\text{O}_3 \end{aligned}$$

$$\Lambda \geq 0.8$$

$$\begin{aligned} \text{Log}C_s = & -0.6261 + 0.4808\Lambda + 0.7197\Lambda^2 \\ & + \left(\frac{1\,697}{T} \right) - \left(\frac{2\,587\Lambda}{T} \right) + 0.0005144 \dots \\ & \times \text{wt} - \% \text{Al}_2\text{O}_3 \end{aligned} \dots\dots\dots (6)$$

Similarly, Taniguchi *et al.*,¹³⁾ defined the relationship presented in Eq. (7) to account for sulphide capacities in the CaO–Al₂O₃–SiO₂–MgO–MnO system. In this system low SiO₂ were investigated. While good adherence is seen with experimental work, validation of the regression was not apparent.

$$\begin{aligned} \text{Log}C_s = & 7.350 + 94.89 \log \Lambda \\ & - \frac{10\,051 + \Lambda(-338(\text{wt} - \% \text{MgO}) + 287(\text{wt} - \% \text{MnO}))}{T} \dots \\ & + 0.2284(\text{wt} - \% \text{SiO}_2) + 0.1379(\text{wt} - \% \text{Al}_2\text{O}_3) \\ & - 0.0587(\text{wt} - \% \text{MgO}) \dots \\ & + 0.0841(\text{wt} - \% \text{MnO}) \dots\dots\dots (7) \end{aligned}$$

In the most recent iteration of the optical basicity based modeling approach, Zhang *et al.*¹⁴⁾ proposed a correction to the molecular optical basicity of FeO, increasing it from 1 to 1.24. Also included in Zhang *et al.*'s model is the inherent assumption that a linear relationship exists between the reciprocal of optical basicity and the pre-exponential factor, A (Eq. (7)). The relationship is shown in Eq. (8). While sufficient for most slag systems, Zhang *et al.*'s model¹⁴⁾ has shown fundamental mismatch at high sulphide capacities for slag compositions in the CaO–SiO₂–Al₂O₃–MgO system.¹⁴⁾

$$\text{Log}C_s = \frac{E}{T} + A \dots\dots\dots (8)$$

$$\text{Log}C_s = -6.08 + \frac{4.49}{\Lambda} + \frac{\left(15\,893 - \frac{15\,864}{\Lambda} \right)}{T} \dots\dots\dots (9)$$

While exhaustive, these models do not capture the system in its entirety (covering the entire temperature and compositional ranges pertinent to FeNi refining). Furthermore, a common downfall to these models has been the method in which they have been formulated. The model development is extensive; however, the model validation often employs a non-split data set and thus may not suffice as a thorough assessment of the correlation.

3.2. Short Range Order

Similarly, sulphide capacity modeled from short range order comes with a host of its own problems. Some of the more famous frameworks in this area have been incorporated into proprietary software such as FactSage (developed by CRCT and GTT Technologies),^{16,17)} and ThermoSlag (developed by the Royal Institute of Technology (KTH)).¹⁸⁾

Comprising many different databases as well as frameworks, FactSage is an all-purpose thermodynamics software. It's sulphide capacity framework, developed by Kang and Pelton,¹⁶⁾ is based on the extrapolation of all permutations of cations and anions on a pseudo-lattice. The resulting model has shown great correlation to experimental results in a wide range of slag systems.¹⁶⁾ The advantages of FactSage are clear, however the formulation of a model such as this requires many building blocks such as an already established method for handling and accounting for randomness.¹⁹⁾ Much like the formulation, the use of FactSage is equally complicated. Furthermore, inherent to its formula-

tion, the inputs into FactSage are set and are invariable.

In the KTH model, a specially defined interaction coefficient (ξ) was proposed to model high FeO slags. This new metric is defined in Eq. (10) and an integral part in its prediction mechanism. A major advantage of this model is its ability to predict slags of lower order therefore, an application to ferronickel refining slags plausible. However, similar to FactSage, the formulation of the interaction parameters requires advanced experimental work. Given the associated costs and difficulties with experiments, some interaction parameters may not be measurable or estimated accurately. An excerpt of the formulation applicable for ferronickel refining is shown in Eq. (11).

$$\frac{a_{O^2-}}{f_{S^2-}} = \exp\left(-\frac{\xi}{RT}\right) \dots\dots\dots (10)$$

$$\begin{aligned} RT(\log_e C_s) = & 58.8157 \times T - 118535 \dots \\ & - (X_{Al_2O_3} \times 157705.28 - X_{CaO} \times 33099.43 \\ & + X_{SiO_2} \times 168872.59 + X_{MgO} \times 9573.07) \dots \\ & - \left(\xi_{Al_2O_3-CaO} + \xi_{Al_2O_3-SiO_2} + \xi_{CaO-SiO_2} + \xi_{MgO-SiO_2} + \xi_{Al_2O_3-CaO-MgO} \dots \right. \\ & \left. + \xi_{Al_2O_3-CaO-SiO_2} + \xi_{Al_2O_3-MgO-SiO_2} + \xi_{CaO-MgO-SiO_2} \right) \dots\dots\dots (11) \end{aligned}$$

Although these fundamental models are much more accurate than their empirical counterparts, these models frequently require advance fundamental knowledge which are at times derived from experimental work. These myriad of inputs constitute a model unattractive for general use in pyrometallurgical plants.

4. Artificial Neural Network Development

4.1. Visualization

Data on sulphide capacities of the CaO–SiO₂–Al₂O₃–MgO system have been gathered from 15 different studies.^{9,13,20–32} These studies examine different aspects of the slag system and thus represent a good staging point for the modeling of the whole spectrum of sulphide capacities. The variables of interest were those that are chemically pertinent to the determination of sulphide capacity ($LogC_s$). These are: 1) wt-% CaO, 2) wt-% SiO₂, 3) wt-% MgO, 4) wt-% Al₂O₃, 5) wt-% (S)_{slag}, 6) $\sqrt{(P_{O_2}/P_{S_2})}$, and 7) temperature. Further, as recommended by Fincham *et al.*,⁹ only the studies using an equilibrium time higher than 4.5 hours were considered. The entire data set used in this work is illustrated in a scatter plot shown in Fig. 1.

Figure 1, has been laid out in such a way to promote insight gathering. The relative correlation strengths are shown in Fig. 1 in different shades of colour. Red (high),

Fig. 1. Scatterplot of CaO–SiO₂–Al₂O₃–MgO system, compiled from^{9,13,20–32} and ordered by relative correlation strength. Red, blue and beige represent high, medium and low correlation. (Online version in color.)

blue (medium), and beige (low), are used as the basis of ordering in the figure. The correlation strength quantifies the strength of a linear relationship between two variables. When the correlation is weak there is no tendency for the variables to linearly move with respect to each other. Correlation can be calculated between the variances of two variables and its covariance. These are shown in Eqs. (12) and (13) respectively. The correlation can then be defined in Eq. (14).

$$\text{Var}(i) = \frac{\sum (i - \bar{i})^2}{n} \dots\dots\dots (12)$$

$i = x, y$

$$\text{Covar}(x, y) = \frac{\sum (x - \bar{x}) \times (y - \bar{y})}{n} \dots\dots\dots (13)$$

$$r = \frac{\text{Covar}(x, y)}{\sqrt{\text{Var}(x) \times \text{Var}(y)}} \dots\dots\dots (14)$$

The ordering of Fig. 1 is expected based on the chemical relationship (Eqs. (1) and (2)) derived from the initial sulphide capacity work by Fincham *et al.*⁹⁾ shown in Eqs. (3) and (4). Strong correlation is seen with wt-% CaO, wt-% SiO₂, wt-% Al₂O₃ and wt-% (S)_{slag}. One small caveat of the analysis shown in Fig. 1 is that the ordering is relative. Thus factors such as the root of the pressure ratio and temperature may not appear to be highly correlated but in reality are important factors to consider in the scope of absolute correlation.

4.2. ANN Fundamentals

Instead of a singular equation, an ANN uses numerous equations which account for the relationships between nodes, synapse weights, output and inputs (**Fig. 2**). This complex interplay of relationships can be described through matrices. Defining $X_{n,i}$ as a matrix representing the inputs

Fig. 2. Basic ANN architecture with a single hidden layer. Where (—) are synapses. (Online version in color.)

(columns as variables, rows as unique data points) and $W_{i,k}$ as a matrix which represents the synapse weights, the relationship at the hidden layer for n data points in the set (training or testing), i variables (**Table 1**) and k number of hidden nodes can then be described as:

$$f^{(1)} \left(\begin{bmatrix} X_{1,1} & \dots & X_{1,i} \\ \vdots & \ddots & \vdots \\ X_{n,1} & \dots & X_{n,i} \end{bmatrix} \times \begin{bmatrix} W_{1,1}^{(1)} & \dots & W_{1,k}^{(1)} \\ \vdots & \ddots & \vdots \\ W_{i,1}^{(1)} & \dots & W_{i,k}^{(1)} \end{bmatrix} \right) = A \dots\dots (15)$$

Where f is an activation function and, A is the resulting matrix which will be passed on to the synapses post-hidden nodes. Similarly, the outputs of the hidden nodes are weighted and then transformed a final time before a prediction can be formed:

$$f^{(2)} \left(\begin{bmatrix} A_{1,1} & \dots & A_{1,k} \\ \vdots & \ddots & \vdots \\ A_{n,1} & \dots & A_{n,k} \end{bmatrix} \times \begin{bmatrix} W_1^{(2)} \\ \vdots \\ W_k^{(2)} \end{bmatrix} \right) = \hat{Y} \dots\dots\dots (16)$$

Optimization occurs after each prediction, thus the weights are constantly changing. The method in which weights are corrected will vary depending on the learning algorithm used. To exploit the convex nature of quadratic equations, sum of squared errors is commonly used as a cost function to optimize the weights. This is defined as Eq. (17):

$$J = \frac{1}{2} \sum \left(Y - f^{(2)} \left(f^{(1)} \left(X_{n,i} W_{i,k}^{(1)} \right) W_k^{(2)} \right) \right)^2 \dots\dots\dots (17)$$

Brute forcing this equation is near impossible for high dimensional problems thus a gradient descent algorithm is utilized to solve the optimized path. Akin to dropping a ball and letting it roll towards the path of least resistance, the gradient descent algorithm solves the cost function by finding the path of lowest slope. The direction of the ball is determined iteratively with each step coming from each data point from the training set. This method shortens the time needed to identify a minimum. Once a minimum is found, the weights can then be used as a basis for a predictive model. At this point the testing portion of the dataset can be used. The entire process is depicted in **Fig. 3**.

Table 1. Different ANN models that were considered for the current investigation.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Hidden Nodes	5	5	5	3	7	3	7
Wt-% MgO			X			X	X
Wt-% CaO	X	X	X	X	X	X	X
Wt-% SiO ₂	X	X	X	X	X	X	X
Wt-% Al ₂ O ₃	X	X	X	X	X	X	X
Wt-% (%S) _{slag}	X	X	X	X	X	X	X
$\sqrt{\frac{P_{O_2}}{P_{S_2}}}$		X	X	X	X	X	X
T		X	X	X	X	X	X

Fig. 3. Flowchart for an artificial neural network model. (Online version in color.)

Fig. 4. Errors associated with gradient descent use. Adapted from reference 34.³⁴⁾ (Online version in color.)

Although faster, the gradient descent algorithm has major drawbacks. Unless the cost function has been solved in all domain and range (at sufficient resolution), errors may be masked by the results. These possible errors are 1) detection of a faulty minima 2) quasi-stand still and 3) the departure from a suitable solution (Fig. 4). Any of these errors will lead to improper training of the ANN and thus a faulty model. To avoid errors such as these, the step size or learning rate must be carefully defined.

4.3. Model Formulation

Currently there are many software products available both commercially and open source, which can provide basic prediction architecture. Because of its accessibility, the open source R programming language,³³⁾ known hereafter as R, is employed as the basis of the modeling approach.

The modeling approach in the current work utilizes a sigmoidal-regression artificial neural network model using a globally resilient backpropagation algorithm developed by Anastasiadis *et al.*³⁴⁾ Other learning algorithms such as the traditional back propagation algorithm, which uses a set learning rate have proven unreliable due to many factors.^{35,36)} Its most frequent issue is that it is extremely time-consuming and often results in a partial training of the dataset due to a local minimum convergence of the error function. Aiding this issue requires advanced knowledge of the optimal hidden nodes, starting weights and optimal learning rate. However; these metrics are rarely available thus using a backpropagation algorithm is unattractive. The method developed by Anastasiadis³⁴⁾ uses an adaptive learning rate which aids in better convergence speed and stability. The theorem and corresponding proof of this method can be found in reference 34.³⁴⁾

Only one hidden layer (between output and input layers)

will be considered in the current work. The number of nodes forming the single layer will vary according to Table 1. To introduce non-linearity into the system, a sigmoidal activation function, the logistic function, will be used (Eq. (18)). W will change depending on the location of calculation ($W_{i,k}$ for synapse weights prior to the hidden node and W_k for synapse weights after the hidden node).

$$f(X_{n,i}W) = \frac{1}{1 + e^{-(X_{n,i} \times W)}} \dots \dots \dots (18)$$

With reference to Fig. 3, the output ($\hat{Y}$) is the sulphide capacity of each data point while the inputs (I) will vary in accordance with Table 1. Similar to any regression method, the inputs will consist of easily accessible metallurgical inputs such as dissolved sulphur in slag or chemical composition.

4.4. Model Validation Approach

The simplest form of cross validation, the hold out method, is used to validate the ANN models. In this method, a certain percentage of the dataset is sampled to train the algorithm (training dataset) while the remaining data (testing dataset) is used to validate the model. To avoid over-training and because the dataset is sufficiently large, the split ratio between testing and validating data sets will be 50:50. As samples are taken randomly, careful measures were enacted to ensure sampling quality.

As prescribed by past work in the evaluation of different ANN models,³⁷⁾ the principal framework that will be used to compare different models studied here, are:

- 1) Coefficient of multiple determination (R^2): is a measure of goodness of fit between predicted and actual values. This can be found through the squaring of the correlation coefficient
- 2) Correlation coefficient (r): is derived in Eqs. (1) to (3) and is a measure of linear strength between predicted and actual values
- 3) Root mean square of errors (RMSE): is the absolute average distance between predicted and actual value and is defined as:

$$RMSE = \sqrt{\frac{\sum (Y - \hat{Y})^2}{n}} \dots \dots \dots (19)$$

Often overlooked or outright ignored, the computation time is an important factor to consider, when qualifying an algorithm for industrial scale use. Although computation time for a dataset of the current size is irrelevant within a

small dataset, in industry the requirements are much different. The computational speed will increase significantly with increasing sensors, compositions, temperatures and sampling intervals. In the work presented here, time will be defined as the time it takes for the optimization of matrix W and will be the final criterion used to evaluate different ANN models presented here.

5. Results and Discussion

5.1. ANN Models

The ANN models developed in accordance with Table 1, was benchmarked under the 4 previously criteria. They have been trained using the training dataset and validated using the testing dataset. The adequacy of each model has been presented in Table 2. Based on only the highly correlated variables shown in Fig. 1, model 1 performed the worst. Without including $\sqrt{(P_{O_2}/P_{S_2})}$, and temperature, an artificial neural network trained using the remaining variables was unable to accurately predict changes in the sulphide capacity. This was also evident in Fincham *et al.*'s original sulphide capacity relationship.⁹⁾

Models 3 and 4 tested the need of MgO. The correlation between MgO and sulphide capacity is a highly contested matter. From literature, the deployment of MgO was found to be highly model dependant. In Taniguchi *et al.*'s model,¹³⁾ MgO was positively correlated with sulphide capacity. Similarly, work by Kang *et al.*¹⁶⁾ using FactSage predicted a positive correlation but only at high mole fractions (>0.6). Conversely, KTH's Thermoslag,¹⁸⁾ showed a negative correlation.¹³⁾ Both models by Sommerville *et al.*¹¹⁾ and Young *et al.*¹²⁾ showed negligible effect of MgO. In the work presented here, MgO was shown to be trivial in the prediction. However, it is important to note that the maximum MgO considered in the dataset was 25 wt-% thus in this work, the assumption that MgO is negligible is only valid for amounts lower than 25 wt-%.

The consideration of MgO also resulted in faster computational time. This suggests that mathematically, MgO aids in finding the steepest gradient in the gradient descent algorithm. This reduction in model formulation time has also been corroborated in Models 6 and 7. Models 4 to 7 investigated the optimal number of nodes in the hidden layer. Without MgO (Models 4 and 5), the increase in dimensionality not only allowed for better predictions but also faster computational times. The increase in synapses potentially allowed for better description of the dataset resulting in faster convergence of the cost function. Similarly, with MgO (Models 6 and 7), faster and more accurate predictions were the result of more hidden nodes in the hidden layer. Further increasing the number of nodes beyond

7, may also result in better predictions; however, was not investigated further due to the possibility of overfitting the data. Furthermore, MgO seems to have a relatively more pronounced effect for 3 and 7 node networks. The RMSE has shown a slight decrease with MgO present as a predictor. This suggests one of two things: 1) convergence of the cost function to a local minima or 2) global translation of the cost function upwards. The resemblance of the models with and without MgO suggests that any convergence or vertical translation may be only minimal. Because of this, a network which does not deploy MgO is preferred.

The predicted LogCs values have been compared with the actual LogCs (testing dataset) values from Figs. 5–11. Also included is the ideal predictor line. This line, described by a slope of 1 and an intercept of (0,0), shows a model's fidelity to an ideal model (Actual LogCs = Predicted LogCs). By inspection, Models 2, 3, 5, and 7 are much closer to ideal than the rest. Therefore, on the basis of the current dataset and validation approach, any of these architectures are applicable.

5.2. Benchmarking

For further verification of the neural network models, the predicted sulphide capacities using established optical

Fig. 5. ANN Model 1 performance. (Online version in color.)

Fig. 6. ANN Model 2 performance. (Online version in color.)

Table 2. Validation Results of various models.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
R^2	0.883	0.936	0.934	0.924	0.943	0.920	0.939
R	0.939	0.967	0.966	0.961	0.971	0.959	0.969
RMSE	0.276	0.200	0.201	0.221	0.189	0.225	0.196
time (s)	0.186	0.272	0.223	0.355	0.232	0.308	0.194

Fig. 7. ANN Model 3 performance. (Online version in color.)

Fig. 10. ANN Model 6 performance. (Online version in color.)

Fig. 8. ANN Model 4 performance. (Online version in color.)

Fig. 11. ANN Model 7 performance. (Online version in color.)

Fig. 9. ANN Model 5 performance. (Online version in color.)

Table 3. Optical basicity values of CaO, SiO₂, Al₂O₃, and MgO.³⁸⁾

CaO	SiO ₂	Al ₂ O ₃	MgO
1	0.48	0.605	0.78

these models are shown in **Table 4** and **Figs. 12–16**.

The comparison between past empirical models and the neural network based models have yielded favourable results. The ANN models demonstrated higher accuracy, and better cohesion to the ideal predictor. Sommerville *et al.*'s¹¹⁾ model showed a general trend towards the ideal model but because it was not discretized, exhibited large variance at higher sulphide capacities (Fig. 12). Also, because Young *et al.*¹²⁾ and Zhang *et al.*¹⁴⁾ used similar methodologies, their respective models showed fundamental deviation in the same region (Figs. 13 and 14). Although, Taniguchi *et al.*'s model performed the best out of this cohort its R^2 , r and RMSE values are still slightly lower than those of the neural network (Fig. 15).

Benchmarking against FactSage has also displayed the strength of a neural network-based prediction system against a short range order model. There are many fundamental differences between both models. As mentioned previously, the formulation of FactSage required advance knowledge

basicity models^{11–14)} have been calculated using the full data set (training and testing) with the values from **Table 3**. It is assumed that the dataset employed here has not been used in the formulation of the previous models, and thus is credible for model re-validation. Predictions from FactSage's Equilibrium module, using the FactOxid database and Eq. (4) as the custom function input, are also included. The results of

Table 4. Results of different sulphide capacity models.

	Sommerville ¹¹⁾	Young ¹²⁾	Taniguchi ¹³⁾	Zhang ¹⁴⁾	FactSage ¹⁶⁾
R^2	0.895	0.851	0.878	0.884	0.813
R	0.946	0.922	0.937	0.940	0.901
RMSE	0.379	0.873	0.210	0.299	0.282

Fig. 12. Sommerville¹¹⁾ model performance. (Online version in color.)

Fig. 13. Young¹²⁾ model performance. (Online version in color.)

Fig. 14. Taniguchi¹³⁾ model performance. (Online version in color.)

Fig. 15. Zhang¹⁴⁾ model performance. (Online version in color.)

Fig. 16. FactSage¹⁶⁾ model performance. (Online version in color.)

of certain thermodynamic systems. Its basis, the modified quasi-chemical model,¹⁹⁾ allows for accurate thermodynamic predictions but only if certain variables are known. One of such variables is the absolute pressures in the sulphide system (Eq. (4)). Many studies opted for only relative pressures ($\sqrt{P_{O_2}/P_{S_2}}$), thus a portion of the studies were unpredictable using FactSage (Fig. 16). Being much more adaptable, the ANN models could utilize the ratios or the absolute values (if re-formulated to do so).

6. Conclusion

Empirical models such as the optical basicity models tend to breakdown when new system parameters are introduced and as a result, unfeasible in a dynamic environment. The idiosyncrasies associated with the use of short range order models such as FactSage, severely limits its application to only situations in which all its input requirements are met. This divide has created a need for a new type of model which, has the accuracy level of FactSage, but simple to formulate and does not deviate with the introduction of new data.

The results gathered in this work have shown the viability of implementing a neural network based prediction system

for CaO–SiO₂–Al₂O₃–MgO sulphide capacities. The development of the model showed that an MgO-based architecture (Model 3, 6 and 7) had acceptable accuracy and also a lower computation time. This however, was offset by the possibility of the cost function converging to a local minimum or shifting the cost curve entirely. An MgO-less design showed higher accuracy but slightly higher computational time and thus is preferred.

Using the same comparative metrics (*i.e.* R², r, and RMSE), the neural network models were much more robust in their prediction compared to other existing models.

Acknowledgements

The comments and helpful suggestions from the members of the PM²G group are gratefully acknowledged.

Nomenclature

a _{O²⁻}	Activity of ionic oxygen
Λ	Optical basicity
A	Pre-exponential factor (used in Zhang model)
ANN	Artificial neural network
Covar (x,y)	Covariance function
C _S	Sulphide capacity
ξ	Interaction coefficient (used in KTH model)
E	Enthalpy change for the reaction of ionic sulphur between gas and slag
f ⁿ ()	Activation function where <i>n</i> is the layers of synapses between nodes
f _{S²⁻}	Activity coefficient of ionic sulphur
ī	Population average of an arbitrary variable
J	Cost function for the gradient descent algorithm
K _e	Equilibrium constant for the reaction of ionic sulphur between gas and slag
Log	Base 10 logarithmic function
P _{O₂}	Partial pressure of oxygen gas
P _{S₂}	Partial pressure of sulfur gas
r	Correlation between actual and predicted values
R ²	Coefficient of multiple determination
RMSE	Root mean square error
Wt-%	Weight percent
T	Temperature
V _i	Variable count (<i>i</i> = 1,2,3...)
Var (i)	Variance function
W _{i,k}	Matrix of synapse weights prior to hidden node where <i>i</i> is the number of input variables and <i>k</i> is the amount of hidden nodes
W _k	Matrix of synapse weights post-hidden nodes where <i>k</i> is the number of hidden nodes
x	Arbitrary variable
X _{n,i}	Dataset matrix where <i>n</i> represents the amount of data points (<i>n</i> = 1,2,3...) and <i>i</i> the amount

of variables

X _i	Mole fraction (<i>i</i> = Al ₂ O ₃ , CaO, MgO, SiO ₂)
y	Arbitrary variable
Y	Actual sulphide capacity
Ŷ	Predicted sulphide capacity value

REFERENCES

- 1) T.-L. Chen and F.-Y. Chen: *Inf. Sci.*, **346–347** (2016), 261.
- 2) J. Sangeetha and S. Jothilakshmi: *Int. J. Comput. Vis. Robot.*, **6** (2016), 113.
- 3) T. Kerdphol, Y. Qudaih, M. Watanabe and Y. Mitani: *Energy Sustain. Soc.*, **6** (2016), 1.
- 4) A. González-Marcos, J. Ordieres-Meré, F. Alba-Elías, F. J. Martínez-De-Pisón and M. Castejón-Limas: *Ironmaking Steelmaking*, **41** (2014), 262.
- 5) J. Mori and V. Mahalec: *Expert Syst. Appl.*, **49** (2016), 1.
- 6) J. Mori and V. Mahalec: *Comput. Chem. Eng.*, **79** (2015), 113.
- 7) E. Palaneeswaran, G. Brooks and X. B. Xu: *Metall. Mater. Trans. B*, **43** (2012), 571.
- 8) L. Z. Yang, R. Zhu, K. Dong, W. J. Liu and G. H. Ma: *Proc. 3rd Int. Conf. Chemical and Material Engineering, ICCMME, Zhuhai, China*, (2013), 1540.
- 9) C. J. B. Fincham and F. D. Richardson: *Proc. R. Soc. A*, **223** (1954), 40.
- 10) J. A. Duffy and M. D. Ingram: *J. Non-Cryst. Solids*, **21** (1976), 373.
- 11) I. D. Sommerville and D. J. Sosinsky: *Metall. Trans. B*, **17B** (1986), 331.
- 12) R. W. Young, J. A. Duffy and Z. Xu: *Ironmaking Steelmaking*, **19** (1992), 201.
- 13) Y. Taniguchi, N. Sano and S. Seetharaman: *ISIJ Int.*, **49** (2009), 156.
- 14) G.-H. Zhang, K.-C. Chou and U. Pal: *ISIJ Int.*, **53** (2013), 761.
- 15) C. Shi, X. Yang, J. Jiao, C. Li and H. Guo: *ISIJ Int.*, **50** (2010), 1362.
- 16) Y.-B. Kang and A. D. Pelton: *Metall. Mater. Trans. B*, **40B** (2009), 979.
- 17) Center for Research in Computational Thermochemistry: FactSage 7.0, www.factsage.com, (accessed 2016-05-28).
- 18) S. Mostaghel, T. Matsushita, C. Samuelsson, B. Björkman and S. Seetharaman: *Trans. Inst. Min. Metall. Sect. C Miner. Process. Extr. Metall.*, **122** (2013), 42.
- 19) A. D. Pelton and P. Chartrand: *Metall. Mater. Trans. Phys. Metall. Mater. Sci.*, **32** (2001), 1355.
- 20) E. Drakaliysky, D. Sichen and S. Seetharaman: *Can. Metall. Q.*, **36** (1997), 115.
- 21) M. A. Andersson, P. G. Jönsson and M. Hallberg: *Ironmaking Steelmaking*, **27** (2000), 286.
- 22) M. Hino, S. Kitagawa and S. Ban-Ya: *ISIJ Int.*, **33** (1993), 36.
- 23) M. Ohta, T. Kubo and K. Morita: *Tetsu-to-Hagané*, **89** (2003), 742.
- 24) C. Allertz and D. Sichen: *Metall. Mater. Trans. B*, **46B** (2015), 2609.
- 25) M. Gonerup and O. Wijk: *Scand. J. Metall.*, **25** (1996), 103.
- 26) E. Drakaliysky, N. Srinivasan and L. Staffansson: *Scand. J. Metall.*, **20** (1991), 251.
- 27) K. Karsud: *Scand. J. Metall.*, **13** (1984), 144.
- 28) M. Kalyanram: *J. Iron Steel Inst.*, **195** (1960), 58.
- 29) K. Abraham and F. D. Richardson: *J. Iron Steel Inst.*, **196** (1960), 313.
- 30) R. A. Sharma and F. D. Richardson: *J. Iron Steel Inst.*, **198** (1961), 386.
- 31) J. Cameron, T. B. Gibbons and J. C. Taylor: *J. Iron Steel Inst.*, **204** (1966), 1223.
- 32) P. T. Carter and T. G. Macfarlane: *J. Iron Steel Inst.*, **185** (1957), 54.
- 33) R. Gentleman and R. Ihaka: *The R Project for Statistical Computing*, University of Auckland, (1997).
- 34) A. D. Anastasiadis, G. D. Magoulas and M. N. Vrahatis: *Neurocomputing*, **64** (2005), 253.
- 35) R. Liu, G. Dong and X. Ling: *Proc. 34th IEEE Conf. on Decision and Control, IEEE, Piscataway, NJ*, (1995), 1278.
- 36) M. Gori and A. Tesi: *IEEE Trans. Pattern Anal. Mach. Intell.*, **14** (1992), 76.
- 37) E. Palaneeswaran, P. E. D. Love, M. M. Kumaraswamy and T. S. T. Ng: *Build. Res. Inf.*, **36** (2008), 450.
- 38) J. A. Duffy and M. D. Ingram: *J. Non-Cryst. Solids*, **144** (1992), 76.