Data based modeling approach to iron and steel making processes

Sachin Santosh[1], Ajay Kumar Shukla[2]

[1]M.Tech Student, Department of Metallurgical and Materials Engineering, IIT Madras, Chennai, 600036, India

[2]Asst. Professor, Department of Metallurgical and Materials Engineering, IIT Madras, Chennai, 600036, India

Abstract:

Iron and steel making processes are very complex in nature and we need prediction tools which can act as a guideline to control them. Various modeling techniques have been adopted in order to develop good prediction models. These models are the part of automation control systems in a steel plant. These models could be fundamental in nature based upon physical and chemical laws of the process on one hand and empirical approach on the other hand. Subject to the condition that there could be lot of variations due to error in input measurements and other uncertain factors beyond control, the actual process will always have some degree of uncertainty. Therefore models which are based upon actual plant data are more reliable as compared to the fundamental models. Even fundamental models could also be used in association with data based models where various relationships and coefficients of uncertainty are evaluated based upon actual plant data. In this paper data based modeling approach is demonstrated for BOF steelmaking process in particular. A comparative study has been done for combination of various approaches like ANN (Artificial Neural Networks), MTS (Mahalanobis Taguchi systems) and PCA (Principal Component Analysis) and MLR (multivariate regression analysis) to develop prediction models based upon industrial data.

Introduction:

Currently, oxygen steelmaking accounts for 65% of worldwide crude steel production and is thus the predominant steelmaking process. The oxygen converter utilizes oxygen as an oxidation source for reacting with other elements to convert iron into steel and increase the bath temperature. These reactions are characterized by a high reaction rate, short residence time, numerous influencing factors and complicated reaction processes. In BOF steelmaking process, composition and temperature of steel bath can't be measured continuously and operation conditions vary frequently, which makes it difficult to control the BOF end-point bath precisely. Actually, it often happens that operators have to reblow the steel bath due to the low control precision of end-point bath. So improving the control precision of BOF steelmaking end-point is quite important. Earlier, the steel industry used to rely on fundamental heat and mass balance methods to predict parameters like the temperature and blow time and input weights required like tonnes of oxygen required. But this process is highly time-consuming and in many cases, inaccurate, owing to the factors mentioned above. Due to the coexistence of several phases and the complex flow conditions with mass and heat transfer inside, a steel making furnace is very difficult to model. For many years, furnace operators have been aware of the fact that there are no universally accepted methods for accurately controlling complex iron and steelmaking operation and predicting the outcome. Our task is to develop a predictive model which could be developed using the operational data of steelmaking process. The models developed under this category uses data based techniques, particularly multiple linear regression and artificial neural networks (ANN) along with reduction in dimensionality of the problem using Mahalonobis method (MTS) and Principal component analysis (PCA). In this paper the application and the advantages of using these techniques is explained in detail in the next section.

BOF Steel making Process

The Basic Oxygen Steel-making (BOS) process converts hot metal, from blast furnace, and scrap into steel by exothermic oxidation of metalloids dissolved in the iron. Oxygen also combines with carbon, eliminating the impurities by gas collection. The main purpose of this process is the carbon percentage decrease: from approximately 4% in hot metal to less than 0.08% in liquid steel. BOF steelmaking process is executed to raise the bath temperature and reduce the impurity level by blowing proper volume of oxygen into the steel bath surface and adding appropriate amount of flux and coolant into the bath. The main raw materials of the process include main materials (such as hot metal, scrap, pig iron) and sub-material (oxygen, iron ore, lime, dolomite and etc.), and the product is the steel bath of which the temperature and composition are required to hit the tapping aim window.

Need for process control in a BOF steel making furnace

The quantity of oxygen utilized plays an important role in determining the steel quality. Specifically, if the amount of oxygen injected is too small, the endpoint carbon content will exceed the required value or the endpoint temperature may be too low. If the amount of oxygen is too large, the molten steel will be over-oxidized, the consumption of alloys will increase, the temperature may be too high and the yield of liquid steel will decrease. Therefore, determination of the exact oxygen blowing quantity has tremendous influence on the steelmaking process. According to the characteristics of BOF steelmaking process, the control method combining the Static Process Control with Dynamic Process Control is popularly used. Static Process Control determines the gross requirement of oxygen and coolant for the each heat based on the initial information, when sub-lance SL1 measurement is processed successfully in the posterior period, Dynamic Process Control is started to adjust the dynamic requirement of oxygen and coolant based on the measurement result of bath [C]

Possible data driven approaches

1. **Fundamental approach**

   The BOS is a very complex chemical batch process. The amount and quality of scrap iron change from batch to batch; the grades of steel produced can change frequently and also changes the vessel shape during the campaign lifetime. A first principles model—called charge balance or static model—which is a complete heat, mass and chemistry balance of the steel-making process is used to predict total oxygen blow necessary to each batch. However, model mismatches and the unsteady-state nature of decarburization rate lead to a poor control in end-point temperature and carbon percentage.

2. Linear regression

The multiple linear regression model is based on the utilization of a large amount of production data; therefore data from nearly 1 000 heats of the same campaign need to be collected from steel plants. Before incorporating the production data into the model, the data is filtered and treated. The principles of filtration and treatment include removal of the variables which do not affect the model and omission of abnormal values of the variables so that the production data meets the actual requirements. The selection of independent variables plays a key role in establishing the model. The reactions that occur in the molten steel bath of a converter are very complex, and end-point manganese content is affected by numerous interacting factors. Therefore, in order to provide an adequate description of the entire melting process and clarify the model, the multiple linear regression models employs those factors which change dramatically and play a key role in the BOF steelmaking process as the basic variables. Broadly speaking a regression model assigns certain weights for each contributing factor in such a way that the equation of a straight line is satisfied for the maximum number of points. Say if we wish to predict the end point Manganese and using some contributing factors. Linear regression fits a straight line into the plot for the graph of contributing factors vs. manganese % plot. The line with the best fit (or highest R squared value) gives us the most accurate curve. Of course, there is no justification for the choice of the particular form of relationship. This and other difficulties associated with ordinary linear regression analysis can be summarized as follows: (a) A relationship has to be chosen before analysis. (b) The relationship chosen tends to be linear, or with non-linear terms added together to form a pseudo linear equation. (c) The regression equation, once derived, applies across the entire span of the input space.

3. Artificial Neural Networks (ANN)

With the development of artificial intelligence, some control methods based on neural network or neural network combined with algorithms have been widely used in BOF end-point control[1-5]. ANNs represent an alternative computational paradigm in which the solution to a problem is learned from a set of examples. The concept of ANN originally comes from the mechanisms for information processing in human brain system. ANN models has been applied to the wide range of complex metallurgical processes[1-5] and proved to be successful due to its ability to develop non linear relationships. ANNs are the mathematical patterns constructed by several neurons arranged in different layers interconnected through the complex networks. The layers are defined as input layer, output layer and at least one hidden layer. A multilayer feed forward back propagation ANN network has been used in present work. The typical ANN topology is presented in Fig. 1.

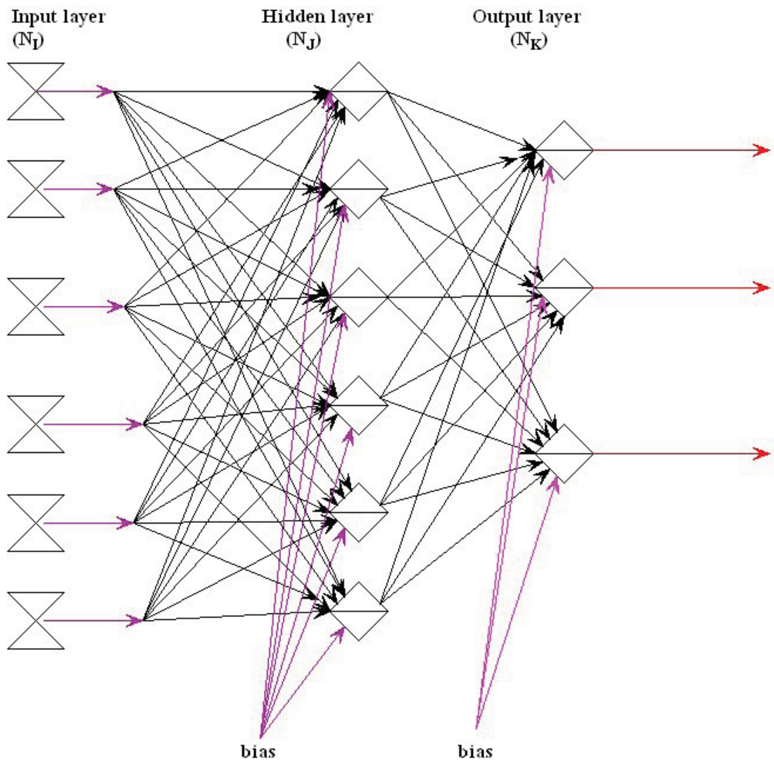


Fig. 1: Architecture of feed-forward back propagation ANN

The output of a neuron (k) in the network (y_k) is the summation of all signals from previous layer multiplied by weights ($w_{k,j}$) and a bias (b_k) which is activated by a transfer function (tanh sigmoid) in the following way:

$$y_k = f\left(\sum_{j=1}^{N}\left(w_{k,j}.x_j\right)+b_k\right) \text{ where } f(z)=\frac{2}{1-\exp(-2.z)}-1 \qquad (1)$$

The sum of the square of the errors (between the training output data and output data obtained using ANN) are minimized for getting the correct values of weights.

4. Principal Component Analysis (PCA)

Principal component analysis is done for reducing the dimensionality of data set. The Principal components are calculated which are orthogonal to each other and all variables can be defined by principal components. Finally only those principal components are considered for analysis which have more than 90% cumulative sum of variances.

5. Mahalanobis Taguchi System (MTS)

Mahalanobis-Taguchi system is used to minimize the number of variables (or control factors) required to predict the performance of a system. It is based upon the calculation of Mahalanobis distance, Mahalanobis space to be used to discriminate between normal and abnormal data followed by reduction using orthogonal array and signal to noise ratio to calculate the effect of each variable. The reduction in dimensionality of the problem is based on Mahalanobis distances and signals to noise ratios[6-8].

Result analysis of data driven model developed for phosphorus prediction in BOF steelmaking process:

Data drive based models have been developed for the prediction of end point phosphorous for BOF steelmaking process. Table 1 gives the details of the steel plant data (400 in numbers) used for calculation.

Table 1: Range, mean and standard deviation of the data set used for investigation

Parameters	Maximum	Minimum	Mean	Standard deviation
LIME (wt of lime (tons))	20.7	5.80	12.04	1.90
HMASI ((Hot Metal Silicon (wt%))	1.61	0.07	0.87	0.16
HMP (Hot Metal Phosphorous (wt %))	0.29	0.20	0.24	0.02
HMA_TEMP (Hot Metal Temperature (^oC))	1360	1201	1291	28
HMWT_ACT (Hot Metal Weight (tons))	158	113	135	9.34
SCP_ACT (Wt of Scrap (tons))	26.40	0.0	15.71	9.15
ORE (Wt of Iron ore (tons))	10.70	0.90	4.52	2.20
OXY_ACT (Oxygen blown (NM3))	7760	5813	6699	357
SL_FE (Fe Level of the slag (Wt %))	26.50	14.60	19.50	2.00
EB_TEMP (Temperature at End of Blow (^oC))	1749	1611	1671	25
SL_P2O5 (P2O5 in slag (Wt %))	5.00	2.44	3.60	0.40
CaO/SiO2 (Basisity of the slag)	4.50	2.90	3.90	0.25
TDP (turndown phosphorous wt%)	0.020	0.010	0.010	0.003

The following Steps were performed for the calculations:

1. Estimation of Correlation matrix among variables.
2. Estimation of Mahalanobis distances and Gain for each variable
3. Selection of most significant variable based upon the gain value

4. Performing Multiple linear regression analysis with the selected variables.
5. Selection of significant variables by ‘t’ test
6. Repeat of step 4 and 5 till we get most significant variables by t test (**MTS-MLR method**).
7. Repeat of Step 4 to 6 with the selection of all variables (**MLR method**).
8. Performing ANN calculations with different network topologies for variables finally selected in step 6 (**MTS-MLR-ANN method**).
9. Performing ANN calculations with different network topologies for variables finally selected in step 7 (**MLR-ANN method**).
10. Performing ANN calculations with different network topologies with selection of all variables (**ANN method without reduction in dimensionality**).
11. Performing Principal component analysis (PCA) calculations and deriving relevant Principal components for all data variables.
12. Performing Multiple linear regression analysis followed by stepwise regression based upon‘t’ test (**PCA-MLR method**).

The correlation matrix is given in following table:

Table 2: Correlation matrix

	LIME	HMA_SI	HMA_P	HMA_TI	HMWT_	SCP_A	ORE	OXY_ACT	EB_TE	SL_FE	SL_P2(	CaO/S	TDP
LIME	1.00												
HMA_SI	0.56	1.00											
HMA_P	0.01	0.05	1.00										
HMA_TEMP	-0.13	-0.17	-0.03	1.00									
HMWT_ACT	0.50	-0.05	0.02	0.07	1.00								
SCP_ACT	-0.49	0.04	0.00	-0.01	-0.88	1.00							
ORE	0.56	0.25	0.10	0.20	0.75	-0.81	1.00						
OXY_ACT	0.06	0.00	-0.06	-0.17	-0.18	0.38	-0.37	1.00					
EB_TEMP	0.09	-0.13	-0.08	-0.15	0.13	-0.16	-0.04	0.29	1.00				
SL_FE	0.15	0.01	-0.06	-0.08	0.22	-0.24	0.18	0.04	0.00	1.00			
SL_P2O5	-0.33	-0.39	0.27	0.11	-0.07	0.11	-0.17	0.03	0.08	-0.49	1.00		
CaO/SiO2	0.03	-0.37	-0.08	0.02	0.22	-0.22	0.02	0.10	0.13	0.12	0.23	1.00	
TDP	-0.21	-0.16	0.03	-0.10	-0.18	0.17	-0.27	0.15	0.54	-0.13	0.14	-0.10	1.00

As it can be seen that phosphorous has got strongest correlation with EB_TEMP followed by OXY_ACT and SCP_ACT. The interdependence among different variables is also evident from above table.

In MTS-MLR method first of all MTS run was done. In MTS run following variables were selected (for variables having positive gain values as given in Table 3.

Table 3: The selected variables and their positive gain values after MTS run

HMA_P	HMA_TEMP	HMWT_ACT	OXY_ACT	EB_TEMP	SL_FE	SL_P2O5	CaO/SiO2
1.9467	0.7808	0.2189	0.7244	0.5844	0.5217	0.0313	0.5307

Multiple Regression analysis using above selected variables followed by step-wise regression and successive reduction of variables after 't' test finally selects LIME, HMWT_ACT. EB_TEMP and CaO/SiO2 (statistical performance given in Table 4):

Table 4: Statistical performance of MTS-MLR model

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	-0.095478988	0.008351118	-11.43307878	2.32082E-26
LIME	-0.000324903	7.41834E-05	-4.379719093	1.52226E-05
HMWT_ACT	-4.15385E-05	1.54942E-05	-2.680895095	0.007648581
EB_TEMP	7.48028E-05	5.02834E-06	14.87624737	4.26512E-40
CaO/SiO2	-0.001740103	0.000497778	-3.495740179	0.000526106

The predictive performance of MTS-MLR model is plotted in Fig. 2.

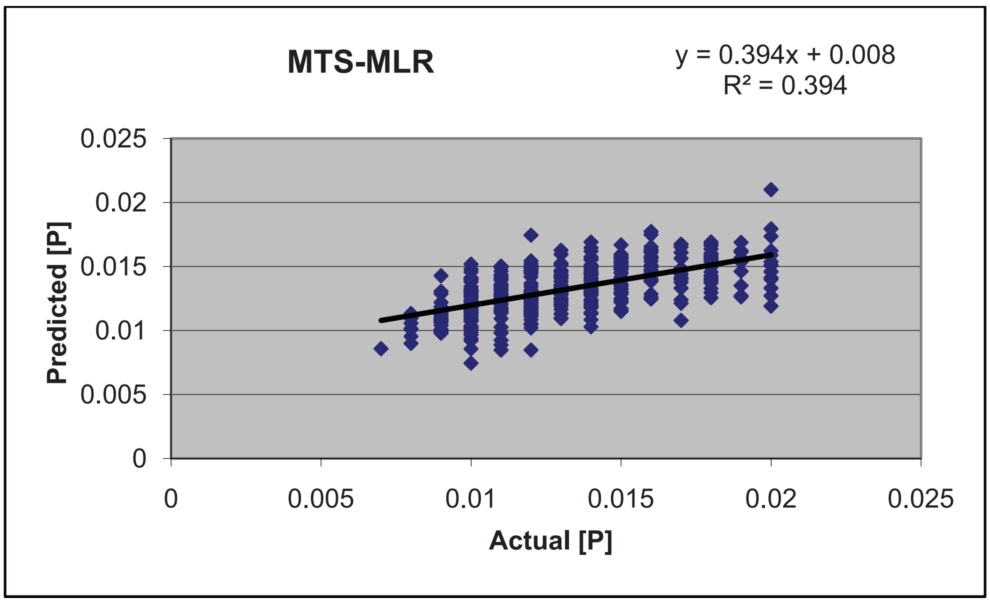


Fig. 2: Predictive performance of MTS-MLR model

In MTS-MLR-ANN method, Neural network model was developed by using finally selected variables in MTS-MLR method. Predictive performance of MTS-MLR-ANN model is plotted in Fig. 3.

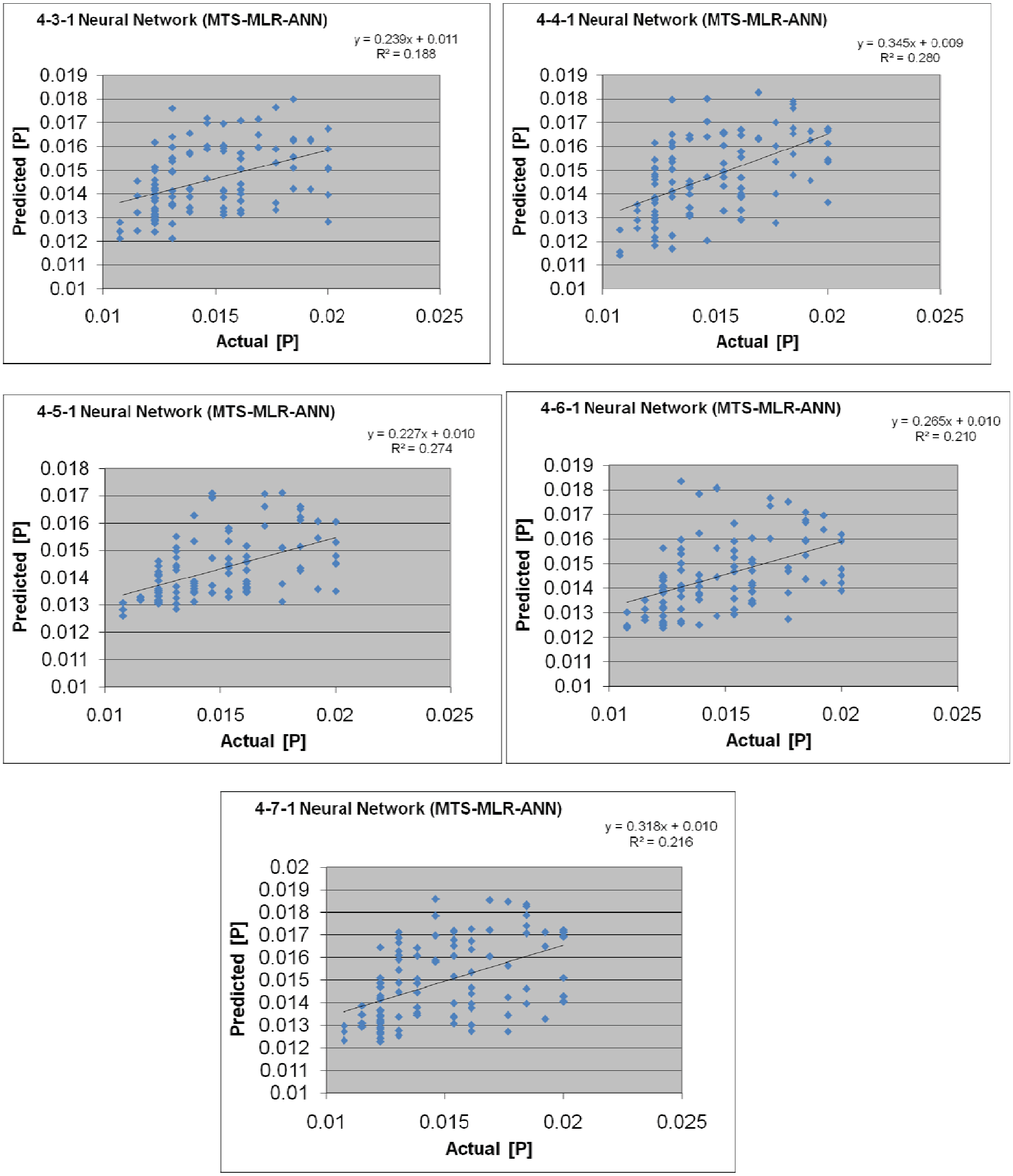


Fig. 3: Predictive performance of MTS-MLR-ANN models

In MLR-ANN method, first of all Multiple linear regression is performed using all 12 variables followed by step-wise regression and successive reduction of variables after ‘t’ test which finally selects SCP_ACT, OXY_ACT. EB_TEMP and CaO/SiO2 (statistical performance given in Table 5).

Table 5: Statistical performance of MLR model

	Coefficients	Standard Error	t Stat	P-value
Intercept	-0.111081634	0.008684	-12.7908	1.31E-31
SCP_ACT	0.000103661	1.59E-05	6.537584	1.92E-10
OXY_ACT	-1.20841E-06	4.1E-07	-2.94548	0.003414
EB_TEMP	8.11189E-05	5.5E-06	14.75673	1.34E-39
CaO/SiO2	-0.001239299	0.000508	-2.43904	0.015164

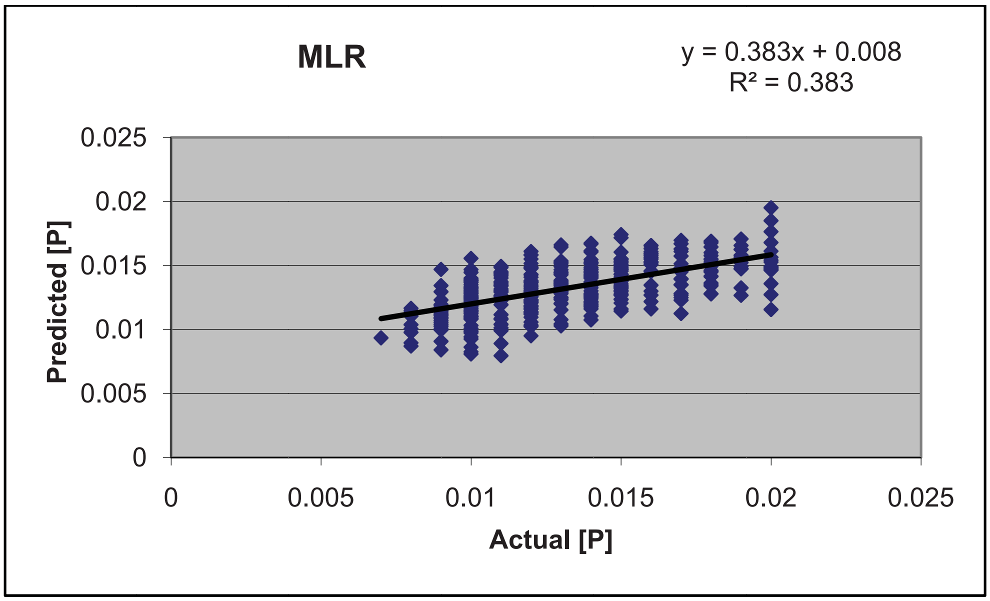
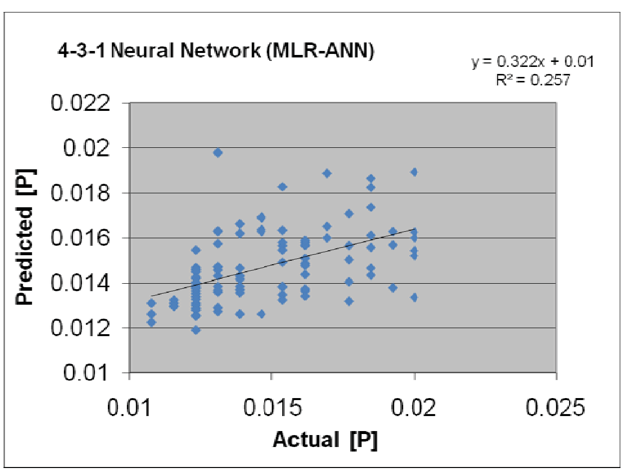
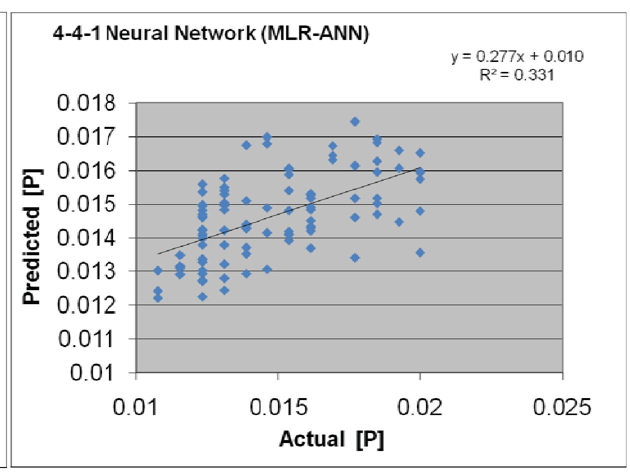


Fig. 4: Predictive performance of MLR model

ANN model is developed using variables finally selected by MLR method. Predictive performance of MLR-ANN model is plotted in Fig.





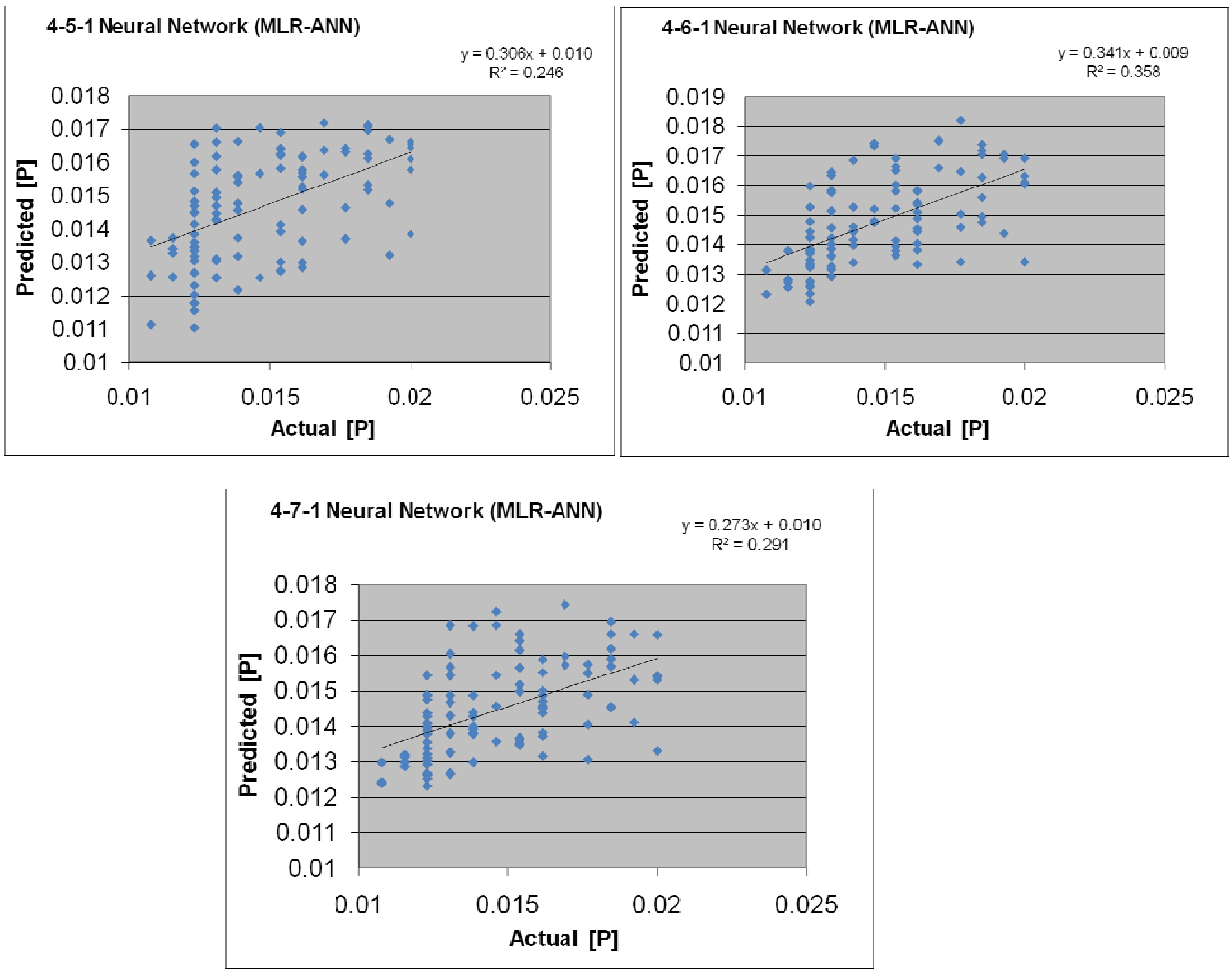
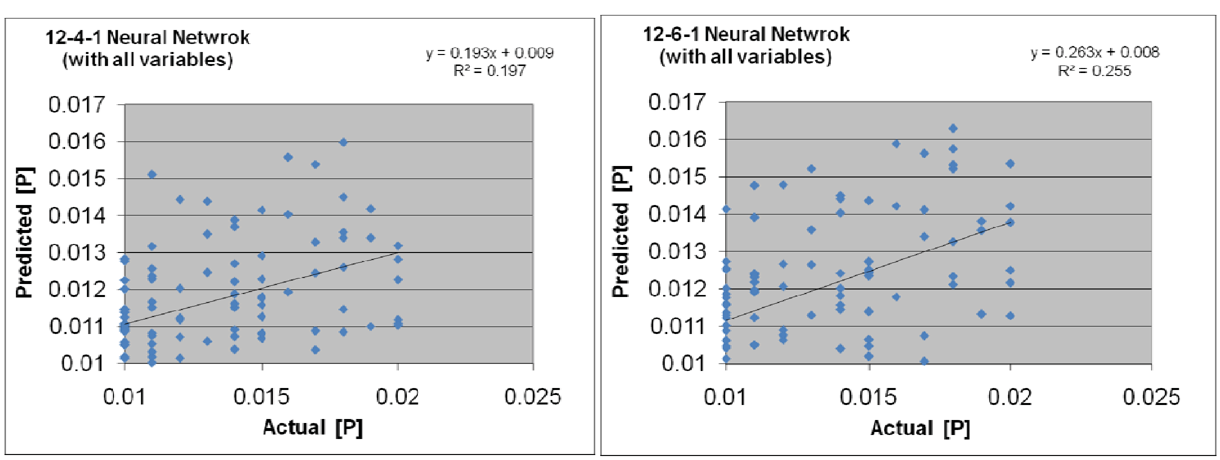


Fig. 5: Predictive performance of MLR-ANN models with different ANN topologies

ANN models were also developed without reduction in dimensionality of the problem (using all 12 variables). The predictive performance of ANN model (without dimensionality reduction is plotted in Fig. 6.



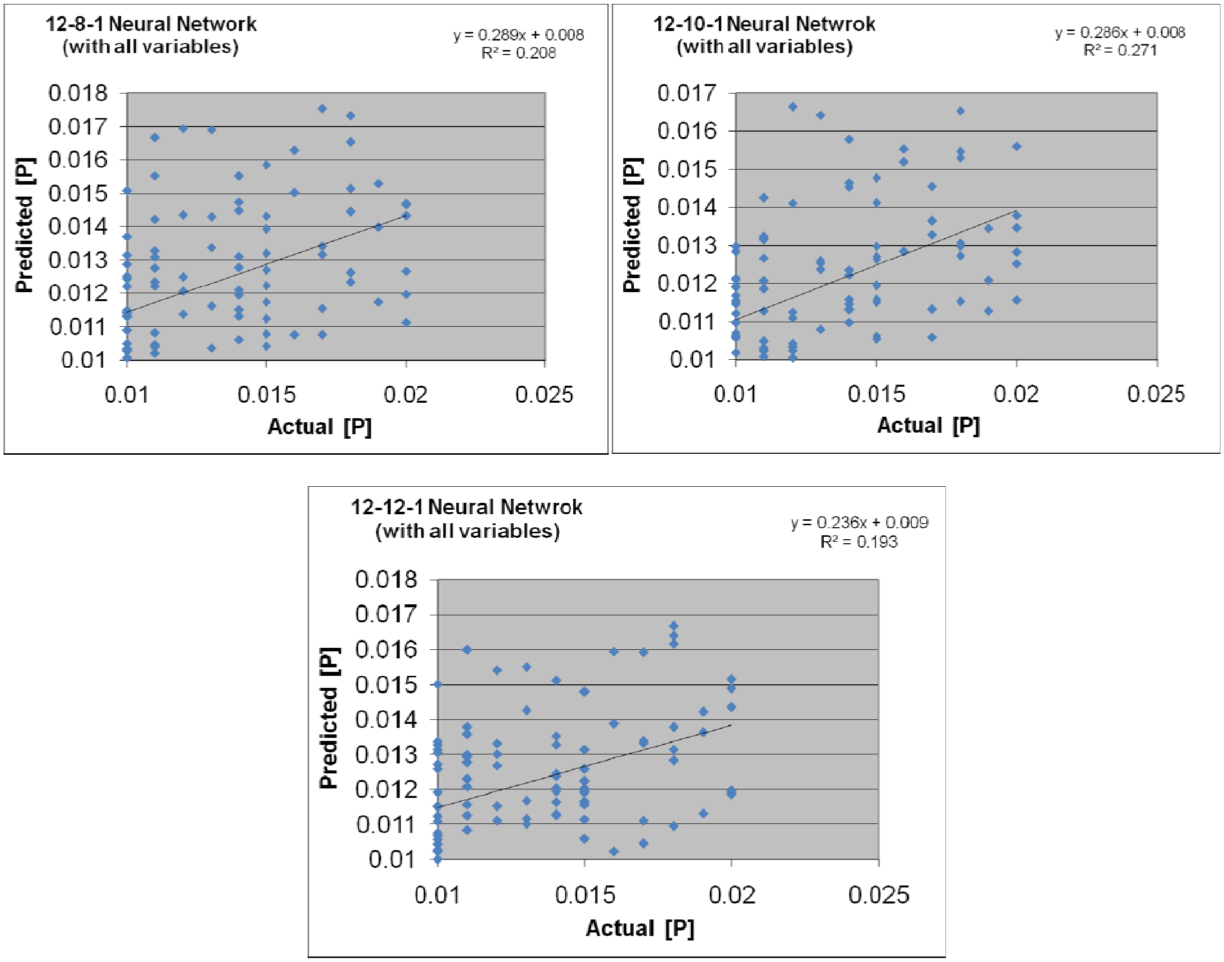


Fig. 6: Predictive performance ANN models with all variables for different topologies

Based upon Principal Component analysis using MATLAB for given data set, the following principal components are calculated for more than 90% cumulative variance (Table 6).

Table 6: Selected Principal Components (for > 90% cumulative variance)

	Coeff_Princomp1	Coeff_Princomp2	Coeff_Princomp3	Coeff_Princomp4
LIME	-0.000298829	-0.010145147	-0.006702508	0.089812428
HMA_SI	3.30E-07	-0.000713899	0.001339821	0.00026098
HMA_P	3.06E-06	-1.34E-05	5.60E-05	3.46E-05
HMA_TEMP	0.0130596	0.950162146	-0.309832054	-0.018530731
HMWT_ACT	0.004706269	-0.008619643	-0.114085367	0.731035508
SCP_ACT	-0.00981226	0.047189526	0.130666009	-0.636634756
ORE	0.0022988	0.00798071	-0.01753033	0.136681841
OXY_ACT	-0.999655022	0.01804258	0.01268418	0.013206178
EB_TEMP	-0.019884965	-0.307198914	-0.934555081	-0.175068211
SL_FE	-0.00024973	-0.005020839	0.001335654	0.048200327
SL_P2O5	-3.54E-05	0.001304467	-0.001737459	-0.004459078
CaO/SiO2	-7.20E-05	5.39E-05	-0.001447738	0.00501806

Multiple linear regression is performed using these principal components as variables are giving following results (Table 7):

Table 7: Statistical performance of PCA-MLR model using selected principal components

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.013216104	0.000133358	99.10259195	4.0634E-282
Princomp1	-831133312.8	375415963.2	-2.213899765	0.027403336
princomp2	831133312.8	375415963.2	2.213899765	0.027403336
princomp3	-2.19824E-05	4.77063E-06	-4.607860563	5.49031E-06
princomp4	-5.90027E-05	5.73514E-06	-10.28793552	3.65163E-22

Predictive performance of PCA-MLR model is plotted in Fig. 7.

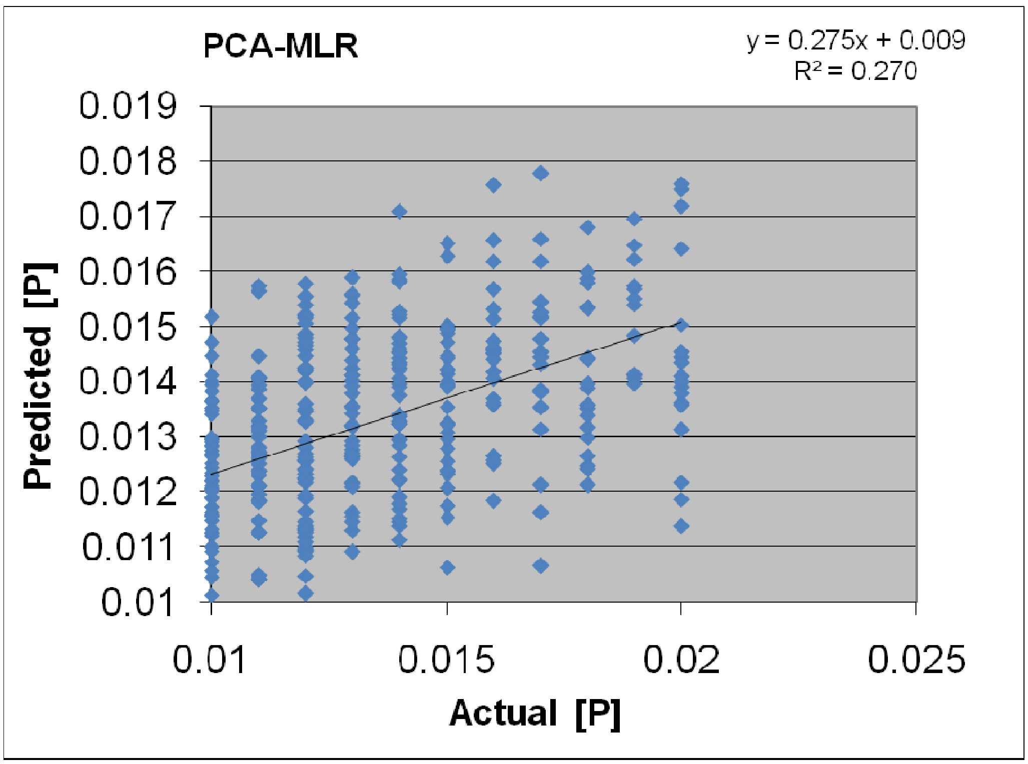


Fig. 7: Predictive performance of PCA-MLR model

Conclusions:

Data driven models for the prediction of phosphorus using industrial data are developed using ANN, MLR, MLR-ANN, MTS-MLR-ANN, PCA-MLR approaches. The relative performances of all these models are given in Table 8:

Table 8: Predictive performance (R^2) of various data driven models

Model	Predictive performance (R^2)
MLR	0.383
MTS-MLR	0.394
MTS-MLR-ANN	0.280 (best using 4-4-1 ANN network)
MLR-ANN	0.358 (best using 4-6-1 ANN network)
ANN (with all variables)	0.271 (best using 12-10-1 network)
PCA-MLR	0.270

Based upon the performance of various data based models, performance of MTS-MLR approach is found to be best followed by MLR and MLR-ANN (4-6-1). Reduction of dimensionality of the problem using MTS or PCA approach is always suggested to deal with lesser number of control variables. The performance of any data based model depends upon the distribution of data of the concerned process. In general linear regression models should work better if range of variation is not so large for different variable data which is the case for most of industrial steelmaking processes which are operated in well defined and small domain of variations. Application of ANN models does not yield better performance due to noise and chaotic nature of the process.

References:

1. Rajesh N., Khare M.R. and Pabi S.K.: *Materials Research*. 2010, vol. 13(1), pp 15.

2. Shukla A.K., Deo B. and Robertson D.G.C.: *Metallurgical and Materials Transactions B*, July 2013.

3. Fileti A.M.F., Pacianotto T.A. and Cunha A. P.: *Engineering Applications of Artificial Intelligence*, 2006, vol. 19, pp 9.

4. Das A., Maiti J., Banerjee R.N.: *Expert Systems with Applications*, 2010, vol. 37, pp 1075.

5. Cox I.J., Lewis R.W., Ransing R.S., Laszczewski H., Berni G.: *Journal of Materials Processing Technology* , 2002, vol. 120, pp 310.

6. Taguchi G. and Rajesh J: *Sankhya-The Indian Journal of Statistics*, 2000, vol. 62(B), pp 233.

7. Cudney E.A., Hong J., Jugulum R., Paryani K., Ragsdell K.M. and Taguchi G.: *Journal of Industrial and Systems Engineering*, 2007, vol. 1(2), pp 139.

8. Bagchi T.P.: *Taguchi Methods Explained*, Prentice Hall of India, New Delhi (1993).