Analysis of splashing in Basic Oxygen Furnace through systematic modelling

Jari Ruuska*. Aki Sorsa* Seppo Ollila Kauko Leiviskä***

** University of Oulu, Control Engineering Laboratory*
P.O.Box 4300, FIN-90014 University of Oulu, Finland
*** Raahe Steel Works, Ruukki Metals Oy, Raahe, Finland*

Abstract: This paper presents some preliminary results of using a systematic approach for identifying the interactions between splashing and the variables measured from the basic oxygen furnace. Splashing is an undesired phenomenon and thus its analysis is important. Earlier the analysis is carried out mainly manually while the systematic approach used in this paper uses forward-selection for selecting the significant variables and multivariable linear regression as a modelling technique for identifying a model between splashing and process variables. The results show that the procedures used are able to find a variable subset that can be used for explaining some changes in splashing. Despite the promising results the process and the studied approach needs more research in the future. Especially, the procedure used needs to be complemented with data selection.

Keywords: steel making, basic oxygen furnace, splashing, automatic variable selection

1. INTRODUCTION

Basic oxygen furnace (BOF) is a sub-process in steelmaking where hot metal is converted into molten steel by reducing the carbon content. To burn carbon, pure oxygen is blown from above while the batch is stirred using an inert nitrogen gas blown from the bottom of the vessel. Controlling of the process is challenging because different additional materials are fed to the vessel during processing. These additions lead to unstable burning, for example, the addition of silicon causes disturbances. The consequence of unstable burning is material losses due to splashing. Splashing can be monitored by using image-based measurement.

The analysis of the causes of splashing is significant to identify and to avoid processing conditions leading to splashing. This analysis is earlier carried out mainly manually by clustering (Ruuska et al. 2006, Ruuska 2012). Some clustering criteria applied are converter type, heat size, target carbon content and amount of silicon added in the early stages of the process. The results obtained are promising but even better results may be obtained if automated procedures are used. Such an automatic procedure may be, for example, an algorithm which detects the significant process variables by identifying a model between splashing and process variables. This paper presents some preliminary results obtained from a study where models are identified between process variables and splashing.

The accuracy of the model obviously depends on the model structure and thus on the input variables. Thus the careful selection of input variables is very significant. It has been reported that excess variables lead to, for example, deterioration of model performance (Alexandridis et al. 2005), increased time consumption in model training (Guyon and Elisseeff 2003) and more difficult interpretation of the developed model (Smit et al. 2008).

In the literature, many methods have been proposed for selecting the appropriate variable subset. These methods can be roughly divided into filters and wrappers (Kohavi and John 1997). Filters are computationally efficient methods where variables are added to the model according to some ranking. However, the modelling technique applied also has an influence on variable selection. Wrapper methods take this into account and include the model identification procedure into the selection procedure. Typical wrapper methods are, for example, forward-selection, backward-elimination and genetic algorithms.

In this study, an automatic procedure is used for analysis of splashing. The results given are, however, preliminary. The algorithm uses a simple deterministic variable selection and multivariable linear regression models to map the interactions between splashing and process measurements. Instead of finding variables that individually explain splashing the aim is to find a subset of process variables able to predict splashing accurately.

2. MATERIALS AND METHODS

2.1 Basics of BOF

The idea of basic oxygen furnace in its simplest form (Fig. 1) is performed in a vessel with a basic refractory lining and an off-gas cleaning system. The vessel is tilted in order to charge the predetermined and weighed amounts of liquid hot metal and solid recycled steel. Gaseous oxygen is blown onto the metal bath until the estimated chemical composition and temperature are achieved. Fluxes such as burned lime and dolomite are added from the top of the vessel to regulate the process while for example iron ore is added as a coolant and ferrosilicon is added to produce additional heat. In normal cases, steel samples are taken and steel bath temperatures are measured from the tilted converter only after the end of the heat. When the heat has ended, the vessel is tilted to the other side and steel is tapped through a tap hole into a steel ladle. Slag is then tapped into a slag pot and the converter is ready for the next batch.

Specific aims for BOF include precisely described end point values for steel weight, temperature and composition. Carbon, phosphorus and sulphur and often also nitrogen, manganese and hydrogen concentrations need to be within the target windows. Typical end point temperatures rose by 50 °C to 1650 - 1700 °C or even higher depending on the steel grade and the subsequent process stages. Stricter quality demands resulted that sulphur limits had to be lowered to prevent for example crack formation. Modern steel quality also required better phosphorus removal and the oxygen content in the liquid steel became far more critical. Moreover, it became more critical to get the heat ready at the right time for the following process phases. (Boom 2003)

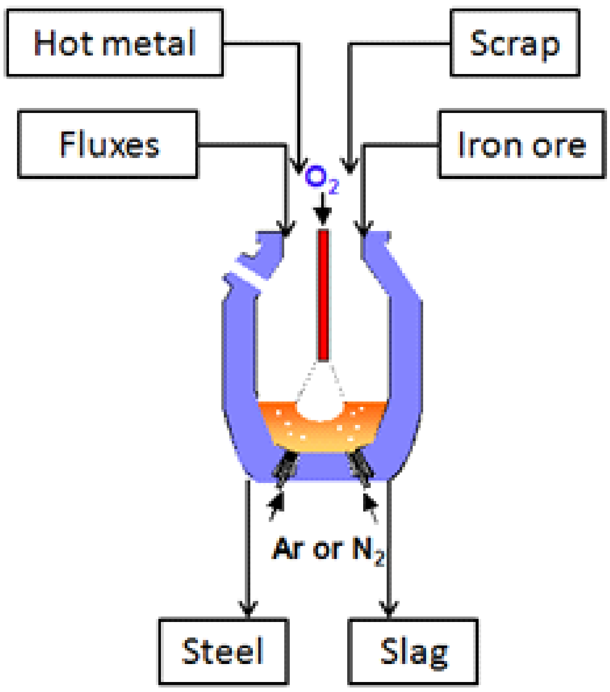

Fig. 1. Basic Oxygen Furnace.

2.2 Splashing

Splashing is a serious problem in BOFs. It causes economic losses and has therefore been widely researched. The negative effects of splashing are well-known, for example lower yield, different kinds of skull formation; in the lance, upper BOF cone, BOF mouth and gas hood; hard blowing, poor dephosphorisation and desulphurisation (Bock et al. 2000). To prevent splashing, the factors that cause foaming need to be known. A significant amount of research has been carried out on this field. Jung and Fruehan (2000) investigated the effects of FeO content, basicity, TiO_2, MgO and temperature of slag on foaming. Koch et al. (1993) determined the critical amount of blown gas when splashing starts, depending on the jet impulse, depth of crater and surface tension. Tang et al. (2008) studied the effects of lance height and bottom stirring flow rate on the mixing time, the amount of splashing, the penetration depth and level fluctuation using a water model. These studies found optimum levels for the parameters in the different phases of the heat. Luomala et al. (2002) investigated the effects of the following variables: lance height, gas flow rate, lance nozzle angle, bottom blowing, lance position and foamy slag. The reduction of the lance nozzle angle increased the total amount of splashing. The usage of bottom blowing increases splashing on the lower parts of the converter.

2.3 Measurements and Data Sets Used

The measurement data of different heats are obtained from the Ruukki's Raahe Steel Works' database. An example of a heat is given in Fig. 2. The analysis of these heat trajectories in time-domain may reveal the interactions between process conditions and splashing. However, such an analysis is complex and significant interactions may not be found. The complexity arises from the great number of variables and their individual and combined effects on splashing. The dynamics of the effects of different variables also vary which makes the analysis even more complex. The time-dependency of the process is neglected in this study and the variables are averaged. The models are systematically identified for the averaged values from the whole data set and from the data sets corresponding to different phases of the heat shown in Fig. 2. In this study, the first 20% of the data represents the ignition phase and the last 10% the end of heat phase. The remaining 70% represents the actual heat phase. Splashing occurs mainly in the set of actual blowing phase as shown in Table 1 which shows the average splashing in each data set. In later studies this splitting may need to be revised.

Splashing measurement has been developed at Ruukki's Raahe Steel Works. It is based on image analysis. At Ruukki's Raahe steel works, there are video cameras for monitoring purposes under the BOFs. These existing cameras were utilised when the splashing measurement was developed. The picture pixels were analysed from a snapshot captured with the camera. The amount of splashing was investigated inside a predefined area by counting the ratio of bright and dark pixels; the limiting value of brightness was predefined. The ratio of bright and dark pixels gives a numerical value for the splashing. The splashing integral for the whole heat was calculated from the momentary values. The variation of splashing within the heats can be seen from Fig. 3. (Ruuska 2012)

The data sets include the measurements of 8 continuous and 9 batch variables from 397 heats. The continuous measurements are flow rates of oxygen, nitrogen and argon, off-gas temperature, FeSi and burnt lime added and lance height. The batch information includes, for example, heat size and content of main chemical components.

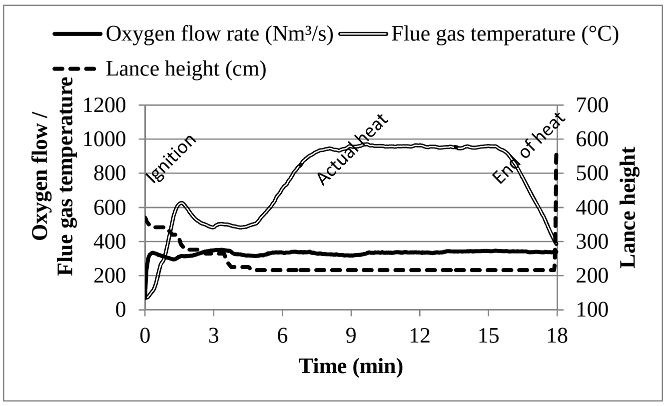

Fig. 2. A typical BOF heat and different phases.

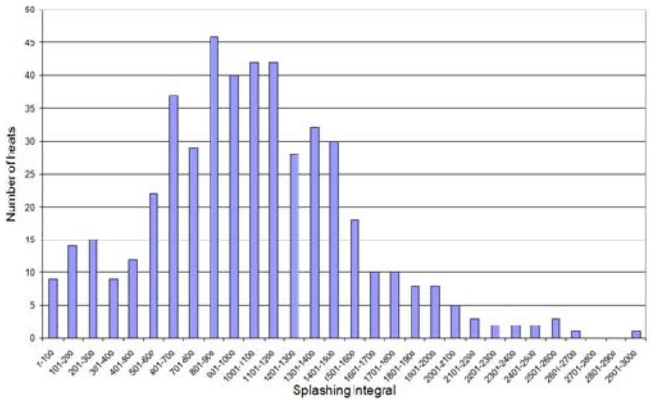

Fig 3. Variation of splashing in different heats.

Table 1. The average splashing in the data sets

Data set	Splashing
Ignition	0.10
Actual heat	2.91
End of heat	0.04
Whole data	2.06

2.4 Multivariable Linear Regression

A multivariable linear regression (MLR) model is given by (Harrel 2001)

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\mathbf{b}}, \tag{1}$$

where $\mathbf{y}$ is the output variable ($N\times1$), $\mathbf{X}$ is the input variable matrix ($N\times M$) and $\mathbf{b}$ is the vector of regression coefficients ($M\times1$). N is the number of data points and M is the number of variables. It should be noticed that the equation above is written for a single output variable. The regression coefficients are obtained as the least squares solution given by

$$\hat{\mathbf{b}} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)\mathbf{X}^{\mathrm{T}}\mathbf{y}\,. \tag{2}$$

The MLR model structure is capable of capturing the major interactions even though it is limited to linear relationships. The obtained model can be easily interpreted and thus significant information about the underlying interactions between process conditions and splashing can be identified.

2.5 Stepwise Variable Selection

The data sets used include many variables. The identification of appropriate models needs the selection of significant variables from the set of candidate variables and the identification of the regression coefficients of the MLR model. The variable selection methods reported in the literature can be roughly divided into filters and wrappers (Kohavi and John 1997). Filters are based on the ranking according to which variables are added to the model. Typical ranking criterion is correlation. The filter approaches seldom lead to optimal result even though their computational efficiency makes their use intriguing (Guyon and Elisseeff 2003). This study, however, uses wrapper approaches and thus filter approaches are not further discussed.

In the wrapper methods, variable selection and model parameter identification are carried out simultaneously which makes them computationally more expensive. The wrapper approaches are further divided into deterministic and stochastic. Stochastic approaches such as genetic algorithms lead to better results but are computationally more expensive. Deterministic methods are usually stepwise algorithms where one variable at a time is added or removed from the candidate subset of features. Forward-selection is the simplest one starting from an empty set of variables. The variable that leads to greatest improvement in model behaviour is added next. The additions are continued until model behaviour no longer improves. The drawback of forward-selection is that once a variable is selected its relevance is never questioned. Thus the algorithm may be trapped into local optima (Guyon and Elisseeff 2003). Despite the drawback, forward-selection is used in this study.

The suitability of the candidate variable subset is evaluated with the root-mean-squared error (RMSE) of prediction. It is obtained from

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}_{i,CV}\right)^2}\,, \tag{3}$$

where $\hat{y}_{i,CV}$ is the predicted output obtained through leave-one-out (LOO) cross-validation procedure. Before the selection the variables are normalised so that the interpretation of their significance is easier.

3. RESULTS AND DISCUSSION

3.1 Variable Selection Results

The variable selection is carried out with the forward-selection method as described in Section 2.4. The number of variables selected greatly depends on the data set. For the data set corresponding to the ignition phase only two variables are selected. The resulting model is poor having only a correlation coefficient of 0.20 between the measured and predicted splashing. This is due to the fact that splashing hardly occurs in the ignition phase as shown in Table 1. If splashing in the ignition phase is to be modelled, the data should first be filtered so that only the heats with significant splashing are used. Minor splashing is also observed in the end of heat data set. The variable selection and subsequent modelling of splashing still produces better results compared with the ignition phase data set. This is likely to be due to the fact that in the end of heat the lance height is kept constant and also otherwise the process is more stable than in the ignition phase meaning that identifying of process circumstances is less challenging.

The variables selected for the actual heat phase data set are somewhat interesting. Altogether 9 variables are selected but only six of them are significant. Significance is evaluated from the value of the corresponding regression coefficient. First it is noticed that the amount of FeSi is not selected even though it is usually associated with splashing. Another interesting variable is the amount of P which is selected and significant but its influence on splashing is unknown. These observations which are not in correspondence with the prior knowledge need a lot more research. Anyway, the variables that are selected may be strong together with some other variable even though they are weak alone. It is also possible that they give indication about some important phenomenon which cannot be directly measured from the process.

The forward-selection used in this study may be trapped into local optima and thus select only a sub-optimal subset of variables. However, it still gives reasonable models as shown, for example, in (Sorsa et al. 2012). The weaknesses of forward-selection are that it easily gets trapped into local optima as mentioned earlier and it also tends to find variables that are strong alone but miss variable combinations that are strong together. These issues can be solved by using, for example, so called floating search or genetic algorithms. Floating search uses sequential forward-selection, backward-elimination and variable replacement steps to avoid local optima and to find the appropriate subset of variables (Nakariyakul and Casasent 2009). Floating search is a deterministic method but computationally much more expensive than forward-selection. Genetic algorithm is a stochastic methods mimicking evolution. It is computationally even more expensive but is usually reported to give better results than deterministic methods. Genetic algorithm is used for variable selection, for example, in (Sorsa et al. 2013).

3.2 Modelling Results

The modelling results are given in Table 2. The table shows that the correlation between the measured and predicted splashing is not very high. Still the present authors find the results promising because the main thing is that the algorithm found a subset of variables that is able to explain some of the variation in splashing. As mentioned earlier, the data used is very challenging because it includes quite a lot of heats with practically no splashing. This leads to a data set where the interactions cannot be unanimously detected and identified. To overcome this challenge, the approach studied here may be complemented with a data selection procedure. This means that the data must be selected so that the interactions are observable. From such data, it is possible to identify and analyse the conditions leading to splashing. An approach where both variable and data selection are considered is reported in (Isokangas et al. 2005).

Table 2. The performance of the identified models

Data set	RMSE	R
Ignition	0.19	0.20
Actual heat	1.15	0.43
End of heat	0.04	0.55
Whole data	0.83	0.38

4. CONCLUSIONS

Splashing is an undesired phenomenon in a basic oxygen furnace leading to material losses. Thus the analysis of the reasons leading to splashing is important. Earlier the analysis is carried out mainly manually. This paper presented some preliminary results of using a systematic approach for identifying the interactions between splashing and the variables measured from the basic oxygen furnace vessel. The systematic approach used forward-selection for selecting the significant variables and multivariable linear regression as a modelling technique for identifying a model between splashing and process variables. The results showed that the procedures used were able to find a variable subset that holds information for explaining some changes in splashing. The selected variables were interesting but their role was not analysed in this study. However, this issue should be studied in the future to better validate the models obtained. The prediction accuracy of the models identified was not very good but it is due to the challenging data set. Thus a careful selection of data needs to be also included in the procedure in order to find the real interactions between splashing and process variables.

REFERENCES

Alexandridis, A., Patrinos, P., Sarimveis, H., and Tsekouras G. (2005). A two-stage evolutionary algorithm for variable selection in the development of RBF neural network models. *Chemometrics and Intelligent Laboratory Systems*, 75, 149-162.

Bock, M., Schoop, J., and Oehler, C. (2000) The influence of blowing techniques, bottom stirring and lance design in the LD-processes on slag formation, the formation of lance and converter skulls and converter lining life. *3rd European Oxygen Steelmaking Conference*, Birmingham, United Kingdom, 163-177.

Boom, R. (2003) Mastering the heat in the Fe-C-O converter: Evolution of process control in fifty years of oxygen steelmaking, *4th European Oxygen Steelmaking Conference*, Graz, Austria, 1-19.

Guyon, I. and Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3, 1157-1182.

Harrell, Jr, F.E. (2001) *Regression modelling Strategies with applications to linear models, logistic regression, and survival analysis*. Springer, New York.

Isokangas, A., Ruusunen, M., and Leiviskä, K. (2005) Systematic Approach to Model-Based Data Survey. *Proceedings of the ICINCO 2005, 2nd International conference on informatics in control, automation and robotics*, Barcelona, Spain, 14-17 September 2005.

Jung, S., and Fruehan, R.J. (2000) Foaming Characteristics of BOF Slags. *ISIJ International* 40, 348-355.

Koch, K., Falkus, J., and Bruckhaus, R. (1993) Hot metal experiments of the metal bath spraying effect during the decarburization of Fe-C melts through oxygen top blowing. *Steel Research* 64, 15-21.

Kohavi, R., and John, G.H. (1997) Wrappers for feature subset selection. *Artificial Intelligence* 97, 273-324.

Luomala, M., Fabritius, T., Virtanen, E., Siivola, T., and Härkki, J. (2002) Splashing and spitting behaviour in the combined blown steelmaking converter. *ISIJ International* 42(9), 944-949.

Nakariyakul, S., and Casasent, D.P. (2009) An improvement on floating search algorithms for feature subset selection. *Pattern Recognition* 42, 1932-1940.

Ruuska, J., Ollila, S., and Leiviskä, K. (2006) Splashing in LD-KG-converter and variables having effect on it. *IFAC Workshop - MMM'2006*, Cracow, Poland, 32-37.

Ruuska, J. (2012) *Special measurements and control models for a basic oxygen furnace (BOF)*. D.Sc. Thesis, University of Oulu Graduate School. http://herkules.oulu.fi/isbn9789514298028/isbn9789514298028.pdf

Smit, S., Hoefsloot, H.C.J., and Smilde, A.K. (2008) Statistical data processing in clinical proteomics. *Journal of Chromatography B* 866, 77-88.

Sorsa, A. (2013) *Prediction of material properties based on non-destructive Barkhausen noise measurement*. D. Sc. Thesis, University of Oulu Graduate School. http://herkules.oulu.fi/isbn9789526200682/isbn9789526200682.pdf

Sorsa, A., Leiviskä, K., Santa-aho, S., and Lepistö, T. (2012). Quantitative prediction of residual stress and hardness in case-hardened steel based on the Barkhausen noise measurement. *NDT&E International*, 46, 100-106.

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

Tang, P., Yu, Y., Wen, G., Zhu, M., Zhou, L., Long, Y., and Liang, Q. (2008) Study on the optimization of the combined blown converter process in Chongqing Iron and Steel company. *Journal of University of Science and Technology Beijing* 15, 5-9.