

MODELING OF SELECTIVE CO OXIDATION OVER $\text{CuO}_x\text{-CoO}_x\text{-CeO}_2$
CATALYSTS USING ARTIFICIAL NEURAL NETWORKS

by

Ufuk Kantar

B.S. in Ch.E., Boğaziçi University, 2004

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Chemical Engineering
Boğaziçi University

2008

ACKNOWLEDGEMENTS

Firstly, I want to express my truthful gratitude to my thesis supervisor Assoc. Prof. Ramazan Yıldırım, throughout my all thesis, his quidance and encouragement was invaluable for me. Also, I wish thank to Prof. Erhan Aksoylu and Prof. Fikret Gürgen for the time they devoted to me, it was a luck for me to benefit from their experience and knowledge.

Our department, has dedicated academicians, without their sacrifice the success would not be possible. My intimate thanks are to Prof. Uğur Akman, Prof. Zeynep ilsan, Prof. Öner Hortaçsu, and to all lecturers and researchers.

My special thanks are to Ph.D. Students Göktuğ Özyönüm and Assistant Erdem Günay for their friendship and suggestions. I wish them success in their doctorate study.

Cordial thanks are for Bilgi Dedeoğlu, Nurettin Bektaş and Yakup Bal, every achievement in the department have a part from their labor.

My Boğaziçi University desire have started with my cousin Alper Gelin (AD'04), prior to ÖSS Exam, we had come here and looked around together. And after 8 years, I am leaving with a master degree in chemical engineering. I thank Alper for all his suggestions and help through my university education.

Finally, my thanks are to my family for their patience and support. My sister Selen and brother Can encouraged me all time with their moral support.

ABSTRACT

MODELING OF SELECTIVE CO OXIDATION OVER CuO_x-CoO_x-CeO₂ CATALYSTS USING ARTIFICIAL NEURAL NETWORKS

This study is aimed to model design and reaction parameters of CO oxidation over CuO_x-CoO_x-CeO₂ catalyst using neural and modular neural networks. Since there are many possible architectures for modular networks, the best network is searched using the measures of correlations such as R^2 , R^2_{adj} , and RMSE. In all models, first a small network is constructed and enlarged without over fitting the data. Then the best networks were optimized with Quasi – Newton method using the MATLAB® Optimization Toolbox. The significance of input parameters and their effects on CO conversion were also analyzed. The data was also modeled using the multiple regression method for comparison.

The effects of precipitation pH and temperature were modeled with a neural network of two neurons in the hidden layer with an R^2 value of 0.970. The precipitation and reaction conditions were also modeled together with a neural network of 3-1 structure, and modular neural networks of 2-3-1 and 1-1-1 structures with the R^2 values of 0.981, 0.989 and 0.971 respectively indicating that both models can be used. However, it was found that the modular neural networks have some advantages; they feed the similar input parameters into the same module which enhances the modeling power, and decreases the possibility of over-fitting by lowering the number of connections for the same number of data points. Similar models were developed for various combinations of catalyst preparation conditions (temperature and pH), target metal loadings (Cu, Ce and Co weight per cent) and reaction parameters (temperature, W/F and time on stream) with considerable success. It was also observed that the effects of measured catalyst properties (actual metal loadings and total surface area) on CO conversion can also be modeled successfully using both neural and modular neural networks.

ÖZET

YAPAY SINİR AĞLARI İLE $\text{CuO}_x\text{-CoO}_x\text{-CeO}_2$ KATALİZÖRDE SEÇİMLİ CO OKSİDASYONUNUN MODELLENMESİ

Bu çalışmada sinir ve modüler sinir ağları kullanılarak CO oksidasyonu için $\text{CuO}_x\text{-CoO}_x\text{-CeO}_2$ katalizörlerinin tasarım ve reaksiyon parametreleri modellenmiştir. Modüler sinir ağları değişik yapılarda kullanılabileceğinden, en iyi ağ mimarisi korelasyon ölçümleri kullanılarak aranmış, her ağ için R^2 , R^2_{adj} ve RMSE değerleri hesaplanarak karşılaştırılmıştır. Her model için önce küçük bir ağ kurulmuş ve veriyi ezberlemeyeceği şekilde performansını artıracak ölçüde genişletilmiştir. En iyi ağlar MATLAB® Optimizasyon Kiti kullanılarak Quasi-Newton metodu ile optime edilmiştir. Ayrıca, CO dönüşümü üzerine girdi parametrelerinin önemi ve etkileri de analiz edilmişlerdir. Karşılaştırma yapabilmek amacıyla deneysel veriler çoklu regresyon metodu ile de modellenmiştir.

Çöktürme pH'ı ve sıcaklığı, saklı katmanında iki nöronu bulunan sinir ağı ile modellenmiş ve 0.970'lik R^2 değeri elde edilmiştir. Çökeltme ve reaksiyon şartları 3-1'lik sinir ağı ile 2-3-1 ve 1-1-1'lik modüler ağlar kullanılarak modellenmiş ve sırasıyla 0.981, 0.989 ve 0.971'lik R^2 değerleri elde edilmiştir. Bu sonuçlar her iki modelleme biçiminin de kullanılabileceğini gösterse de modüler yapay sinir ağlarının, değişik girdi gruplarını ayırması ile modelleme gücünü artırma avantajı olduğu belirlenmiştir. Ayrıca, modüler yapı nöronlar arasındaki bağlantı sayısını indirmekte ve böylece modelin veriyi ezberleme olasılığını da azalmaktadır. Katalizör hazırlama koşullarının (çökeltme pH'ı ve sıcaklığı), hedef metal yüzdelerinin (Cu, Co ve Ce ağırlık yüzdeleri) ve reaksiyon parametrelerinin (sıcaklık, W/F ve kalış süresi) çeşitli kombinasyonlarını için benzer modeller makul başarı ile geliştirilmiştir. Ayrıca ölçülmüş katalizör özelliklerinin (metal yüzdeleri ve katalizör yüzey alanı) CO dönüşümü üzerine etkilerinin de sinir ve modüler sinir ağları ile başarılı bir şekilde modellenildiği görülmüştür.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	iii
ABSTRACT.....	iv
ÖZET.....	v
LIST OF FIGURES.....	ix
LIST OF TABLES.....	xii
LIST OF SYMBOLS / ABBREVIATIONS.....	xiv
1. INTRODUCTION.....	1
2. LITERATURE SURVEY.....	3
2.1. CO Oxidation.....	3
2.1.1. CO Oxidation Catalysts.....	4
2.1.2. Co Oxidation Over $\text{CuO}_x\text{-CoO}_x\text{-CeO}_2$ Catalysts.....	5
2.2. Artificial Neural Network (ANN) Modeling.....	6
2.2.1. Backpropagation Algorithm.....	7
2.2.2. Levenberg-Marquardt Algorithm.....	10
2.3. Modular Neural Network (MNN) Modeling.....	11
2.4. Modeling of CO Oxidation Using ANN.....	11
2.5. Multiple Regression.....	12
2.5.1. Quadratic Response Surface Models.....	13
2.6. Measures of Regression.....	14
2.6.1. Residual Analysis.....	14
2.6.2. Sum of Squares due to Error (SSE).....	15
2.6.3. Root Mean Square Error (RMS).....	15
2.6.4. R-Square.....	15
2.6.5. Degrees of Freedom Adjusted R-Square.....	17
2.6.6. k-Fold Cross Validation.....	17
2.7. Optimization of Neural Network.....	18
2.7.1. Constrained Optimization.....	19

2.7.2. Sequential Quadratic Programming (SQP).....	20
3. COMPUTATIONAL DETAILS.....	22
3.1. Experimental Data.....	22
3.1.1. Effects of Preparation Conditions (Precipitation pH and Temperature).	22
3.1.2. Effects of Reaction Conditions (Reaction Temperature, W/F Ratio, Sampling Time).....	23
3.1.3. Metal Percent Optimization of $\text{CuO}_x\text{-CoO}_x\text{-CeO}_2$ Catalyst.....	24
3.2. Computational Approach.....	24
4. RESULTS AND DISCUSSION.....	27
4.1. Neural and Modular Neural Network Modeling of CO Conversion.....	27
4.1.1. Modeling the Effects of Preparation Conditions on CO Conversion....	27
4.1.2. Modeling the Effects of Preparation and Reaction Conditions Together on CO Conversion.....	31
4.1.3. Modeling the Effects of Target Metal Content on CO Conversion.....	38
4.1.4. Modeling the Effects of Reaction Conditions on CO Conversion.....	41
4.1.5. Modeling the Effects of Design and Reaction Studies Together on CO Conversion.....	44
4.1.6. Modeling of CO conversion from Measured Catalyst Properties.....	48
4.1.7. Modeling of CO Conversion from Measured Catalyst Properties and Reaction Conditions.....	51
4.1.8. Modeling of CO Conversion from Design and Reaction Studies Together with Measured Metal Content.....	54
4.2. Optimization of Artificial Neural Networks.....	57
4.2.1. Optimization of Precipitation pH and Temperature.....	57
4.2.2. Optimization of Target Metal Contents.....	58
4.3. Input Significance Analysis of the Neural Models.....	59
4.3.1. Input Significance of the Preparation Conditions.....	59
4.3.2. Input Significance of the Preparation and Reaction Conditions.....	59
4.3.3. Input Significance of the Design Studies and Reaction Conditions.....	60
4.4. Analysis of Factor Effects for Modular Neural Models.....	61

5. CONCLUSIONS AND RECOMMENDATIONS.....	63
5.1. Conclusions.....	63
5.1.1. Conclusions for Modeling of Preparation Conditions.....	63
5.1.2. Conclusions for Modeling of Preparation and Reaction Conditions.....	64
5.1.3. Conclusions for Modeling of Catalyst Metal Content and Total Surface Area.....	64
5.1.4. Conclusions for Modeling of Reaction Conditions.....	65
5.1.5. Conclusions for Modeling of Design and Reaction Studies Together....	65
5.1.6. Conclusions for Optimization of Neural Networks.....	65
5.1.7. Conclusions for Input Significance and Factor Effects Analysis.....	66
5.2. Recommendations.....	66
REFERENCES.....	68

LIST OF FIGURES

Figure 2.1.	Schematic representation of a neuron.....	6
Figure 2.2.	Backpropagation algorithm.....	9
Figure 4.1.	Schematic representation of the NN 2-1 model.....	28
Figure 4.2.	Comparison of NN 2-1 and quadratic multiple regression results with experimental data.....	29
Figure 4.3.	Residual analysis for the (a) NN 2-1 and (b) quadratic multiple regression models.....	30
Figure 4.4.	Schematic representation of the (a) NN 3-1 and (b) MNN 2-3-1 models	32
Figure 4.5.	Comparison of the (a) NN 3-1 and (b) MNN 2-3-1 results with experimental data.....	34
Figure 4.6.	Residual analysis for the NN 3-1 and MNN 2-3-1 models.....	34
Figure 4.7.	Schematic representation of the MNN 1-1-1 model.....	35
Figure 4.8.	Residual analysis for the MNN 1-1-1 and quadratic MR models.....	37
Figure 4.9.	Schematic representation of the ANN 2-1 model.....	38
Figure 4.10.	Comparison graphs for the (a) NN 4-1 and (b) quadratic multiple regression results with experimental data.....	40
Figure 4.11.	Residual analysis of the NN 2-1 and multiple regression models.....	40

Figure 4.12.	Schematic representation of the MNN 2-1-2-1 model.....	41
Figure 4.13.	Comparison graphs for the (a) NN 4-1 and (b) MNN 2-1-2-1 results with experimental data.....	43
Figure 4.14.	Residual analysis for the NN 4-1, MNN 2-1-2-1 and multiple regression models.....	43
Figure 4.15.	Schematic representation of the MNN 3-3-1 model.....	45
Figure 4.16.	Comparison graphs for the (a) NN 3-1 and (b) MNN 3-3-1 results with the experimental data.....	47
Figure 4.17.	Residual analysis of the NN 3-1 and MNN 3-3-1 models.....	47
Figure 4.18.	Schematic representation of the MNN 2-1-1 model.....	48
Figure 4.19.	Residual analysis for the (a) NN 2-1 and (b) MNN 2-1-1 models.....	50
Figure 4.20.	Residual analysis for the NN 2-1, MNN 2-1-1 and MR models.....	50
Figure 4.21.	Schematic representation of the MNN 1-1-1 model.....	51
Figure 4.22.	Comparison graphs of the (a) NN 2-1, (b) MNN 1-1-1 results with experimental data.....	53
Figure 4.23.	Residual analysis for the NN 2-1 and MNN 1-1-1 models.....	53
Figure 4.24.	Schematic representation of the MNN 2-2-1 model.....	54
Figure 4.25.	Comparison of the (a) NN 3-1 and (b) MNN 2-1-1 results with the experimental data.....	56

Figure 4.26. Residual analysis for the ANN 3-1 and MNN 2-1-1 models.....	56
Figure 4.27. Factor effects graphs for preparation and reaction condition.....	62

LIST OF TABLES

Table 3.1.	Experimental total surface area, metal percents, conversion and selectivity according to precipitation pH and temperature data.....	23
Table 3.2.	Experimental conversion and selectivity values for the respective reaction temperature and weight of catalyst over flow rate ratio values.	23
Table 3.3.	Experimental total surface area, conversion and selectivity data for the corresponding target and measured metal percents.....	24
Table 4.1.	Results for the NN 2-1 and quadratic multiple regression models.....	28
Table 4.2.	k-fold cross validation analysis for the NN 2-1 model.....	30
Table 4.3.	Comparison of regression coefficients of different network architectures for preparation and reaction conditions modeling.....	31
Table 4.4.	Results for the NN 3-1 and MNN 2-3-1 models.....	33
Table 4.5.	Results for the MNN 1-1-1 and quadratic multiple regression models...	36
Table 4.6.	k-fold cross validation analysis for the MNN 1-1-1 model.....	37
Table 4.7.	Results for the NN 2-1 and quadratic multiple regression models.....	39
Table 4.8.	Result obtained from the NN 4-1 and MNN 2-1-2-1 models.....	42
Table 4.9.	Comparison of regression coefficients of different network architectures for design and reaction parameters modeling.....	44
Table 4.10.	Results for the NN 4-1 and MNN 3-3-1 models.....	46

Table 4.11.	Results obtained from the NN 2-1 and MNN 2-1-1 models.....	49
Table 4.12.	Results for the NN 2-1 and MNN 1-1-1 models.....	52
Table 4.13.	Results for the NN 3-1 and MNN 2-2-1 models.....	55
Table 4.14.	Optimization results of the NN 2-1 for precipitation pH and temperature.....	57
Table 4.15.	Optimization results of the NN 2-1 for target metal contents.....	58
Table 4.16.	Input significance analysis for the NN 2-1 model.....	59
Table 4.17.	Input significance analysis for the MNN 1-1-1 model.....	60
Table 4.18.	Input significance analysis for the MNN 3-3-1 model.....	60

LIST OF SYMBOLS / ABBREVIATIONS

$\%X$	Percent Conversion
a_k	Parameters of any nonlinear function
Cal.	Calculated
e	Vector of network errors
Exp.	Experimental
$f(x)$	Objective function
g	Gradient
g_k	Current gradient
$G(x)$	Vector function
$G_i(x)$	Equality constraints
H	Hessian matrix
J	Jacobian matrix
k	Subset number
$L(x,\lambda)$	Lagrangian function
m	Number of fitted coefficients
MR	Multiple Regression
n	Number of response values
N	Neuron
N_{hidden}	Number of hidden neurons
N_{input}	Number of input nodes
N_{output}	Number of output nodes
N_{sample}	Number of training data sets
net_j	Net input
o_j	Net output

P.T.	Precipitation temperature
r	Residual
R^2	Coefficient of determination
R^2_{adjusted}	Adjusted coefficient of determination
RMSE	Root mean square error
R.T.	Reaction Temperature
S_e	Standard Error
SSE	Sum of Squares due to Error
T.S.	Time on Stream
TSA	Total Surface Area
vol%	Volume percent
w_0	Initial vector of weights
w_{ij}	Vector of weights
wt% Ce	Cerium weight percent
wt% Co	Cobalt weight percent
wt% Cu	Copper weight percent
W/F	Weight of catalyst and flow rate ratio
x_k	Vector of current weights and biases
x_l	Lower bound
x_u	Upper bound
$X_k(x)$	Nonlinear function
y	Experimental conversion data point
\hat{y}	Calculated conversion
\bar{y}	Mean of experimental conversion values
$y(x)$	General linear model
α_k	Learning rate

β	Degree of over-determination
δ_k	Error of each unit
λ_i	Lagrange multipliers
ν	Degrees of freedom
σ	Variance
χ^2	The error sum of squares
ANN	Artificial neural network
KT	Tucker
MNN	Modular Neural Network
PEMFC	Polymer electrolyte membrane fuel cell
RSM	Response surface methodology
SOP	Sequential quadratic programming

1. INTRODUCTION

In the last decade, the limitation of petroleum reserves has become clearer for the communities and the researches on the alternative energy technologies have been increased. One of the promising energy sources for transportation vehicles and small/medium size stationary applications such as houses and small businesses is the fuel cell technology.

Although, the fuel cell technology has been subjected a considerable development since its initiation, there are some problems concerning this technology. One of them is the catalyst used in the anode of the cell which has two disadvantages (Panzera et al., 2004, İnce et al., 2005):

- i. The feed stream to the fuel has to be free from carbon monoxide. CO poisons the Pt catalyst in the anode.
- ii. The platinum is a precious metal and has a limited supply. For instance there isn't enough platinum to be used in every vehicle produced in the world.

Since the safe storage of the hydrogen is not well established, it should be produced on-site using a fuel processor for the transportation vehicles and the other small-medium size applications like houses. However, hydrogen from the fuel processor contains some amount of CO, which must be eliminated. Although noble metal catalyst (especially Pt based catalysts) are common studied for this purpose, $\text{CuO}_x\text{-CeO}_2$ catalysts also seem to be a promising choice (Park et al., 2004). Indeed 20wt% Cu – 20wt% Co – 60wt% Ce catalysts prepared by co-precipitation method was investigated for CO oxidation in H_2 rich environment in our group and found to be successful. The effects of pH and temperature which are the most important parameters for precipitation were studied using response surface method and the optimum pH and temperature values were found to be 10.0 and T 37.5 °C respectively. The effects of reaction conditions such as reaction temperature, W/F, CO, O_2 , CO_2 and H_2O concentration in feed stream were also studied in the same work. (Kibar, 2005). In a similar study the effects of the metal loadings were also studied experimentally by Özdemir (2006).

In this study the results of the experimental works by Kibar (2005) and Özdemir(2006) were modeled via both Neural Networks (NN) and Modular Neural Networks (MNN). The tangential sigmoid function was used for both models as the activation function while the Marquardt – Levenberg algorithm was employed as the learning algorithm to adapt weights. MATLAB® Neural Network Toolbox was used for computations of the NNs and MNNs. The data were also fitted by quadratic multiple regression for comparison. The goodness of the fit were measured using R^2 , R^2_{adjusted} and Root Mean Squared Error while k-fold cross validation method was used for validation.. The best neural network models were also optimized with MATLAB® Optimization Toolbox using constrained nonlinear optimization function with Quasi-Newton optimization method. Finally the the input significance analysis for the best neural network model of each subset was also performed to see the relative impacts of the design parameters

In the 2nd chapter, a literature survey of the CO oxidation, catalyst design, neural and modular neural models structure, measures of regression and optimization method are given. Chapter 3 presents the experimental data and computational methods. The results obtained from various artificial neural network models and multiple regression with measures of regression are given and discussed in chapter 4. Finally, the conclusions extracted from modeling techniques and recommendations for further studies on neural networks are listed in Chapter 5 of the thesis.

2. LITERATURE SURVEY

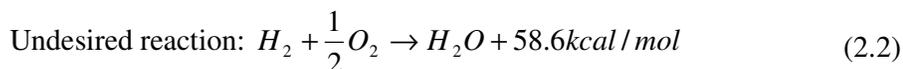
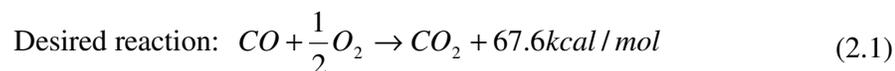
In this chapter of thesis some recent studies on CO oxidation using mixed oxide catalysts will be presented first. Then, the neural network (NN) and modular neural network (MNN) models will be explained in details in section two and three respectively. The text will continue with the modeling of CO oxidation by neural networks in section four. In section five, the method of multiple regression calculations will be presented. Calculation methods of correlation parameters namely R^2 , R^2_{adjusted} and RMSE, input significance computation method, k-fold cross validation and factor effects analysis will be covered in sixth section. Finally, the optimization methods for neural networks will be discussed.

2.1 CO Oxidation

Selective CO oxidation in hydrogen rich streams has been the main focus for many researchers in the recent years. CO is a poison for fuel cells which are considered as the promising energy conversion devices both for home and industrial use in the future. Polymer electrolyte membrane fuel cells (PEMFC), especially, seem to be an attractive choice for small-medium applications such as cars or houses. Due to nature of the reactions, CO is obtained as an impurity in the product stream in most of the H₂ production methods, which involve the use of hydrocarbon source. For instance, H₂ can be produced on-site by steam reforming or autothermal reforming of hydrocarbon fuels; such as gasoline and methanol followed by water gas shift process (Zou et al., 2006). CO concentration is expected to be about 1% in a H₂ feed stream from a fuel processor while the concentration of more than 10 ppm CO in the hydrogen stream can cause deterioration in its energy conversion efficiency via CO-induced poisoning of the anode catalyst (Pt), it should be kept far below a tolerable level. Hence the CO in the feed must be eliminated.

The concentration can be reduced to acceptable levels by catalytic methanation, Pd-based membrane purification and catalytic selective CO oxidation. Of these methods, the selective oxidation of CO with O₂ appears to be the simplest and most effective method for removing CO. The most important requirements of catalysts of selective CO oxidation are

a high oxidation rate of CO and a high selectivity with respect to the side oxidation reaction of H₂ since these two reactions will compete (Park et al., 2004):



Such catalysts should also be active in the presence of CO₂ and H₂O in the feed. (Chen et al., 2006).

2.1.1. CO Oxidation Catalysts

The catalysts proposed in the literature for the selective oxidation of CO are noble metal based, including alumina-supported Pt-group metal catalysts and metal oxide-supported Au catalysts. Gold based catalysts have been found to be markedly more active catalysts than Pt-group metal catalysts at low temperatures (<393 K), but not as resistant to deactivation by CO₂ and H₂O. None of these catalysts can prevent significant losses of hydrogen by oxidation (Avgouropoulos et al., 2002).

Purifying H₂ by selectively oxidizing trace amounts of CO over the precious metal catalysts such as Pt, Ru, Rh and Au could be achieved. Oh et al. (2000) studied the activity and selectivity of Pt, Ru, and Rh catalysts supported on alumina and found Ru and Rh to be very selective compared to Pt/alumina. Korolkikh et al. (2000) investigated the selective oxidation of CO over Pt/alumina and found that metal oxide promoters enhanced the activity of the catalysts even at low temperatures. Igarashi et al. (1997) investigated Pt supported zeolites. Their results showed that the selectivity was affected by supports and Pt/mordenite showed the highest conversion of CO to CO₂. Manasilp and Gulari (2002) showed that a 2% Pt/alumina sol-gel catalyst can clean down to a few ppms (Ren and Hong, 2007). A number of platinum group metal-based catalysts supported on alumina, zeolite or activated carbon have been also studied as potential PROX catalysts. At low oxygen concentrations, the selectivity of Pt based catalysts for CO oxidation in a H₂-rich

environment can be improved by increasing the oxygen supply to Pt sites via promoters like CeO₂ or SnO₂ (Özdemir, 2006, Ince et al 2005, and Şimşek et al., 2007).

CuO-CeO₂ mixed oxide catalysts have been also reported to be very active in the oxidation of CO with a specific activity several orders of magnitude higher than that of conventional Cu-based catalysts, and comparable to Pt-based catalysts. Mixed oxides of CuO-CeO₂ have recently been proposed as good candidates for the selective removal of CO from reformat streams: they can be used in the temperature range of 373-423 K with a selectivity of 95-90% for complete conversion; they are more active and significantly more selective than Pt-based catalysts at a lower reaction temperature; they are less active but much more selective than Au-based catalysts. The use of the mixed oxide CuO-CeO₂ for CO oxidation has recently attracted much attention (Chen et al., 2006).

2.1.2. Co Oxidation over CuO_x-CoO_x-CeO₂ Catalysts

High cost and limited supply of Pt directed the researchers for less expensive more abundant metals for CO oxidation. Cu and Co are promising candidates when they are supported with a metal which has an affinity for O₂ binding. Neither of Cu, Co or Ce is good catalysts for CO oxidation when they are compared with Al₂O₃ supported Pt. Consequently, a combination of these metals should be used for complete conversion. CuO and CeO₂ enhance CO oxidation through a synergistic effect. Cerium oxide is well known to have a high oxygen exchange capacity, which is related to the capacity of cerium to change oxidation states reversibly between Ce⁴⁺ and Ce³⁺ by receiving or giving up oxygen (Schmitz et al., 1993).

It has been observed that CoO_x addition has increased the activity of CuO_x-CeO₂ catalysts. The addition of Co metal until a certain point raises the activity then it started to decrease. When Co is precipitated only with Cu or Ce, the conversion of the catalyst is very low. CoO_x-CeO₂ has the lowest activity among the various combinations of these three metals (Özdemir, 2006). It was found that the catalyst with metal weight percents of 16.67 %Cu, 16.67 %Co and 66.66 %Ce has the highest (100%) conversion (Özdemir, 2006). It was also reported that the presence of Co in the Cu-Ce catalyst is helpful in alleviating the temporal poisoning effect of H₂O by weakening the water and catalyst

surface interaction (Park et. al., 2004). The catalyst preparation conditions (precipitation pH and temperature) were also optimized with data from face centered composite design within the pH range of 8.5 – 11.5 and temperature range of 5.0 – 70.0 °C, and the optimum pH and temperature were calculated as 10.0 and 37.5 °C respectively (Kibar, 2005).

2.2. Artificial Neural Network (ANN) Modeling

Artificial neural network (ANN) is a powerful data modeling tool. The main advantages of this technique are the modeling without any assumptions about the nature of the phenomenological mechanism underlying the process, the ability to learn linear and nonlinear relationships between variables directly from a set of examples, the capacity of modeling multiple outputs simultaneously and a reasonable application of the model to unlearned data (Basheer & Hajmeer, 2000).

Even multiple input – multiple output (MIMO) nonlinear relationships can be approximated simultaneously and easily. Owing to their several attractive characteristics, ANNs have been widely used in chemical engineering applications such as steady state and dynamic process modeling, process identification, yield maximization, nonlinear control, and fault detection and diagnosis (Nandi et al., 2004).

A neural network model consists of an input layer, hidden layer(s) and an output layer. Each layer has elements called neuron; a neuron takes its input from the previous neuron and calculates the output using the weights from a transfer function.

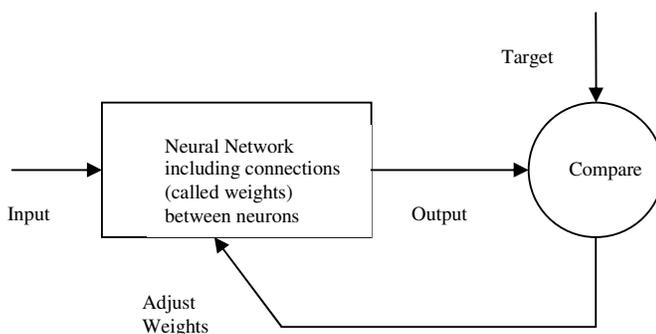


Figure 2.1: Schematic representation of a neuron

To select the number of hidden layers and the number of processing elements (neurons) in hidden layers, it is necessary to make a trial and error procedure until a good behavior of the network is obtained. However, as for the number of hidden layers concerns, it is advisable to use just one layer because use of more layers exacerbates the problem of local minima (Rai et al., 2005).

The number of hidden neurons depends on the number of training patterns, the amount of data noise, and the complexity of the function that ANN is approximating (Hecht-Nielsen, 1987).

The number of hidden neurons that can be used without the risk of over fitting can be estimated according to the equation introduced by Carpenter and Hoffman (1995).

$$N_{hidden} = \frac{N_{sample} / \beta - N_{output}}{N_{input} + N_{output} - 1} \quad (2.3)$$

Where N_{hidden} is the number of hidden nodes, N_{sample} the number of training data sets, N_{input} the number of input nodes and N_{output} is the number of output nodes. The constant β determines the degree of over-determination. It has three values: $\beta < 1$, under-determined; $\beta = 1$, determined and $\beta > 1$, over-determined. In brief, $\beta \geq 1$ is preferred.

2.2.1. Backpropagation Algorithm

Training algorithms for feedforward networks use the gradient of the performance function to determine how to adjust the weights to minimize performance. The gradient is determined using a technique called backpropagation, which involves performing computations backward through the network. The backpropagation computation is derived using the chain rule of calculus. The simplest implementation of backpropagation learning updates the network weights and biases in the direction in which the performance function decreases most rapidly which is the negative of the gradient. An iteration of this algorithm can be written as in equation 2.4 (MATLAB Neural Network Toolbox Help),

$$x_{k+1} = x_k - \alpha_k g_k \quad (2.4)$$

Where x_k , is the vector of weights and biases, g_k is current gradient and α_k is the learning rate.

There are two different ways in which this gradient descent algorithm can be implemented: incremental mode and batch mode. In incremental mode, the gradient is computed and the weights are updated after each input is applied to the network. In batch mode on the other hand, all the inputs are applied to the network before the weights are updated.

The learning rate is multiplied by the negative of the gradient to determine the changes to the weights and biases. The larger the learning rate, the bigger the step is. If the learning rate is made too large, the algorithm becomes unstable. If the learning rate is set too small, the algorithm takes a long time to converge.

The training stops if the number of iterations exceeds epochs, if the performance function drops below the goal, or if the magnitude of the gradient is less than the minimum gradient set before training.

The networks are also sensitive to the number of neurons in their hidden layers. Too few neurons can lead to under-fitting, while too many neurons can contribute to over-fitting, in which all training points are well fitted, but the fitting curve oscillates wildly between these points. The backpropagation algorithm is given Figure 2.2;

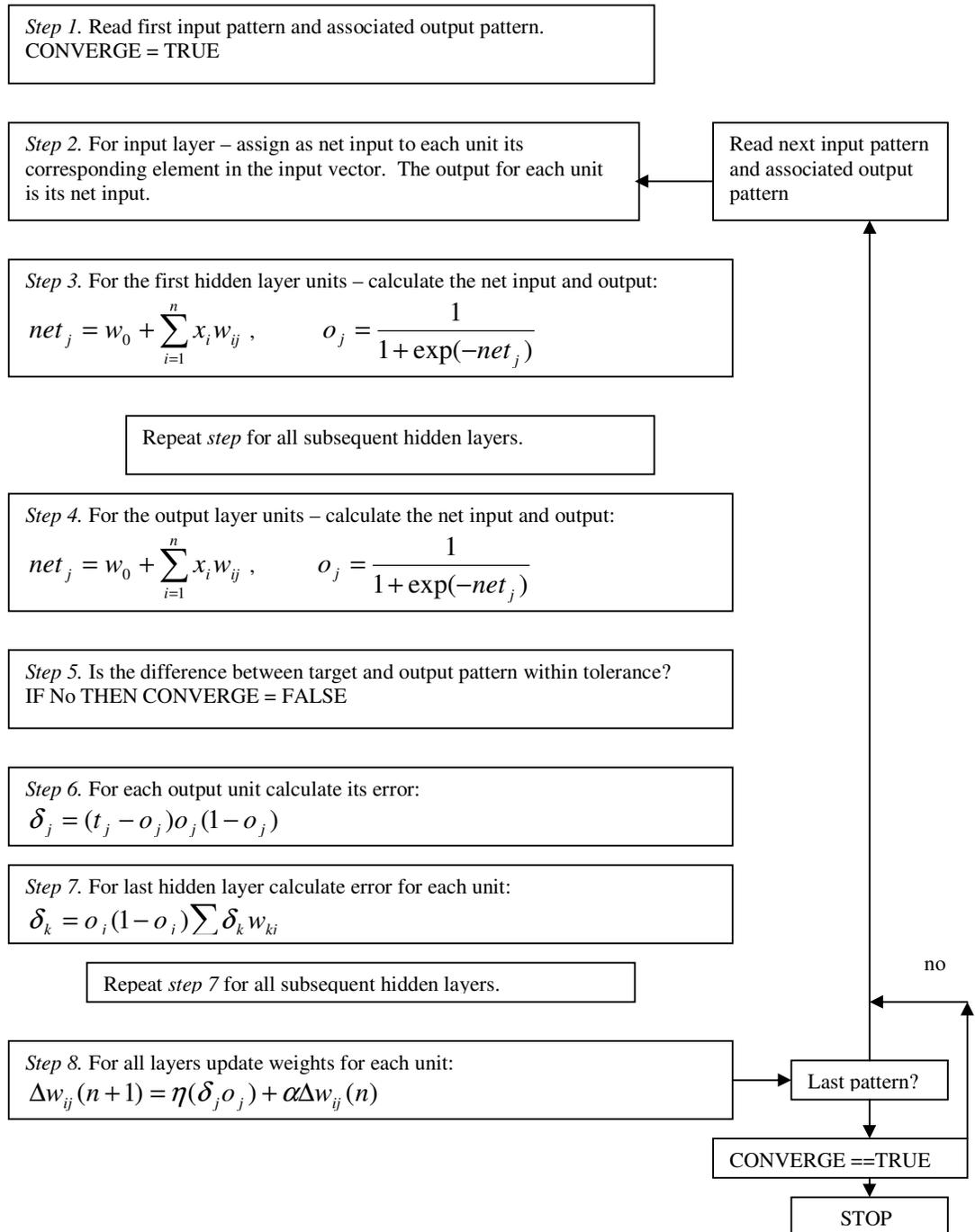


Figure 2.2. Backpropagation Algorithm (Callan, 1999)

The input and target data are separated into groups of vectors for training, validating and testing the network. Validation set is used to stop training early if it

attempts to over-fit the training data, and testing set is used as an independent measure of how the network might be expected to perform on data that it was not trained on.

Typically one epoch of training is defined as a single presentation of all input vectors to the network. The network is then updated according to the results of all those presentations.

2.2.2. Levenberg-Marquardt Algorithm

The Levenberg-Marquardt algorithm was designed to approach second-order training speed without having to compute the Hessian matrix. When the performance function has the form of a sum of squares (as is typical in training feedforward networks), then the Hessian matrix can be approximated as

$$H = J^T J \quad (2.5)$$

and the gradient can be computed as

$$g = J^T e \quad (2.6)$$

Where J is the Jacobian matrix that contains first derivatives of the network errors with respect to the weights and biases, and e is a vector of network errors. The Jacobian matrix can be computed through a standard backpropagation technique that is much less complex than computing the Hessian matrix.

The Levenberg-Marquardt algorithm uses this approximation to the Hessian matrix in the following Newton-like update:

$$x_{k+1} = x_k - [J^T J + \mu I]^{-1} J^T e \quad (2.7)$$

When the scalar μ is zero, this is just Newton's method; the approximate Hessian matrix is used for updating weights and biases. When μ is large, this becomes gradient descent with a small step size. Newton's method is faster and more accurate near an error minimum, so the aim is to shift toward Newton's method as quickly as possible. Thus, μ is

decreased after each successful step (reduction in performance function) and is increased only when a tentative step would increase the performance function. In this way, the performance function is reduced at each iteration of the algorithm (MATLAB, Neural Network Toolbox Help, 2006).

2.3. Modular Neural Network (MNN) Modeling

The term “Modular neural networks” is very fuzzy. It is used in a lot of ways with different structures. Everything that is not monolithic is said to be modular (Melin et al., 2007). A monolithic neural network is a standard artificial neural network in which all inputs is connected to the every neuron in the hidden layer. In a neural network the connections between inputs and neurons in the hidden layer can be established as wished, these special neural network architectures are named as custom neural networks (MATLAB, Neural Network Toolbox Help, 2006). Modular neural networks are artificial neural networks as well.

On the other hand, a neural network is said to be modular if the computation performed by the network can be decomposed into two or more modules (subsystems) that operate on distinct inputs without communicating with each other. The outputs of the modules are mediated by an integrating unit that is not permitted to feed information back to the modules. In particular, the integrating unit both decides (1) how the outputs of the modules should be combined to form the final output of the system, and (2) which modules should learn which training patterns (Haykin, 1994).

2.4. Modeling of CO Oxidation using ANN

Traditionally, the processing and understanding of the experimental outputs (characterization and catalytic performances) was accomplished by the researchers, who applied previous experiences or fundamental knowledge in order to carry out the experimental design and to establish relationships between the different experimental results. In the case of the large number of variables (such as in the case of computational catalysis) in play and the application of complex optimization algorithms for the experimental design make the direct human interpretation of data derived from high

throughput experimentation difficult. Hence recently, data mining techniques have been applied in order to find relationships and patterns between the input and output data derived from accelerated experimentation. For instance, the artificial intelligence (AI) techniques have an important potential for modeling and prediction of complex high-dimensional data. Among these techniques, artificial neural networks (NN) could be useful in the chemical field (Serra et al., 2003).

Omata et al. investigated activity and selectivity of cobalt supported on alkali metal carbonate for PROX of 1 vol% CO using the stoichiometric amount of O₂ in excess hydrogen. Co/SrCO₃ was discovered, and optimized to achieve high performance by using a full factorial design of experiment, an artificial neural network and a grid search (Omata et al., 2005). Günay and Yildirim, (2007) modeled design parameters of Pt-Co-Ce/Al₂O₃ catalyst via artificial neural networks for low temperature CO oxidation.

2.5. Multiple Regression

A general linear model is one expressed as

$$y(x) = \sum_{k=1}^M a_k X_k(x) \quad (2.8)$$

Where the parameters are $\{a_k\}$, and the expression is linear with respect to them, and $X_k(x)$ can be any (nonlinear) functions of x , not depending on the parameters $\{a_k\}$. Then equation 2.9 is obtained,

$$\sum_{i=1}^N \frac{1}{\sigma_1^2} \left[y_i - \sum_{j=1}^M a_j X_j(x_i) \right] X_k(x_i) = 0, k = 1, \dots, M \quad (2.9)$$

This is rewritten as equation 2.10,

$$\sum_{j=1}^M \left[\sum_{i=1}^N \frac{1}{\sigma_1^2} X_j(x_i) X_k(x_i) \right] a_j = \sum_{i=1}^N \frac{y_i}{\sigma_1^2} X_k(x_i) \quad (2.10)$$

Or as

$$\sum_{j=1}^M \alpha_{kj} a_j = \beta_k \quad (2.11)$$

Solving this set of equations gives the regression parameters $\{a_j\}$.

Various global and piecewise polynomials can be used to fit the data. Most approximations are to be used with $M < N$. One can sometimes use more and more terms and calculating the value of χ^2 (error sum of squares) for each solution. Then stop increasing M when the value of χ^2 no longer increases with increasing M (Perry and Green, 1997).

2.5.1. Quadratic Response Surface Models

Response surface methodology (RSM) is a tool for understanding the quantitative relationship between multiple input variables and one output variable. Consider one output, z , as a polynomial function of two inputs, x and y . The function $z = f(x,y)$ describes a two-dimensional surface in the space (x,y,z) . In general, you can have as many input variables as you want and the resulting surface becomes a hyper surface. Also, you can have multiple output variables with a separate hyper surface for each one.

For three inputs (x_1, x_2, x_3) , the equation of a quadratic response surface is (MATLAB Statistics Toolbox Help)

$$\begin{aligned} y = & b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + \dots \quad (\text{Linear Terms}) \\ & + b_{12} x_1 x_2 + b_{13} x_1 x_3 + b_{23} x_2 x_3 + \dots \quad ((\text{Interaction Terms}) \\ & b_{11} x_1^2 + b_{22} x_2^2 + b_{33} x_3^2 \quad (\text{Quadratic Terms}) \end{aligned} \quad (2.15)$$

2.6. Measures of Regression

After fitting a data, goodness of fit statistics should be evaluated. A good fit might be a model; (i) in which the model coefficients can be estimated with little uncertainty, (ii) which explains a high proportion of the variability in the data, and (iii) which is able to predict new observations with high certainty.

Generally speaking, graphical measures are more beneficial than numerical measures because they allow viewing the entire data set at once, and they can easily display a wide range of relationships between the model and the data. The numerical measures are more narrowly focused on a particular aspect of the data and often try to compress that information into a single number. In practice, depending on the data and analysis requirements, one might need to use both types to determine the best fit (MATLAB® Curve Fitting Toolbox Help).

The goodness of fit statistics that may be used to evaluate the model are; (i) Residual analysis, (ii) Sum of squares due to error, SSE (iii) Root mean square error, (iv) R-Square Statistic, R^2 and (v) Adjusted R-Square, R^2_{adj} .

2.6.1. Residual Analysis

Residuals are defined as the difference between the observed values of the dependent variable and the values that are predicted by the model. The residuals approximate independent random errors of a fit that is appropriate for the data.

Mathematically, the residual for a specific predictor value is the difference between the response value y and the predicted response value \hat{y} .

$$r = y - \hat{y} \tag{2.16}$$

Assuming the model you fit to the data is correct, the residuals approximate the random errors. Therefore, if the residuals appear to behave randomly, it suggests that the

model fit the data well. However, if the residuals display a systematic pattern, it is a clear sign that the model fits the data poorly (MATLAB Curve Fitting Toolbox Help).

2.6.2. Sum of Squares Due to Error (SSE)

This statistic measures the total deviation of the response values from the fit to the response values. It is also called the summed square of residuals and is usually labeled as SSE where y_i is experimental and \hat{y}_i is calculated conversion value.

$$SSE = \sum_{i=1} w_i (y_i - \hat{y}_i)^2 \quad (2.17)$$

A value closer to 0 indicates that the model has a smaller random error component, and that the fit will be more useful for prediction.

2.6.3. Root Mean Square Error (RMSE)

This statistic is also known as the fit standard error and the standard error of the regression. It is an estimate of the standard deviation of the random component in the data, and is defined as

$$RMSE = s = \sqrt{MSE} \quad (2.18)$$

Where MSE is the mean square error or the residual mean square and ν is degrees of freedom.

$$MSE = \frac{SSE}{\nu} \quad (2.19)$$

Just as with SSE, an MSE value closer to 0 indicates a fit that is more useful for prediction (MATLAB, Curve Fitting Toolbox Help).

2.6.4. R-Square

This statistic measures how successful the fit is in explaining the variation of the data. Put another way, R-square is the square of the correlation between the response

values and the predictions. It is also called the square of the multiple correlation coefficient and the coefficient of multiple determination.

R-square is defined as the ratio of the sum of squares of the regression (SSR) and the total sum of squares (SST). SSR is defined as

$$SSR = \sum_{i=1}^n w_i (\hat{y}_i - \bar{y})^2 \quad (2.20)$$

SST is also called the sum of squares about the mean, and is defined as

$$SST = \sum_{i=1}^n w_i (y_i - \bar{y})^2 \quad (2.21)$$

Where $SST = SSR + SSE$. Given these definitions, R-square is expressed as

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \quad (2.22)$$

R-square can take on any value between 0 and 1, with a value closer to 1 indicating that a greater proportion of variance is accounted for by the model. For example, an R-square value of 0.8234 means that the fit explains 82.34% of the total variation in the data about the average.

If the number of fitted coefficients is increased in the model, R-square will increase although the fit may not improve in a practical sense. To avoid this situation, degrees of freedom adjusted R-square statistic should be used.

It is possible to get a negative R-square for equations that do not contain a constant term. Because R-square is defined as the proportion of variance explained by the fit, if the fit is actually worse than just fitting a horizontal line then R-square is negative. In this case, R-square cannot be interpreted as the square of a correlation (MATLAB, Curve Fitting Toolbox).

2.6.5. Degrees of Freedom Adjusted R-Square

This statistic uses the R-square statistic defined in section 2.6.2, and adjusts it based on the residual degrees of freedom. The residual degrees of freedom is defined as the number of response values n minus the number of fitted coefficients m estimated from the response values.

$$v = n - m \quad (2.23)$$

v indicates the number of independent pieces of information involving the n data points that are required to calculate the sum of squares. If parameters are bounded and one or more of the estimates are at their bounds, then those estimates are regarded as fixed. The degrees of freedom are increased by the number of such parameters.

The adjusted R-square statistic is generally the best indicator of the fit quality when you compare two models that are nested — that is, a series of models each of which adds additional coefficients to the previous model.

$$R_{adj}^2 = 1 - \frac{SSE(n-1)}{SST(v)} \quad (2.24)$$

The adjusted R-square statistic can take on any value less than or equal to 1, with a value closer to 1 indicating a better fit. Negative values can occur when the model contains terms that do not help to predict the response (MATLAB Curve Fitting Toolbox Help).

2.6.6. k-Fold Cross Validation

Cross validation is a model evaluation method that is better than residuals. The problem with residual evaluation is that they do not give an indication of how well the learner will do when it is asked to make new predictions for data it has not already seen. One way to overcome this problem is to not use the entire data set when training a learner. Some of the data is removed before training begins. Then when training is done, the data that was removed can be used to test the performance of the learned model on “new” data.

This is the basic idea for a whole class of model evaluation methods called cross validation (Schneider, 1997).

The holdout method is the simplest kind of cross validation. The data set is separated into two sets, called the training set and the testing set. The function approximator fits a function using the training set only. Then the function approximator is asked to predict the output values for the data in the testing set (it has never seen these output values before). The errors it makes are accumulated as before to give the mean absolute test set error, which is used to evaluate the model. The advantage of this method is that it is usually preferable to the residual method and takes no longer to compute. However, its evaluation can have a high variance. The evaluation may depend heavily on which data points end up in the training set and which end up in the test set, and thus the evaluation may be significantly different depending on how the division is made (Schneider, 1997).

k -fold cross validation is one way to improve over the holdout method. The data set is divided into k subsets, and the holdout method is repeated k times. Each time, one of the k subsets is used as the test set and the other $k-1$ subsets are put together to form a training set. Then the average error across all k trials is computed. The advantage of this method is that it matters less how the data gets divided. Every data point gets to be in a test set exactly once, and gets to be in a training set $k-1$ times. The variance of the resulting estimate is reduced as k is increased. The disadvantage of this method is that the training algorithm has to be rerun from scratch k times, which means it takes k times as much computation to make an evaluation. A variant of this method is to randomly divide the data into a test and training set k different times. The advantage of doing this is that size of the each test set and could be chosen independently. By doing so the number of the trials is averaged over (Schneider, 1997).

2.7. Optimization of Neural Networks

A neural network can be thought as a function whose variables are present in the input layer. A network with one output can be optimized by any numerical optimization method if the network is fed as a function to the particular multiple optimization method (Günay, 2005).

Optimization techniques are used to find a set of design parameters that can in some way be defined as optimal. In a simple case this might be the minimization or maximization of some system characteristics that is dependent on x . In a more advanced formulation the objective function, $f(x)$, to be minimized or maximized, might be subject to constraints in the form of equality constraints, $G_i(x) = 0$ ($i = 1, \dots, m_e$); inequality constraints, $G_i \leq 0$ ($i = m_e + 1, \dots, m$); and/or proper bounds, x_b, x_u .

$$\begin{aligned}
 & \text{minimize } f(x) \\
 & \quad \mathbf{x} \\
 & \text{subject to} \\
 & G_i(x) = 0, \quad (i = 1, \dots, m_e) \\
 & G_i \leq 0, \quad (i = m_e + 1, \dots, m)
 \end{aligned} \tag{2.25}$$

where x is the vector of length n design parameters, $f(x)$ is the objective function, which returns a scalar value, and the vector function $G(x)$ returns a vector of length m containing the values of the equality and inequality constraints evaluated at x (MATLAB Optimization Toolbox Help).

2.7.1. Constrained Optimization

In constrained optimization, the general aim is to transform the problem into an easier sub problem that can then be solved and used as the basis of an iterative process. A characteristic of a large class of early methods is the translation of the constrained problem to a basic unconstrained problem by using a penalty function for constraints that are near or beyond the constraint boundary. In this way the constrained problem is solved using a sequence of parameterized unconstrained optimizations, which in the limit (of the sequence) converge to the constrained problem. These methods are now considered relatively inefficient and have been replaced by methods that have focused on the solution of the Kuhn-Tucker (KT) equations. The KT equations are necessary conditions for optimality for a constrained optimization problem. If the problem is a so-called convex programming problem, that is, $f(x)$ and $G_i(x)$, $i = 1, \dots, m$, are convex functions, then the KT equations are both necessary and sufficient for a global solution point.

Referring to equation 2.25, the Kuhn Tucker equations can be stated as

$$\begin{aligned} \nabla f(x^*) + \sum_{i=1}^m \lambda_i \cdot \nabla G_i(x^*) &= 0 \\ \lambda_i^* \cdot G_i(x^*) &= 0 \quad i = 1, \dots, m_e \\ \lambda_i^* &\geq 0 \quad i = m_e + 1, \dots, m \end{aligned} \tag{2.26}$$

in addition to the original constraints in equation 2.25.

The first equation describes a canceling of the gradients between the objective function and the active constraints at the solution point. For the gradients to be canceled, Lagrange multipliers ($\lambda_i, i = 1, \dots, m$) are necessary to balance the deviations in magnitude of the objective function and constraint gradients. Because only active constraints are included in this canceling operation, constraints that are not active must not be included in this operation and so are given Lagrange multipliers equal to zero. This is stated implicitly in the last two equations of 2.26.

The solution of the KT equations forms the basis to many nonlinear programming algorithms. These algorithms attempt to compute the Lagrange multipliers directly. Constrained quasi-Newton methods guarantee superlinear (faster convergence than linear) convergence by accumulating second order information regarding the KT equations using a quasi-Newton updating procedure. These methods are commonly referred to as Sequential Quadratic Programming (SQP) methods, since a QP sub problem is solved at each major iteration (MATLAB Optimization Toolbox Help).

2.7.2. Sequential Quadratic Programming (SQP)

SQP methods represent the state of the art in nonlinear programming methods. Given the problem description in equation 2.25 the principal idea is the formulation of a QP sub problem based on a quadratic approximation of the Lagrangian function.

$$L(x, \lambda) = f(x) + \sum_{i=1}^m \lambda_i \cdot g_i(x) \tag{2.27}$$

Here equation 2.25 is simplified by assuming that bound constraints have been expressed as inequality constraints. QP sub problem is obtained by linearizing the nonlinear constraints.

3. COMPUTATIONAL DETAILS

Preparation and reaction conditions of $\text{CuO}_x\text{-CoO}_x\text{-CeO}_2$ for the selective CO oxidation catalyst were modeled via neural and modular neural networks. The multiple regression models were also constructed for comparison. The results were tested statistically using residual analysis and k-fold cross validation. Precipitation pH and temperature and Cu, Co, and Ce metal percents were optimized using constrained optimization method.

3.1. Experimental Data

The experimental data obtained by Kibar (2005) and Özdemir (2006) were used as the training, test and validation data. The catalyst were prepared using co-precipitation technique and tested in a micro flow reaction system in both cases. The data are presented and explained briefly in the following sections, and the details of the experimental work can be found elsewhere (Kibar, 2005, Özdemir, 2006).

3.1.1. Effects of Preparation Conditions (Precipitation pH and Temperature)

Kibar (2005) was studied the effect of precipitation pH and temperature using response surface method. He changed the pH from 8.5 to 11.5 and temperature from 5 to 70 °C, while keeping all other conditions constant with target metal contents of 20 wt % Cu, 20 wt % Co and 60 wt % of Ce. Then he tested the optimum catalyst in a micro flow reactor at 120 °C and with W/F ratio of 1.0 mg min/ml at the feed composition of 1% CO, 1% O₂, 60% H₂, 10% H₂O and 25% CO₂, the reactor outlet is sampled after 3 hours.

Kibar (2005) also develop a model and optimized it to have the optimum pH of 10.0 and temperature of 37.5 leading a conversion of 34.1%. The results are summarized in Table 3.1.

Table 3.1. Experimental total surface area, metal percents, conversion and selectivity depending on the precipitation pH and temperature data (Kibar, 2005)

pH	Temperature (°C)	TSA (m ² /g)	Co (%)	Cu (%)	X (%)	S (%)
8.5	5.0	2.0	28.4	15.5	28.2	31.9
8.5	37.5	37.5	28.6	15.7	32.4	40.1
8.5	70.0	12.6	32.3	16.2	25.2	36.5
10.0	5.0	13.3	29.9	17.0	27.7	35.0
10.0	37.5	36.6	25.7	13.5	32.9	41.9
10.0	37.5	34.6	30.2	15.8	32.9	38.4
10.0	37.5	34.6	32.1	16.0	36.6	39.4
10.0	70.0	24.1	27.2	14.8	28.9	39.7
11.5	5.0	11.5	28.0	15.3	27.1	41.4
11.5	37.5	25.7	18.6	12.8	31.6	37.1
11.5	70.0	25.7	20.6	12.7	29.1	34.4

3.1.2. Effects of Reaction Conditions (Reaction Temperature, W/F Ratio, Time on Stream)

Kibar (2005) also studied the effects of reaction conditions such as reaction temperature, and W/F ratio as given in Table 3.2. The time on stream could be considered as another reaction parameter since the measurements were performed three different times on stream as seen from the same table.

Table 3.2: Experimental conversion and selectivity values for the respective temperature and weight of catalyst over flow rate values (Kibar, 2005).

Exp#	Temperature (°C)	W/F (mg.min/ml)	1 Hour		2 Hours		3 Hours	
			X (%)	S (%)	X (%)	S (%)	X (%)	S (%)
1	120	1	27.1	22.1	36.6	27.3	41.1	29.4
2	140	1	46.3	67.6	53.1	64.3	57.0	45.6
3	160	1	78.2	54.8	75.7	58.1	70.3	60.6
4	140	2.5	71.4	55.7	76.7	54.9	74.3	55.8
5	160	2.5	100.0	38.5	100.0	41.8	100.0	43.4
6	160	1.75	100.0	50.3	98.5	48.6	100.0	47.5
7	150	2.5	99.4	52.0	100.0	60.4	100.0	56.7

3.1.3. Metal Percent Optimization of CuO_x-CoO_x-CeO₂ Catalyst

Özdemir (2005) studied the effects of metal loading on CO conversion using mixture experiment. She changed the target Co, Ce and Cu ratios and measured the total surface area, actual metal loadings, conversion and selectivity (Table 3.3). The reaction temperature was set at 150°C and W/F ratio at 1 mg min/ml. The optimum metal loadings are found as 16.67% Cu, 16.67% Co and 66.66% Ce. The results of this study are given in Table 3.3.

Table 3.3: Experimental Total Surface Area, Conversion and Selectivity for the corresponding target and measured metal percents (Özdemir 2006)

Exp#	Target wt% Cu	Measured wt% Cu	Target wt% Co	Measured wt% Co	Target wt% Ce	Measured wt% Ce	%X	%S	TSA (m ² /g)
1	0.00	0.00	0.00	0.00	100.00	100.00	1.90	0.00	35.80
2	16.67	11.07	16.67	8.05	66.67	80.88	100.00	31.97	29.18
3	33.33	23.33	33.33	28.81	33.33	47.86	87.62	33.81	23.42
4	16.67	10.62	66.67	34.78	16.67	50.60	84.97	83.86	15.08
5	50.00	33.62	0.00	0.00	50.00	66.38	71.67	24.05	24.54
6	50.00	36.18	50.00	26.11	0.00	0.00	2.73	33.72	3.71
7	0.00	0.00	50.00	28.83	50.00	71.17	2.05	0.00	23.43
8	0.00	0.00	0.00	0.00	100.00	100.00	0.70	0.00	34.69
9	0.00	0.00	100.00	100.00	0.00	0.00	6.49	0.00	4.92
10	0.00	0.00	50.00	27.78	50.00	72.22	10.75	43.77	32.32
11	100.00	100.00	0.00	0.00	0.00	0.00	0.23	0.00	2.10
12	0.00	0.00	100.00	100.00	0.00	0.00	8.41	0.00	5.61
13	66.67	37.55	16.67	10.68	16.67	51.77	86.87	28.07	19.13
14	100.00	100.00	0.00	0.00	0.00	0.00	4.45	0.00	1.40

3.2. Computational Approach

Neural network and modular neural network models were constructed for modeling the experimental data showing the effects of the preparation and reaction conditions of CuO_x-CoO_x-CeO₂ catalysts on CO oxidation. MATLAB® Neural Network Toolbox was used for computations. Tangential sigmoid function which gives better results than

logarithmic sigmoid function was used as the activation function. The Widrow-Hoff learning rule was employed as the error correcting rule of Levenberg - Marquardt modification of backpropagation algorithm which was used to adopt weights and biases of the network. The data was divided into three sets before computations as training, validation and testing set; validation set was used for validating network performance during training so that training stops early if it attempts to over-fit the training data, and test data were used for an independent measure of how the network might be expected to perform on data it was not trained on.

The performances of the networks were measured statistically with coefficient of determination (R^2), adjusted coefficient of determination (R^2_{adj}), and root mean square error (RMSE) correlation factors. First, neural and modular neural networks were constructed using as small as possible number of neurons in the hidden layers then the networks were enlarged. The time on stream was also included in the modeling of reaction conditions, (reaction temperature and W/F ratio) to increase the data by 3 fold. For comparison, the experimental data were also modeled via quadratic multiple regression using Excel Solver (MS Excel Xp, 2002).

The model validation was done by constructing curves of experimental results and calculated data, the distribution of data with respect to $y = x$ line was observed. Also, the errors of the models were displayed on the graphs of which y axis was zero error line. The errors shouldn't be much scattered in order to confirm the suitability of the model. k-fold cross validation was used to validate the data, the data set was divided in k set, and each time k-1 set were used to calculate conversion, a balanced distribution of root mean square error was looked for to assign a model good.

The input importance calculations were also done for the best models of each data set. Before each run, one of the input columns was excluded from the data. And, its effect on the root mean square error of the network was observed in order to determine which input parameter has greater importance on conversion for the models tested (Alyuda Help, 2005). Factor effects of the input parameters were also determined by constructing curves each time changing only one input parameter and having constant others.

Neural networks were optimized using MATLAB® Optimization Toolbox with constraint optimization algorithm. The optimum values of precipitation pH and temperature, and Cu, Co, and Ce metal ratios were calculated from the best models.

4. RESULTS AND DISCUSSION

4.1. Neural and Modular Neural Network Modeling of CO Conversion

The CO conversion on $\text{CuO}_x\text{-CoO}_x\text{-CeO}_2$ was modeled with neural networks and modular neural networks. At first, the artificial neural network model was applied starting with a small model and then enlarging the model without the number of points exceeding the number of connections between neurons. The data were also modeled via quadratic multiple regression for comparison.

The neural network modeling was performed for each data set of preparation conditions, metal loading and reaction conditions as well as the combination of them to increase the number of data points. The time on stream (sampling time) was also used for the same purpose. In addition, the effects of the physical properties of the catalyst (weight percents and total surface area) as an intermediate output, on CO conversion were modeled to see the relation between catalyst properties and the activity. Various neural network structures for various combinations of data set are presented in the following sections.

4.1.1. Modeling the Effects of Preparation Conditions on CO Conversion

The preparation conditions of the catalyst, precipitation pH and temperature were modeled first with a neural network of 2 neurons in a hidden layer. The schematic representation of the network is given in Figure 4.1. Since there are only 11 experimental data points and the input parameters can be considered as the same type, the modular neural network was not constructed for this section. For the same reason, only a neural network with one hidden layer and 2 neurons was used with a considerable success.

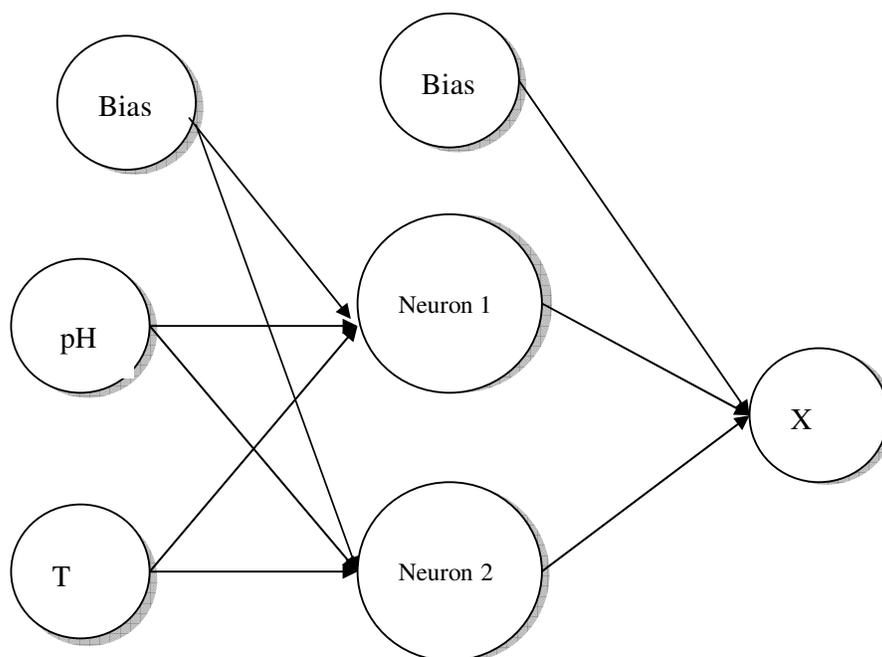


Figure 4.1. Schematic representation of the NN 2-1 model

The experimental and calculated CO conversions by neural network are presented in Table 4.1 with multiple linear regression results. The measures of regression R^2 , R^2_{adj} , and RMSE are also given in the same table.

Table 4.1: Results for the NN 2-1 and quadratic multiple regression models

%X Exp.	Neural Network Results			Multiple Regression Results		
	X Cal.	Error	Error %	%X Cal.	Error	Error %
28.2	28.2	0.00	0.00	29.83	1.63	5.78
32.4	32.6	0.16	0.49	28.61	-3.79	11.69
25.2	25.5	0.28	1.13	27.39	2.19	8.71
27.7	27.7	0.00	0.00	31.75	4.05	14.63
32.9	32.8	-0.12	0.36	31.79	-1.11	3.38
32.9	32.8	-0.12	0.36	31.79	-1.11	3.38
36.6	32.8	-3.82	10.43	31.79	-4.81	13.15
28.9	28.0	-0.86	2.98	31.82	2.92	10.10
27.1	27.7	0.61	2.26	27.99	0.89	3.29
31.6	31.6	0.00	0.00	29.28	-2.32	7.35
29.1	29.8	0.66	2.26	30.56	1.46	5.02
R^2		0.851		R^2		0.573
$R^2_{adjusted}$		0.851		$R^2_{adjusted}$		0.289
RMSE		2.01		RMSE		1.17

As it can be seen from the table, the neural network model represents the experimental data better than multiple regression model considering its R^2 value of 0.851 compare to 0.573 of regression. When the 7th data point which seems to be an experimental error is excluded from the model an R^2 of 0.970 is obtained from the neural network model.

The experimental versus calculated conversion values are given in Figure 4.2. The neural network results are dominantly closer to $y = x$ curve than multiple regression indicating a better fit.

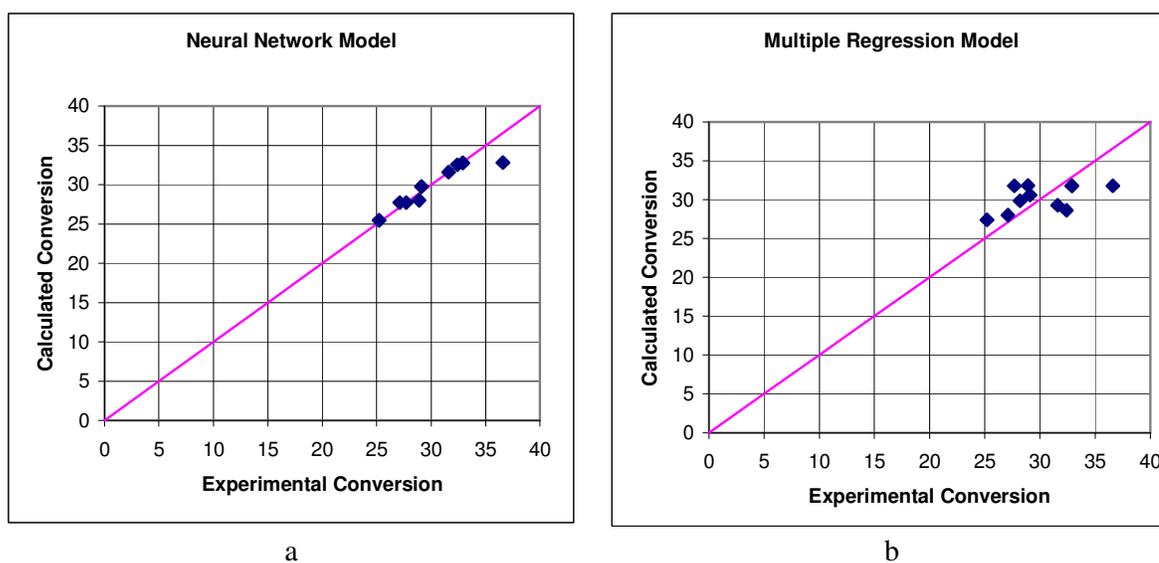


Figure 4.2. Comparison of (a) NN 2-1 and (b) quadratic multiple regression results with experimental data

When the residuals of the models are plotted, the neural network also shows less deviation from zero error line compared to the multiple regression as can be seen in Figure 4.3.

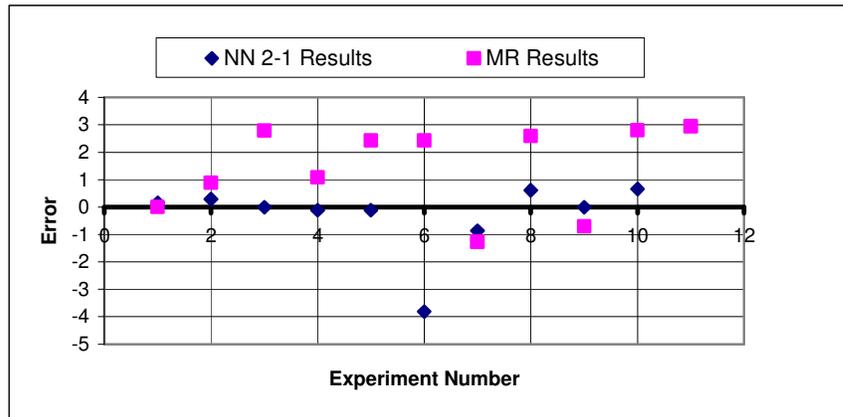


Figure 4.3. Residual analysis for NN 2-1 and quadratic multiple regression models

Since there are 11 data points k-fold cross validation analysis is done by extracting only one data point each time. As can be seen from Table 4.2, RMSE's are usually small and close to each other as a further indicator of the fitness of the model.

Table 4.2: k-fold cross validation analysis for the NN 2-1 model

Subset Number	Experiment Excluded	RMSE
1	1	2.10
2	2	1.99
3	3	2.84
4	4	2.21
5	5	2.15
6	6	2.03
7	7	2.53
8	8	2.06
9	9	2.07
10	10	2.06
11	11	2.46
	Average	2.23

4.1.2. Modeling the Effects of Preparation and Reaction Conditions Together on CO Conversion

The data related to the precipitation pH and temperature were then combined with the data for the effects of the reaction conditions (reaction temperature, W/F ratio and time on stream) and modeled via neural networks as well as modular neural networks and multiple regression. The data were separated into two inputs groups during modular neural network modeling. In the first group catalyst preparation conditions (pH and T) were collected, while the second group contained reaction parameters. Since various modular neural network architectures, with the same or similar number of connections, are possible, only the ones that have reasonable good fit will be presented in details while only the measures of regression will be given for the others. The regression coefficients of ten different network architectures (2 ANN and 8 MNN) are given in Table 4.3 to compare their fitness.

Table 4.3: Comparison of regression coefficients of different network architectures for preparation and reaction conditions modeling

Input Structure	Network Model	R²	R²_{adjusted}	S_e	RMSE
5	NN 4-1	0.982	0.978	3.96	21.7
5	NN 3-1	0.981	0.974	4.15	7.2
5	NN 2-1	0.952	0.917	6.53	9.6
2-3	MNN 2-2-2-1	0.974	0.973	4.88	18.9
2-3	MNN 2-2-1-1	0.949	0.894	6.88	9.1
2-3	MNN 1-1-1	0.971	0.909	5.19	6.1
2-3	MNN 1-2-1	0.839	0.667	12.18	16.2
2-3	MNN 2-1-1	0.971	0.935	5.20	6.7
2-3	MNN 2-2-1	0.958	0.931	6.22	9.4
2-3	MNN 2-3-2-1	0.988	0.988	3.28	18.0
2-3	MNN 2-3-1	0.989	0.988	3.14	7.7
5	MR	0.984	0.976	3.84	6.3

The results of NN 3-1 (3 neuron in one hidden layer) and MNN 2-3-1 (two neuron for the preparation conditions and three neuron for the reaction conditions) were analyzed in details as the sample structures since they represent the experimental data well even though they are quite simple. Although, the modular network (MNN 2-3-1) has two more neurons than the NN (NN 3-1), it has only two more connections due to its structure. The schematic representation of the neural networks are given in Figure 4.4.

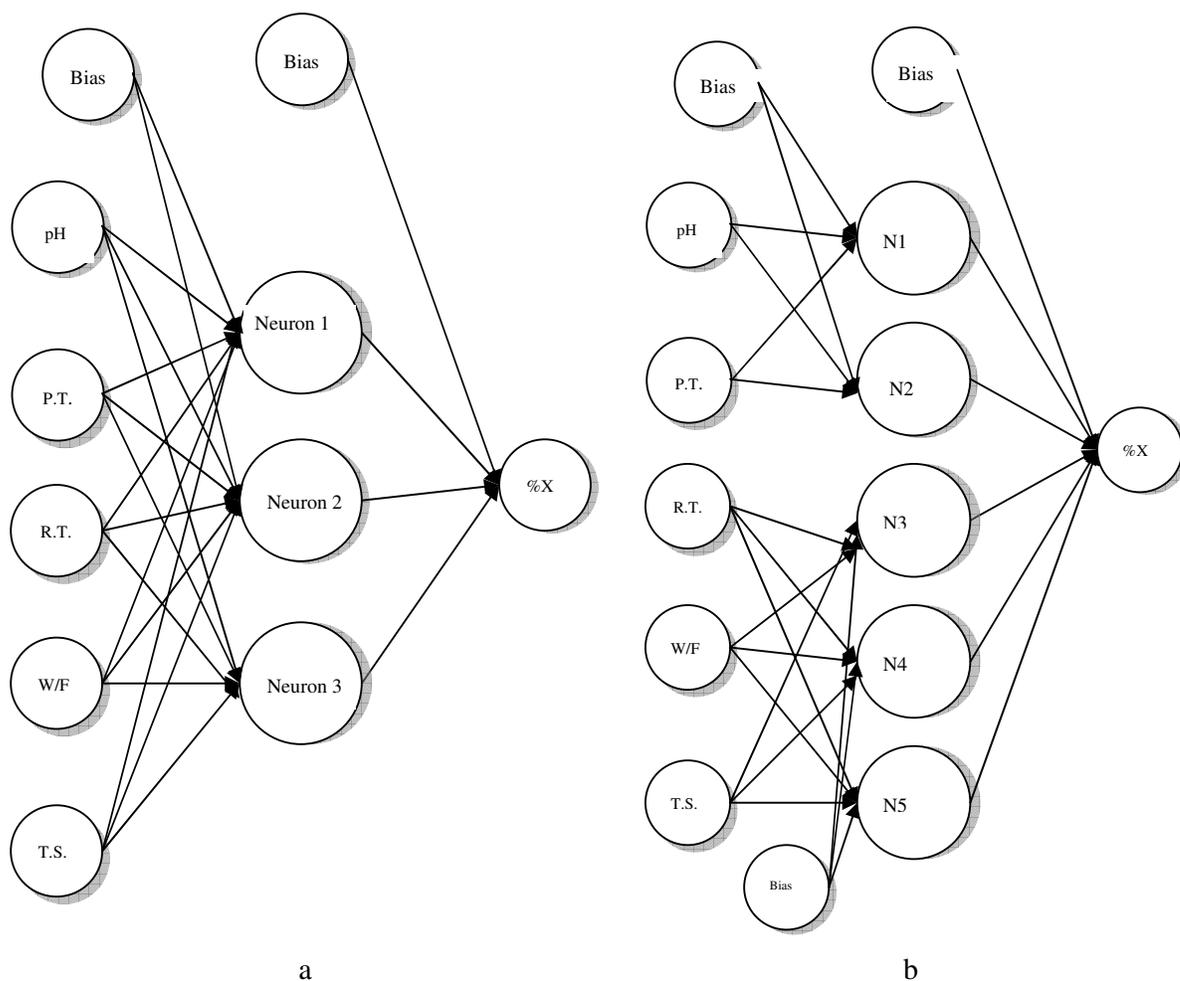


Figure 4.4: Schematic representation of the (a) NN 3-1 and (b) MNN 2-3-1 models

The results obtained from these two models are listed in Table 4.4 indicating that the modular network has slightly better R^2 , and R^2_{adjusted} values.

Table 4.4. Results for the NN 3-1 and MNN 2-3-1 models

%X Exp.	Neural Network Results (3-1)			Modular Neural Network Results (2-3-1)		
	X Cal.	Error	Error %	%X Cal.	Error	Error %
28.2	31.2	3.0	10.5	28.2	0.0	0.12
32.4	29.6	-2.8	8.7	25.6	-6.8	21.02
25.2	28.4	3.2	12.8	25.3	0.1	0.49
27.7	34.2	6.5	23.4	35.5	7.8	28.20
32.9	31.8	-1.1	3.3	34.2	1.3	3.81
32.9	31.8	-1.1	3.3	34.2	1.3	3.81
36.6	31.8	-4.8	13.1	34.2	-2.4	6.68
28.9	30.0	1.1	4.0	29.0	0.1	0.37
27.1	38.5	11.4	42.1	27.2	0.1	0.51
31.6	35.2	3.6	11.2	28.1	-3.5	11.13
29.1	32.5	3.4	11.8	29.1	0.0	0.05
27.1	32.8	5.7	21.2	29.8	2.7	9.85
46.3	48.3	2.0	4.3	50.7	4.4	9.57
78.2	74.3	-3.9	5.0	78.5	0.3	0.38
71.4	75.1	3.7	5.2	80.6	9.2	12.83
100.0	102.6	2.6	2.6	100.0	0.0	0.00
100.0	96.9	-3.1	3.1	99.9	-0.1	0.09
99.4	96.8	-2.6	2.7	99.9	0.5	0.55
36.6	33.5	-3.1	8.4	32.2	-4.4	11.93
53.1	49.8	-3.3	6.3	53.7	0.6	1.06
75.7	74.8	-0.9	1.2	74.4	-1.3	1.69
76.7	75.1	-1.6	2.0	76.7	0.0	0.04
100.0	102.9	2.9	2.9	100.0	0.0	0.00
98.5	96.8	-1.7	1.7	99.7	1.2	1.25
100.0	97.2	-2.8	2.8	99.9	-0.1	0.15
41.1	34.3	-6.8	16.5	35.5	-5.6	13.59
57.0	51.3	-5.7	10.1	56.1	-0.9	1.57
70.3	75.3	5.0	7.1	71.4	1.1	1.55
74.3	75.2	0.9	1.2	74.0	-0.3	0.45
100.0	103.2	3.2	3.2	100.0	0.0	0.00
100.0	96.6	-3.4	3.4	99.2	-0.8	0.78
100.0	97.4	-2.6	2.6	99.6	-0.4	0.41
R²	0.981			R²	0.989	
R²_{adjusted}	0.974			R²_{adjusted}	0.988	
RMSE	7.2			RMSE	7.7	

The graphical comparisons of the results obtained from the neural models with experimental data are given in Figure 4.5. Both models have a very good fit since all the data points are quite close to on the 45° ($y=x$) line.

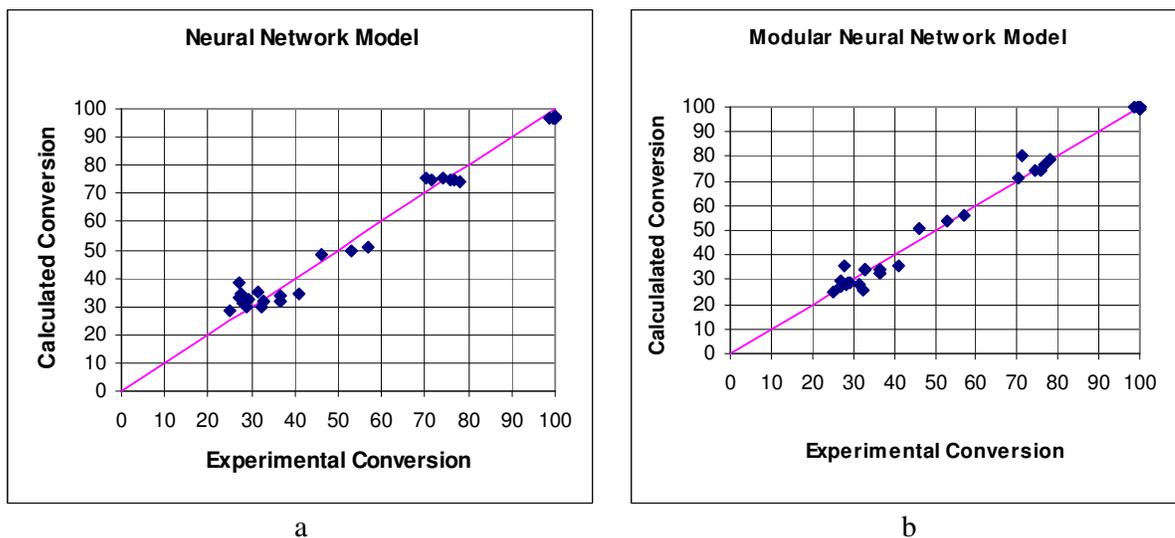


Figure 4.5. Comparison of (a) NN 3-1 and (b) MNN 2-3-1 results with experimental data

The residual analysis indicate that the modular neural network results are less scattered and majority of them are closer to horizontal line compare to monolithic network as seen in Figure 4.6.

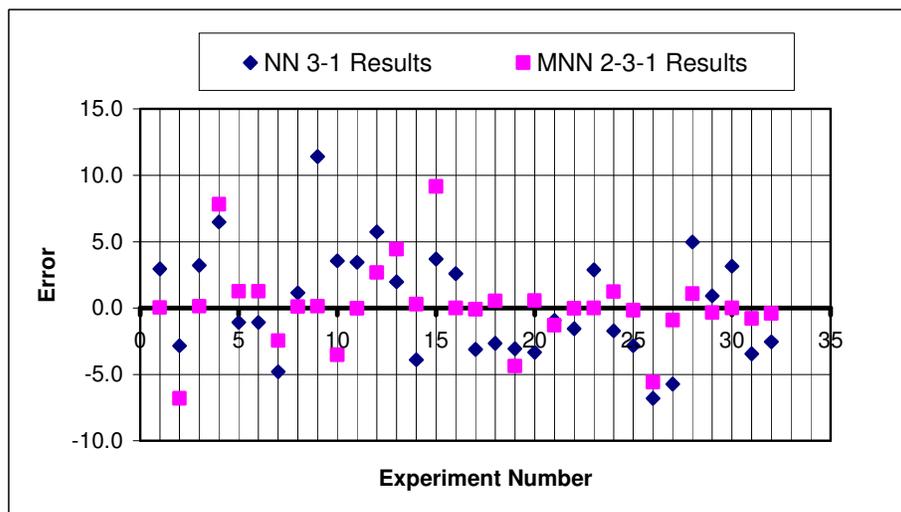


Figure 4.6. Residual analysis for the NN 3-1 and MNN 2-3-1 models

Although, the number of connections did not exceed the number of data points (32) in the modular neural network discussed in the previous section, to prevent any questions of over-fitting, a smaller neural network is searched for modeling the data. For example the data was also modeled with a modular neural network of only 1 neuron in the hidden layer of each input pattern, which is the smallest modular neural model possible to construct. The schematic representation of the model is given in Figure 4.7. The calculated CO conversions has a R^2 value of 0.971 which indicates a sufficient representation of the data without having much complexity compared to the quadratic regression equation (Table 4.5).

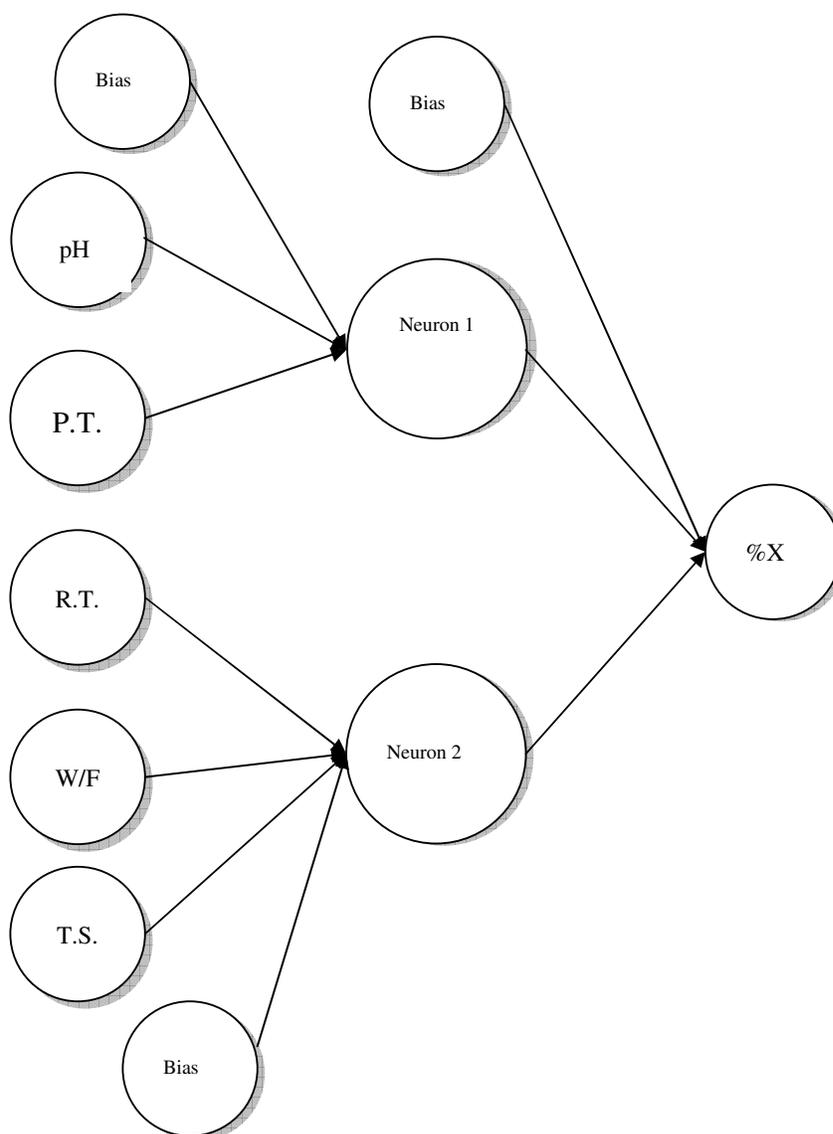


Figure 4.7. Schematic representation of the MNN 1-1-1 model

Table 4.5. Results for the MNN 1-1-1 and quadratic MR models

	Modular Neural Network Results (1-1-1)			Quadratic Multiple Regression Results		
%X Exp.	X Cal.	Error	Error %	%X Cal.	Error	Error %
28.2	34.1	5.9	20.77	27.82	-0.4	1.37
32.4	34.1	1.7	5.11	33.31	0.9	2.80
25.2	26.1	0.9	3.56	24.45	-0.8	2.98
27.7	34.1	6.4	22.95	27.59	-0.1	0.38
32.9	34.1	1.2	3.51	35.99	3.1	9.40
32.9	34.1	1.2	3.51	35.99	3.1	9.40
36.6	34.1	-2.5	6.95	35.99	-0.6	1.66
28.9	27.7	-1.2	3.98	30.05	1.1	3.96
27.1	34.1	7.0	25.67	27.59	0.5	1.82
31.6	34.1	2.5	7.77	31.74	0.1	0.43
29.1	29.7	0.6	2.07	28.70	-0.4	1.39
27.1	34.0	6.9	25.42	26.22	-0.9	3.23
46.3	41.3	-5.0	10.83	51.97	5.7	12.24
78.2	75.8	-2.4	3.09	74.27	-3.9	5.02
71.4	75.0	3.6	5.09	75.40	4.0	5.60
100.0	99.9	-0.1	0.15	104.33	4.3	4.33
100.0	95.8	-4.2	4.21	99.93	-0.1	0.07
99.4	95.6	-3.8	3.79	90.29	-9.1	9.16
36.6	34.0	-2.6	7.04	32.73	-3.9	10.57
53.1	41.4	-11.7	21.96	55.43	2.3	4.39
75.7	76.3	0.6	0.83	74.69	-1.0	1.33
76.7	75.6	-1.1	1.47	79.31	2.6	3.40
100.0	99.9	-0.1	0.13	105.19	5.2	5.19
98.5	96.0	-2.5	2.54	100.57	2.1	2.11
100.0	95.9	-4.1	4.15	92.68	-7.3	7.32
41.1	34.1	-7.0	17.14	35.99	-5.1	12.42
57.0	41.6	-15.4	27.03	55.65	-1.4	2.38
70.3	76.9	6.6	9.35	71.86	1.6	2.22
74.3	76.1	1.8	2.45	79.97	5.7	7.64
100.0	99.9	-0.1	0.12	102.81	2.8	2.81
100.0	96.2	-3.8	3.79	97.97	-2.0	2.03
100.0	96.1	-3.9	3.94	91.82	-8.2	8.18
R²	0.971			R²	0.984	
R²_{adjusted}	0.910			R²_{adjusted}	0.976	
RMSE	6.0			RMSE	6.3	

The error distribution around the zero error line is given in Figure 4.8. Although the MNN errors are seen to be more scattered than those for multiple regression, the neural network can be still considered as successful since the number of weights (10) is small when compared to the 21 coefficients in the multiple regression.

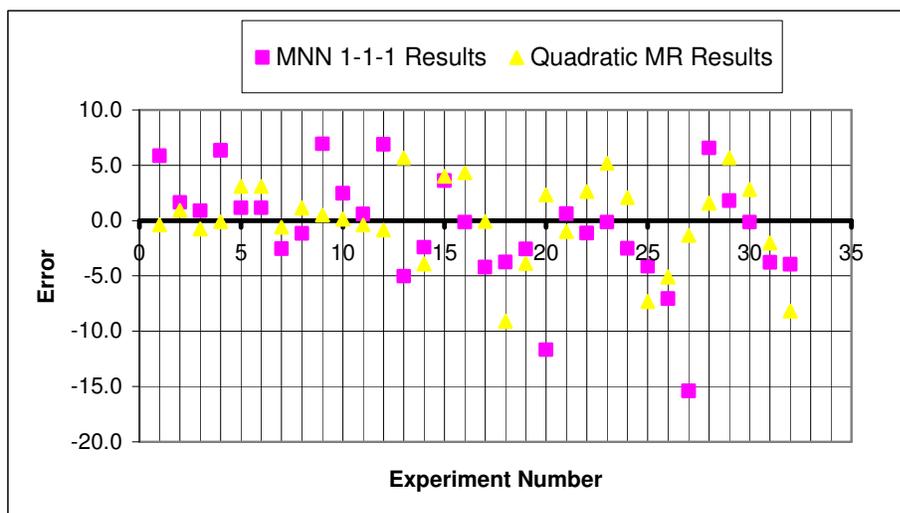


Figure 4.8: Residual analysis for the MNN 1-1-1 and quadratic multiple regression models

k-fold cross validation analysis was also done by dividing the experimental data into 8 ($k = 8$) subsets. Since time on stream has included as a parameter into data set, each subset includes all 3 time on stream values of the same experiment. The change of time stream of the catalyst does not change the main character of it. Hence, in cross validation all three points should be excluded together. Otherwise, the exclusion of one data point will be compensated by the others and cross validation results would seem better than the model actually would have performed (Günay, 2007). The RMSE's of the remaining data points are given in Table 4.6.

Table 4.6: k-fold cross validation analysis for the MNN 1-1-1 model

Subset Number	Experiments Excluded	RMSE
1	2,4,9,11	6.3
2	1,2,19,26	6.6
3	3,13,20,27	6.7
4	5,14,21,28	6.9
5	6,15,22,29	6.3
6	7,16,23,30	6.4
7	8,17,24,31	6.5
8	10,18,25,32	6.1
	Average	6.48

The RMSE's of the subsets are balanced and close to one used for prediction of the entire range of the experimental data, which indicates that the model covers all data points with sufficient precision.

The results obtained from the Modular Neural Network with 1-1-1 Structure are sufficiently accurate with a R^2 of 0.971 and RMSE of 6.1. However, the R^2 's obtained from MNN 1-2-1 and MNN 2-2-1 are 0.839 and 0.958 which suggest that more connections between the parameters are not always mean better fit. Consequently, it is best to start with a small neural network and enlarge it. Also, one should not enlarge the network too much to avoid over-fitting of the data, in which each neuron would memorize a data point and the model will perform well on the given points but will fit poorly on the data that it has not encountered before.

4.1.3. Modeling the Effects of Target Metal Content on CO Conversion

The effects of target Cu, Co and Ce contents of the catalyst were modeled via neural network of 2 neurons (NN 2-1). The schematic representation of the neural network model is given in Figure 4.9. Since there were only 14 data points, and 3 input parameters were similar in nature, a modular approach was decided to be unnecessary. The calculated versus experimental conversions with the neural network model and quadratic multiple regression are given in Table 4.7.

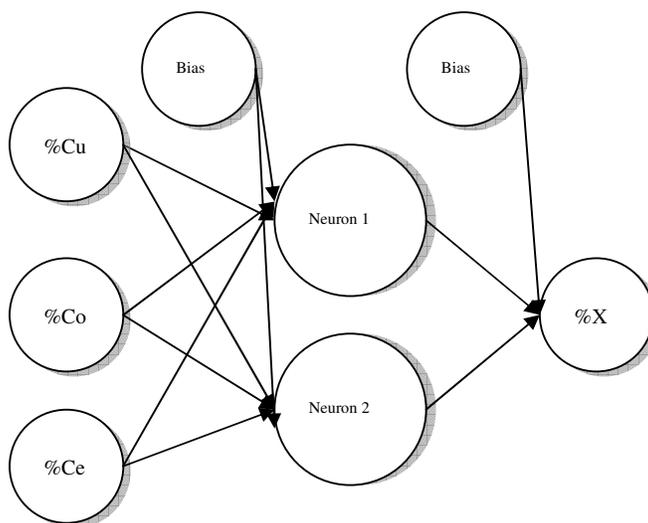


Figure 4.9: Schematic representation of the NN 2-1 model

Table 4.7. Calculated conversions for the NN 2-1 and quadratic multiple regression models

	Neural Network Results (2-1)			Quadratic Multiple Regression Results		
%X Exp.	X Cal.	Error	Error %	%X Cal.	Error	Error %
1.90	0.53	-1.4	72.3	1.36	-0.5	28.56
100.00	64.26	-35.7	35.7	61.13	-38.9	38.87
87.62	64.51	-23.1	26.4	74.38	-13.2	15.11
84.97	61.11	-23.9	28.1	43.69	-41.3	48.59
71.67	64.81	-6.9	9.6	100.00	28.3	39.53
2.73	21.14	18.4	674.4	41.13	38.4	1406.53
2.05	16.33	14.3	696.6	28.39	26.3	1284.78
0.70	0.53	-0.2	24.7	1.36	0.7	93.91
6.49	2.15	-4.3	66.9	7.91	1.4	21.86
10.75	16.33	5.6	51.9	28.39	17.6	164.07
0.23	1.93	1.7	738.9	-0.59	-0.8	356.51
8.41	2.15	-6.3	74.5	7.91	-0.5	5.96
86.87	58.42	-28.4	32.7	64.72	-22.1	25.49
4.05	1.93	-2.1	52.4	-0.59	-4.6	114.57
R²	0.825			R²	0.803	
R²_{adjusted}	0.810			R²_{adjusted}	0.488	
RMSE	44.0			RMSE	22.0	

The neural network model fit the data much better than the multiple regression with a R^2_{adj} of 0.810 rather than 0.488. The comparison graphs for modeling results with experimental data are given in Figure 4.10 for both models. Although the neural network results are quite closer to the $y = x$ line indicating a better fit, they are not as good as the ones that obtained in the previous sections. The similar conclusions can be drawn from the residual analysis in Figure 4.11.

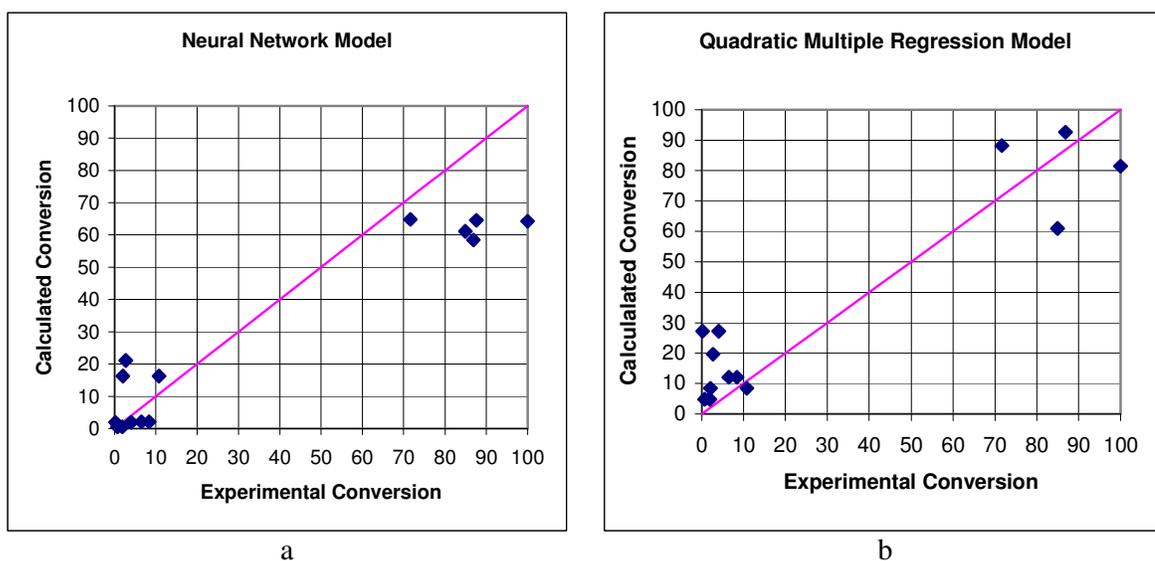


Figure 4.10. Comparison graphs for the (a) NN 2-1 and (b) MR results with experimental data

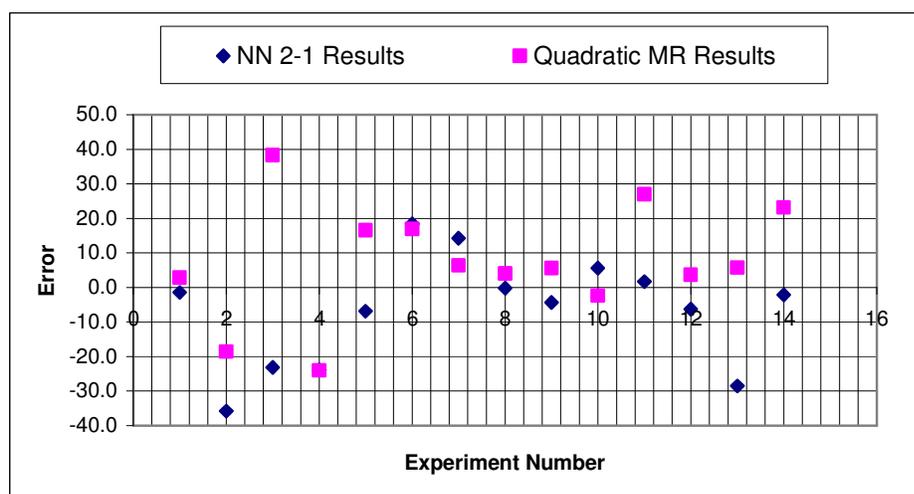


Figure 4.11. Residual analysis of the NN 2-1 and multiple regression models

The experiments for the effects of metal loading were performed in a single time on stream of 3 hours, preventing to increase the number of data points for a better fit. Combining the data related to the metal loading and reaction conditions did not result a good fit both because they are from different experimental work and their common parameters are not as many as in the case of combining precipitation and reaction conditions. Hence the results could not be improved further.

4.1.4. Modeling the Effects of Reaction Conditions on CO Conversion

The reaction temperature and W/F ratio were also modeled, without combining with the preparation parameters, using neural network (NN 4-1), modular neural network (MNN 2-1-2-1) and multiple regression. In order to increase the data number by 3 fold, time on stream is considered as a model parameter. The schematic representation of the MNN 2-1-2-1 is given in Figure 4.12 while the results obtained from this models and NN 4-1 are given in Table 4.8.

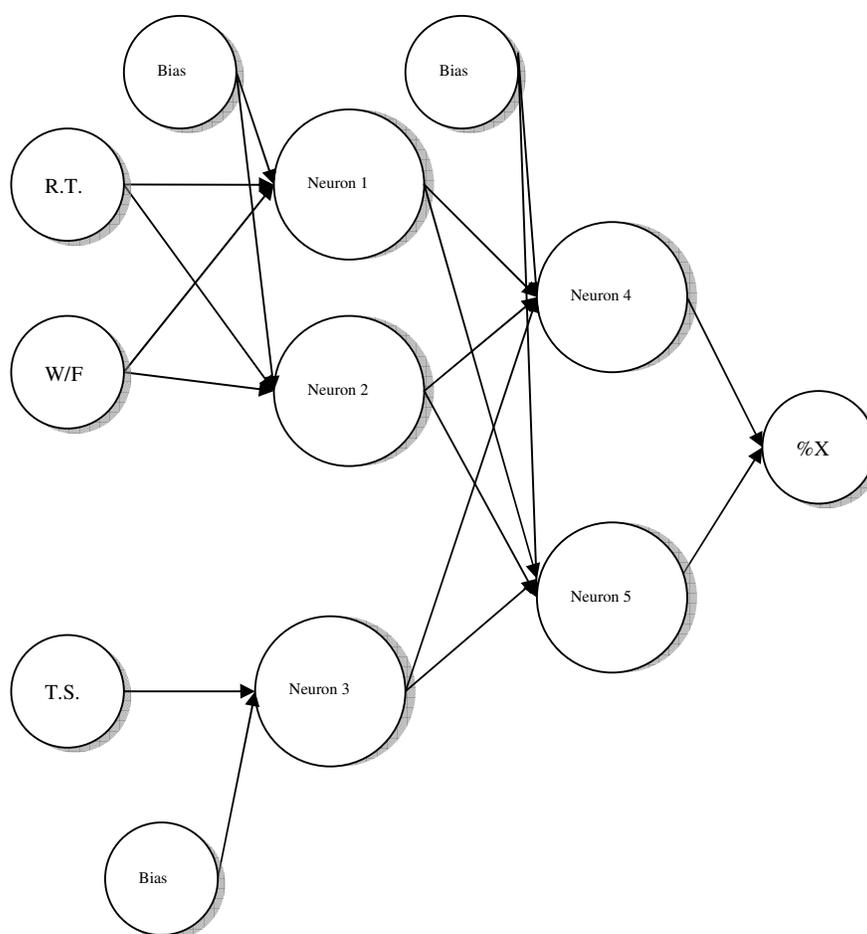


Figure 4.12. Schematic representation of the MNN 2-1-2-1 model

Table 4.8. Results obtained from the NN 4-1 and MNN 2-1-2-1 models

	Neural Network Results (4-1)			Modular Neural Network Results (2-1-2-1)		
%X Exp.	X Cal.	Error	Error %	%X Cal.	Error	Error %
27.1	27.81	0.7	2.6	34.19	7.1	26.16
46.3	40.82	-5.5	11.8	51.24	4.9	10.66
78.2	77.29	-0.9	1.2	75.62	-2.6	3.30
71.4	72.18	0.8	1.1	76.39	5.0	6.98
100.0	99.61	-0.4	0.4	101.34	1.3	1.34
100.0	98.43	-1.6	1.6	100.34	0.3	0.34
99.4	100.98	1.6	1.6	99.93	0.5	0.53
36.6	35.43	-1.2	3.2	34.18	-2.4	6.61
53.1	53.37	0.3	0.5	51.11	-2.0	3.74
75.7	77.82	2.1	2.8	75.14	-0.6	0.74
76.7	74.38	-2.3	3.0	75.89	-0.8	1.05
100.0	100.05	0.1	0.1	100.66	0.7	0.66
98.5	99.31	0.8	0.8	99.66	1.2	1.18
100.0	100.00	0.0	0.0	99.24	-0.8	0.76
41.1	39.11	-2.0	4.8	34.18	-6.9	16.85
57.0	63.66	6.7	11.7	50.99	-6.0	10.54
70.3	80.74	10.4	14.9	74.67	4.4	6.22
74.3	74.37	0.1	0.1	75.41	1.1	1.50
100.0	100.60	0.6	0.6	99.98	0.0	0.02
100.0	100.52	0.5	0.5	98.98	-1.0	1.02
100.0	99.35	-0.6	0.6	98.56	-1.4	1.44
R²	0.983			R²	0.981	
R²_{adjusted}	0.983			R²_{adjusted}	0.980	
RMSE	14.4			RMSE	10.7	

The NN results are slightly better than the MNN ones. However, the modular neural network has an extra hidden layer but 5 less connections than the NN model of 4 neurons in the hidden layer. The R^2 values are close to each other which are 0.983 for NN and 0.981 for MNN. R^2 for multiple regression, on the other hand was 0.965, and R^2_{adj} was 0.930. The small difference in the correlation results suggest that when the input number is low, using MNN model does not make much difference. However, the effect of 5 less connections used in the MNN should not be underestimated. For example it was not possible to lower the number of neurons in the hidden layer in ANN (3 neurons gave a poor fit) while it could be done with MNN.

The comparison graphs of the ANN and MNN are given in Figure 4.13. The graphs are similar to each other since both models gave a good fit of the data.

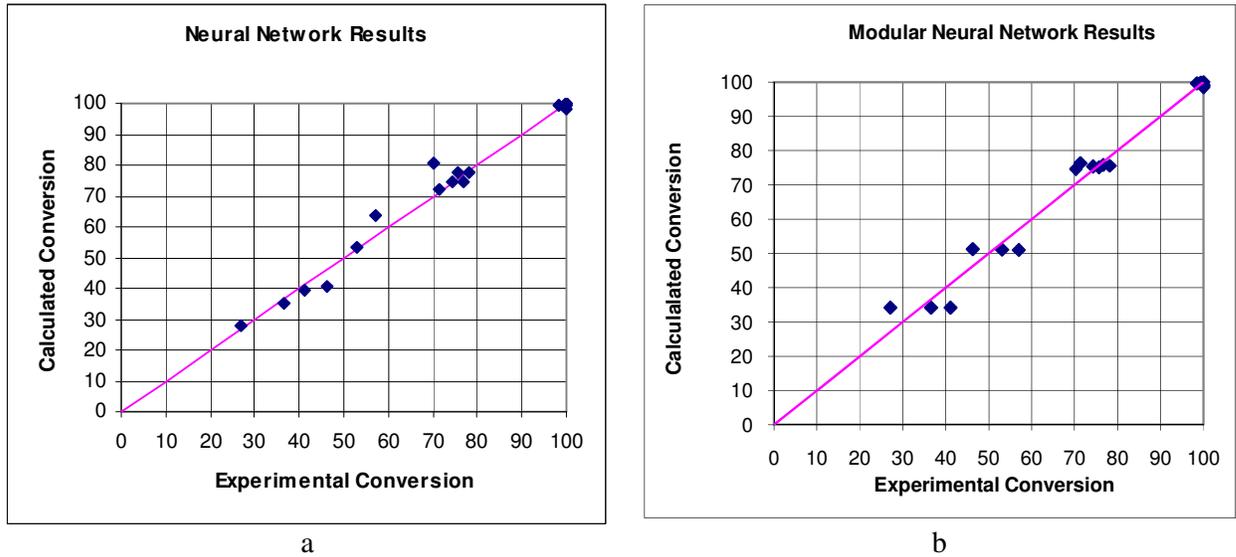


Figure 4.13. Comparison graphs for the (a) NN 4-1 and (b) MNN 2-1-2-1 results with experimental data

The distributions of the errors for both models are also given in Figure 4.14. Both neural models results are close to each other, and they are better than multiple regression model.

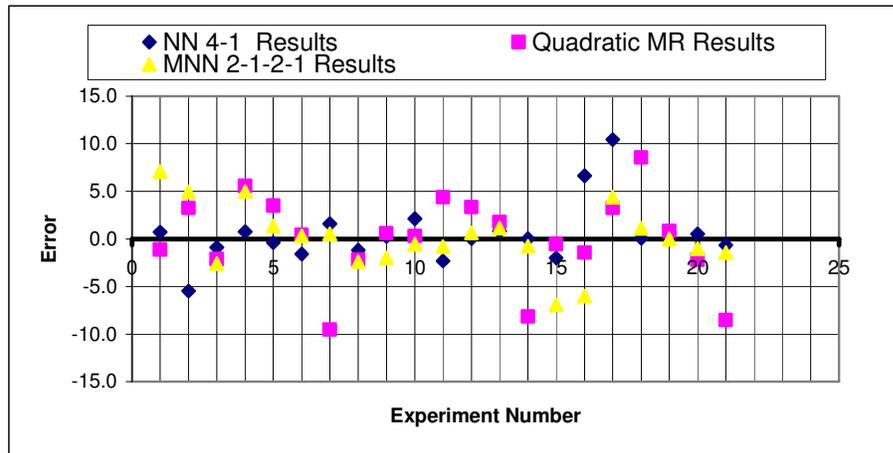


Figure 4.14: Residual analysis for the NN 4-1, MNN 2-1-2-1, and quadratic multiple regression models

4.1.5. Modeling the Effects of Design and Reaction Parameters Together on CO Conversion

The design parameters, namely precipitation pH, precipitation temperature, target Cu, Co, and Ce weight percents are modeled with the reaction parameters; reaction temperature, W/F and time on stream (sampling time) together using both neural and modular neural networks. The catalyst preparation and reaction parameters were fed to the MNN as separate input groups similar to the MNN modeling works discussed in previous sections. The correlation measures of all models tested are given in Table 4.9. Among the two different NN structures the model with 3 neurons in the hidden layer has better correlation coefficients which again supports that more neuron does not mean to give better result. Similarly among three different modular architectures constructed the one with 3 neurons for each input pattern has the best results.

Table 4.9: Comparison of regression coefficients of different network architectures for design and reaction parameters modeling

Input Structure	Network Model	R²	R²_{adjusted}	S_e	RMSE
8	NN 4-1	0.896	0.896	11.61	77.0
5	NN 3-1	0.949	0.935	8.10	16.2
5-3	MNN 4-3-1	0.967	0.937	6.58	30.9
5-3	MNN 3-3-1	0.963	0.959	6.96	20.7
5-3	MNN 2-3-1	0.953	0.937	6.88	15.0

The comparison of the results best NN which has 3 neurons in the hidden layer, and the modular neural network, which has two groups of inputs; design and reaction parameters with three neurons for each in the hidden layer (Figure 4.15) are given in Table 4.10. It should be noted that, although MNN has 3 more neurons in the hidden layer and has only 6 more number of connections. Dividing the input sequence reduced the number of additional connections arising from the increase in the number of neurons.

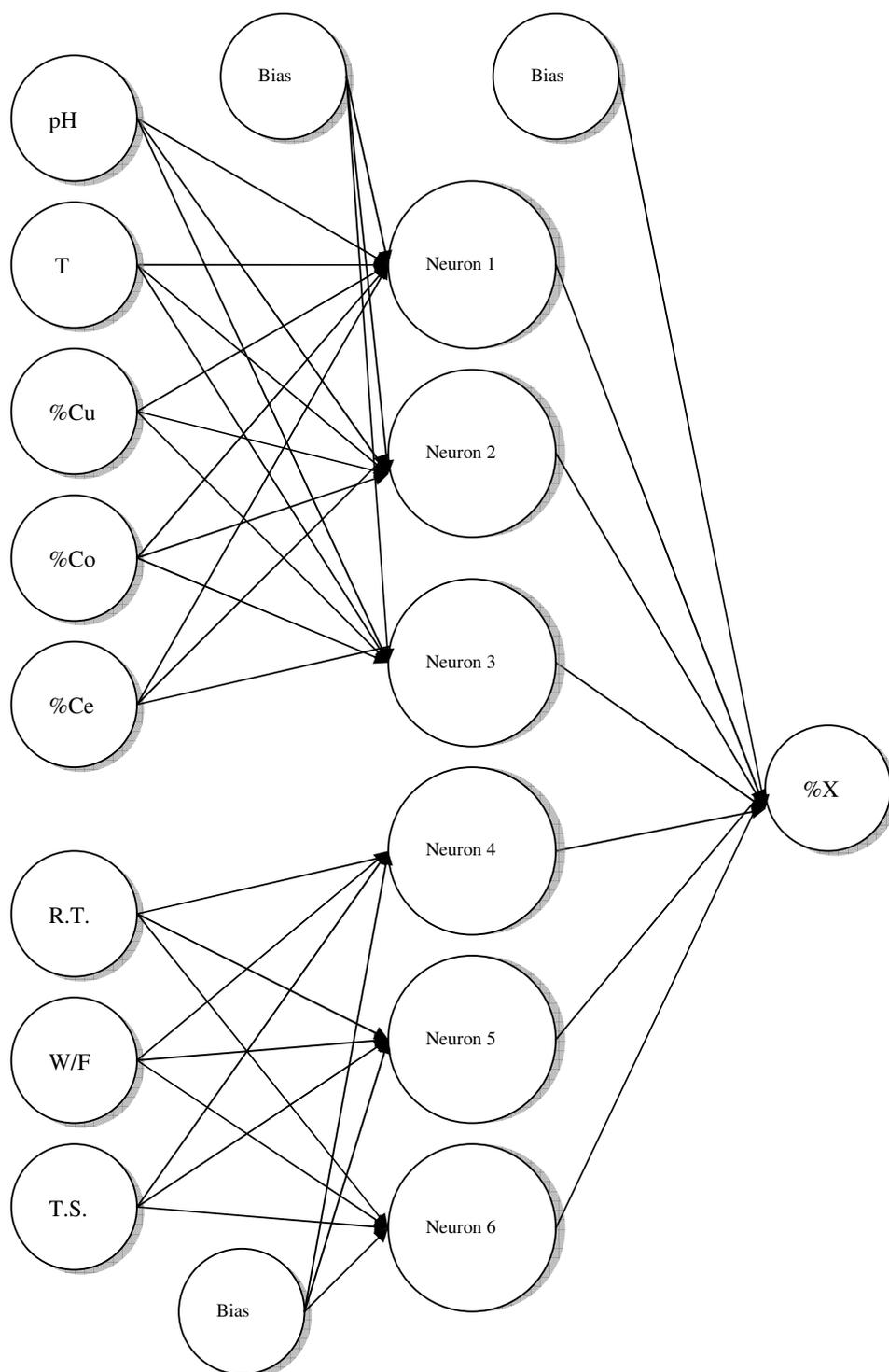


Figure 4.15. Schematic representation of the MNN 3-3-1 model

Table 4.10. Results for the NN 4-1 and MNN 3-3-1 models

%X Exp.	Neural Network Results (3-1)			Modular Neural Network Results (3-3-1)		
	X Cal.	Error	Error %	%X Cal.	Error	Error %
28.2	29.8	1.6	5.7	33.54	5.3	19.0
32.4	29.6	-2.8	8.7	31.83	-0.6	1.8
25.2	29.3	4.1	16.1	27.22	2.0	8.0
27.7	30.2	2.5	9.1	34.87	7.2	25.9
32.9	30.0	-2.9	8.8	33.38	0.5	1.4
32.9	30.0	-2.9	8.8	33.38	0.5	1.4
36.6	30.0	-6.6	18.0	33.38	-3.2	8.8
28.9	29.6	0.7	2.5	30.16	1.3	4.4
27.1	30.7	3.6	13.3	35.48	8.4	30.9
31.6	30.5	-1.1	3.5	34.48	2.9	9.1
29.1	30.0	0.9	3.2	32.29	3.2	11.0
27.1	29.6	2.5	9.1	26.65	-0.4	1.7
46.3	51.5	5.2	11.3	48.82	2.5	5.4
78.2	68.3	-9.9	12.7	82.54	4.3	5.6
71.4	71.6	0.2	0.3	73.76	2.4	3.3
100.0	93.1	-6.9	6.9	103.59	3.6	3.6
100.0	102.3	2.3	2.3	93.15	-6.9	6.9
99.4	99.3	-0.1	0.1	90.66	-8.7	8.8
36.6	29.7	-6.9	18.9	28.65	-8.0	21.7
53.1	53.2	0.1	0.2	53.77	0.7	1.3
75.7	68.0	-7.7	10.2	85.54	9.8	13.0
76.7	74.2	-2.5	3.2	76.88	0.2	0.2
100.0	92.2	-7.8	7.8	101.36	1.4	1.4
98.5	101.4	2.9	2.9	93.77	-4.7	4.8
100.0	100.0	0.0	0.0	91.29	-8.7	8.7
41.1	29.8	-11.3	27.4	31.54	-9.6	23.3
57.0	55.0	-2.0	3.5	59.49	2.5	4.4
70.3	67.6	-2.7	3.8	88.58	18.3	26.0
74.3	76.7	2.4	3.3	80.64	6.3	8.5
100.0	91.2	-8.8	8.8	100.38	0.4	0.4
100.0	100.4	0.4	0.4	94.92	-5.1	5.1
100.0	100.7	0.7	0.7	92.65	-7.4	7.4
1.90	-3.3	-5.2	273.7	-0.05	-2.0	102.8
100.00	90.4	-9.6	9.6	74.86	-25.1	25.1
87.62	89.6	2.0	2.3	83.04	-4.6	5.2
84.97	76.9	-8.1	9.6	76.03	-8.9	10.5
71.67	86.6	14.9	20.8	84.06	12.4	17.3
2.73	19.5	16.7	612.5	5.44	2.7	99.3
2.05	15.9	13.9	677.4	8.42	6.4	310.7
0.70	-3.3	-4.0	571.4	-0.05	-0.8	107.5
6.49	4.4	-2.1	32.6	5.82	-0.7	10.3
10.75	15.9	5.2	48.2	8.42	-2.3	21.7
0.23	5.1	4.9	2138.7	1.98	1.8	761.8
8.41	4.4	-4.0	48.0	5.82	-2.6	30.8
86.87	52.1	-34.8	40.1	83.73	-3.1	3.6
4.05	5.1	1.1	27.1	1.98	-2.1	51.1
R²	0.949			R²	0.963	
R²_{adjusted}	0.935			R²_{adjusted}	0.959	
RMSE	16.2			RMSE	20.7	

The advantage of using modular neural network clearly appears in these models. MNN has an R^2 of 0.963 and RMSE of 20.7. NN with 3 neurons in the hidden layer, however, has a RMSE of 53.7 which indicates a fit as not good as MNN although R^2 of 0.949 is quite close to the one of MNN. The comparison graphs of the models with the experimental data can be seen in Figure 4.16.

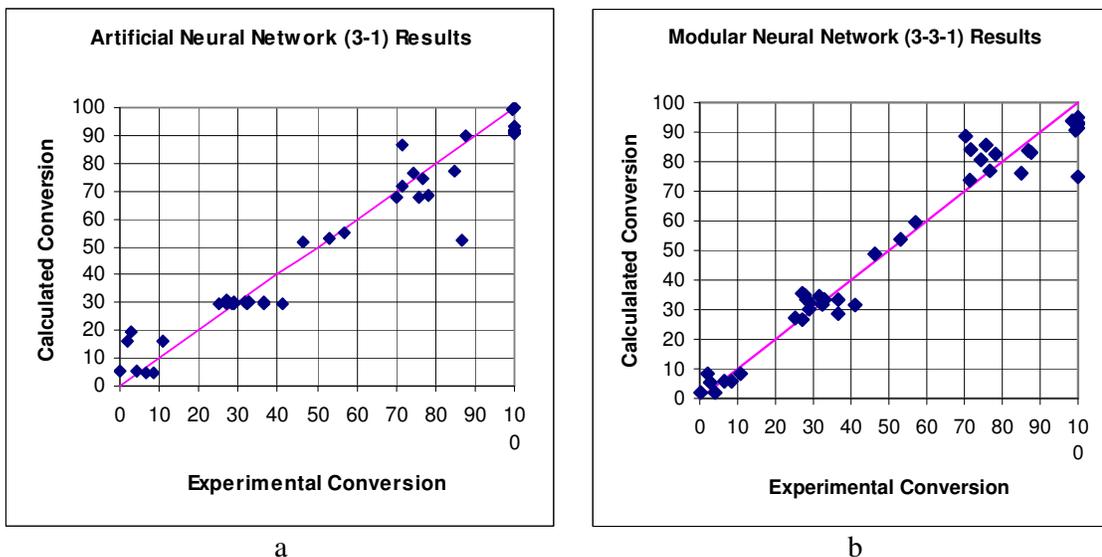


Figure 4.16: Comparison graphs of the (a) NN 3-1 and (b) MNN 3-3-1 results with the experimental data

The residual analysis of the NN 3-1 is given in Figure 4.17. Modular neural network errors have fewer deviations from the zero error line.

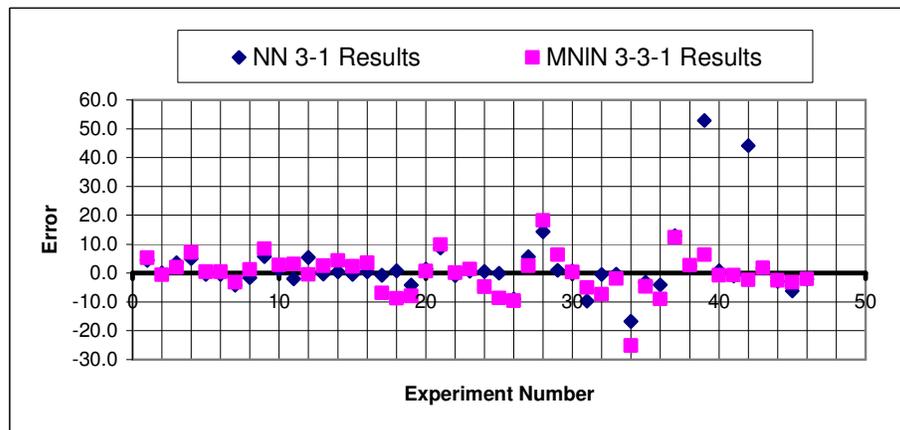


Figure 4.17. Residual analysis for the NN 3-1 and MNN 3-3-1 models

4.1.6. Modeling of CO Conversion from Measured Catalyst Properties

There is a significant difference in target metal contents and measured metal contents in the prepared catalysts due to nature of co-precipitation technique (Kibar 2005, Özdemir 2006). In the sections 4.1.3 and 4.1.5, the metal contents used for modeling of conversion were the target values. Here, the measured Cu, Co, and Ce metal percents, and the total surface area (TSA) measured at precipitation pH, 10.0 and temperature 37.5 °C were used for modeling with a neural network of 2 neurons in the hidden layer and a modular neural network of 2 neurons for metal contents and 1 neuron for TSA. The schematic representation of the MNN is given in Figure 4.18. The results of the neural networks are given in Table 4.11 for comparison.

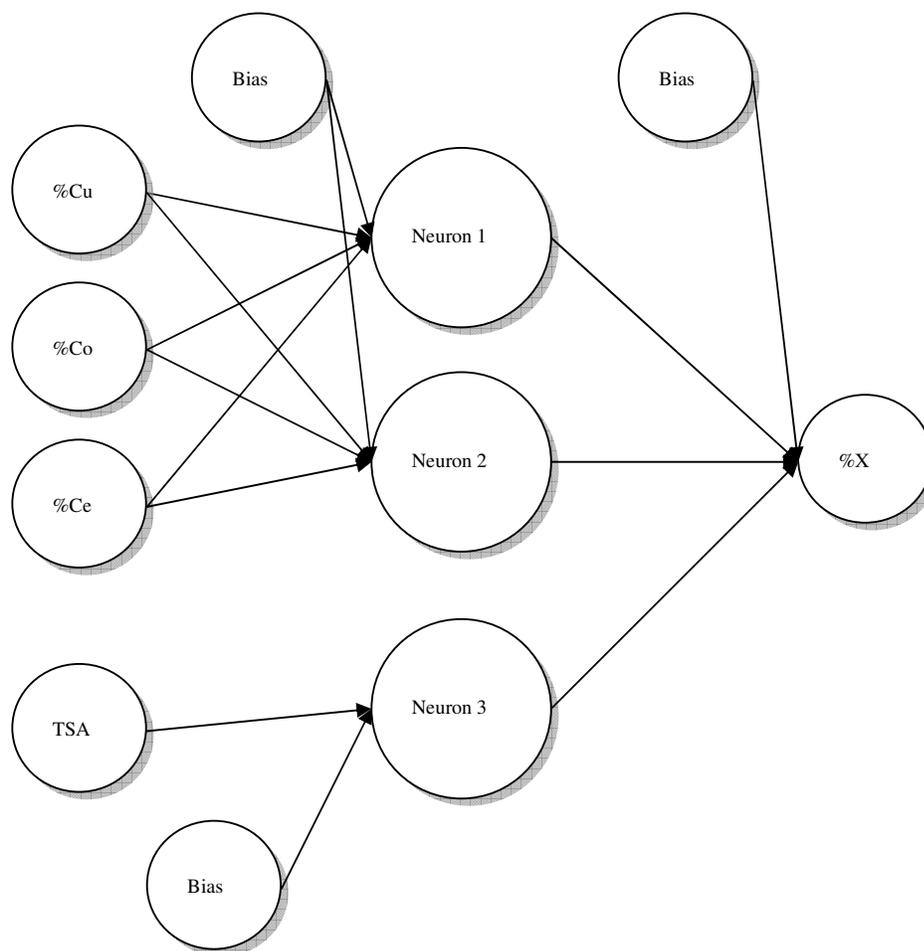


Figure 4.18. Schematic representation of the MNN 2-1-1 model

Table 4.11. Results obtained from the NN 2-1 and MNN 2-1-1 models

	Neural Network Results (2-1)			Modular Neural Network Results (2-1-1)		
%X Exp.	X Cal.	Error	Error %	%X Cal.	Error	Error %
1.9	2.9	0.99	51.95	2.1	0.19	10.00
100.0	99.7	-0.27	0.27	90.1	-9.94	9.94
87.6	87.9	0.25	0.29	87.7	0.09	0.10
85.0	100.9	15.86	18.65	82.4	-2.59	3.05
71.7	71.7	-0.01	0.01	89.1	17.43	24.31
2.7	3.9	1.17	43.32	5.6	2.89	107.04
2.1	0.1	-1.96	93.44	3.3	1.17	55.71
0.7	1.8	1.10	157.03	1.1	0.43	61.43
6.5	7.5	0.99	15.26	9.9	3.43	52.77
10.8	9.6	-1.20	11.07	9.6	-1.25	11.57
0.2	2.3	2.10	1049.75	4.0	3.76	1880.00
8.4	8.4	-0.02	0.26	10.2	1.81	21.55
86.9	86.4	-0.46	0.53	83.0	-3.87	4.45
4.5	2.3	-2.16	47.91	3.7	-0.79	17.56
R²	0.988			R²	0.979	
R²_{adjusted}	0.988			R²_{adjusted}	0.979	
RMSE	16.5			RMSE	21.6	

The R^2 value for the artificial neural network is 0.988 and for the modular neural network is 0.979. Both models approximated the experimental data well. The NN model is slightly better than MNN. Since the number of data points is relatively small from the other models, the modular approach hasn't revealed its advantages. The conversion data was also modeled with multiple regression which gives a R^2 of 0.815. The neural models have a clear advantage on fitting the data. The comparison of the results of the neural models with experimental data is given in Figure 4.19.

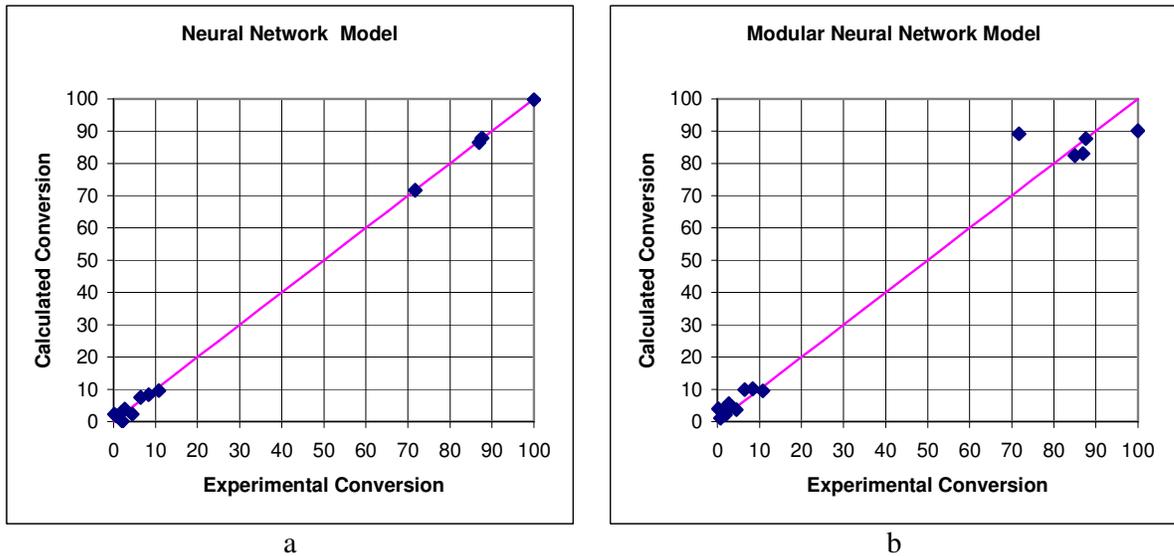


Figure 4.19. Comparison of the (a) NN 2-1 and (b) MNN 2-1-1 results with experimental data

The residual analyses of the three models are given in Figure 4.20. Neural network results have lower errors than multiple regression and closer distribution to the zero error line.

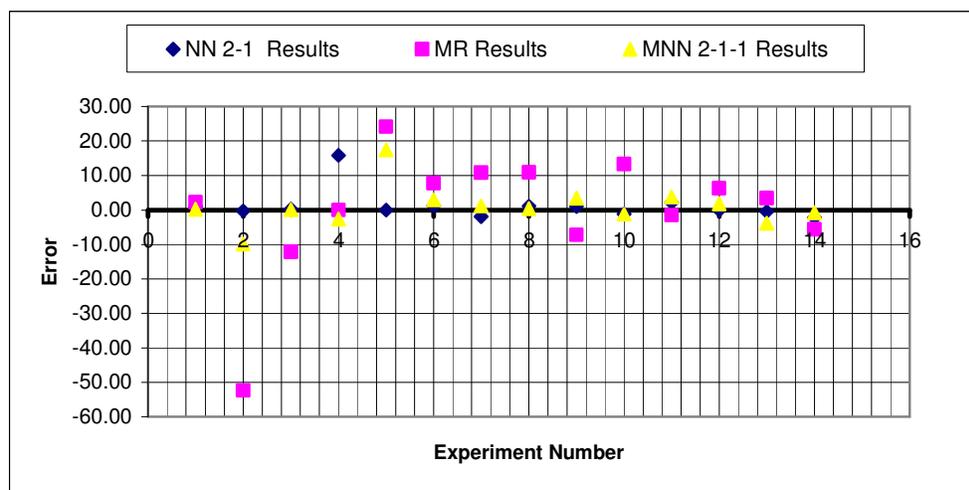


Figure 4.20. Residual analysis for the NN 2-1, MNN 2-1-1 and MR models

4.1.7. Modeling the Measured Catalyst Properties and Reaction Conditions over CO Conversion

Next the measured metal percents and the total surface area were modeled together with reaction conditions (reaction temperature, W/F and time on stream) using an artificial neural network of 2 neurons in the hidden layer and with a modular neural network, which has a two inputs sequence with 1 neurons in each. The schematic representation of the MNN 1-1-1 is given in Figure 4.21.

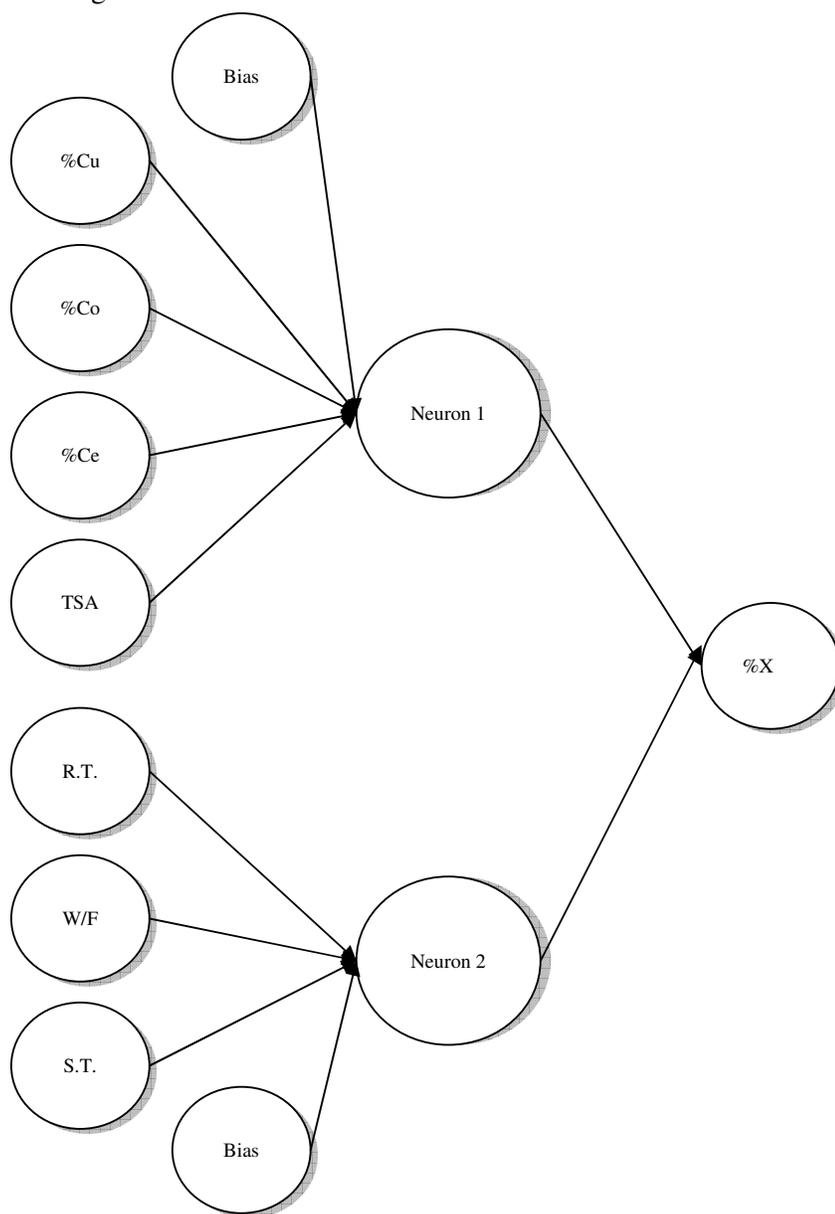


Figure 4.21. Schematic representation of the MNN 1-1-1 model

The predicted and experimental results were compared in Table 4.12. The artificial neural network 2-1 has R^2 of 0.991 and MNN has 0.974. Both models fit the data well. MNN 1-1-1 has only 1 Neuron in each hidden layer and 12 connections which are 7 less than the NN indicating a much lower risk of over-fitting.

Table 4.12. Results for the NN 2-1 and MNN 1-1-1 models

%X Exp.	Neural Network Results (2-1)			Modular Neural Network Results (1-1-1)		
	X Cal.	Error	Error %	%X Cal.	Error	Error %
28.2	26.1	-2.1	7.5	27.1	-1.1	3.80
32.4	33.4	1.0	3.2	32.8	0.4	1.15
25.2	27.2	2.0	7.9	27.4	2.2	8.81
27.7	26.2	-1.5	5.6	27.2	-0.5	1.69
32.9	40.9	8.0	24.3	33.0	0.1	0.31
32.9	33.5	0.6	1.9	32.5	-0.4	1.31
36.6	35.2	-1.4	3.8	34.0	-2.6	7.14
28.9	29.7	0.8	2.7	28.1	-0.8	2.89
27.1	26.8	-0.3	1.2	27.2	0.1	0.47
31.6	29.1	-2.5	7.9	27.4	-4.2	13.25
29.1	31.1	2.0	6.9	27.7	-1.4	4.96
27.1	29.0	1.9	7.0	40.7	13.6	50.12
46.3	56.9	10.6	23.0	40.9	-5.4	11.68
78.2	77.3	-0.9	1.2	76.1	-2.1	2.72
71.4	70.2	-1.2	1.7	75.5	4.1	5.78
100.0	100.4	0.4	0.4	101.1	1.1	1.15
100.0	99.6	-0.4	0.4	99.2	-0.8	0.85
99.4	99.4	0.0	0.0	99.1	-0.3	0.29
36.6	33.9	-2.7	7.5	40.7	4.1	11.16
53.1	57.0	3.9	7.3	40.9	-12.2	23.02
75.7	74.7	-1.0	1.3	74.9	-0.8	1.08
76.7	73.3	-3.4	4.5	74.3	-2.4	3.10
100.0	100.4	0.4	0.4	101.1	1.1	1.14
98.5	99.3	0.8	0.8	99.0	0.5	0.49
100.0	99.4	-0.6	0.6	98.9	-1.1	1.06
41.1	41.7	0.6	1.5	40.7	-0.4	1.02
57.0	57.0	0.0	0.0	40.9	-16.1	28.31
70.3	72.3	2.0	2.9	73.7	3.4	4.80
74.3	73.4	-0.9	1.2	73.1	-1.2	1.60
100.0	100.4	0.4	0.4	101.1	1.1	1.13
100.0	99.0	-1.0	1.0	98.8	-1.2	1.20
100.0	99.2	-0.8	0.8	98.8	-1.2	1.25
R^2	0.991			R^2	0.974	
R^2_{adjusted}	0.987			R^2_{adjusted}	0.946	
RMSE	5.0			RMSE	6.5	

The comparison curves of calculated results from the models and the experimental data is given in Figure 4.21 while the residual analysis was presented in Figure 4.22.

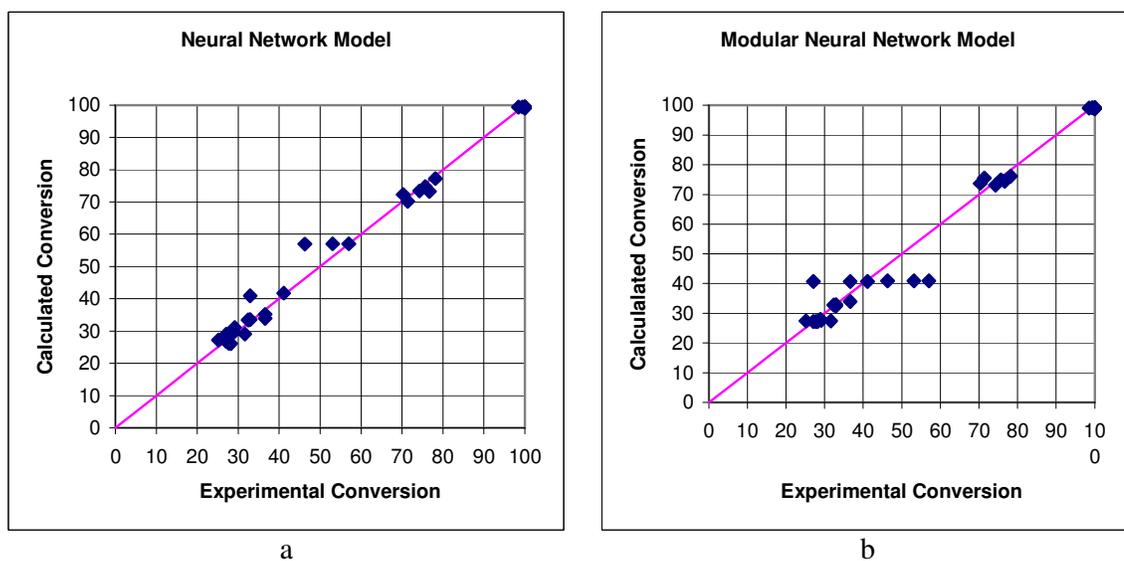


Figure 4.22. Comparison graphs of the (a) NN 2-1 and (b) MNN 1-1-1 results with experimental data

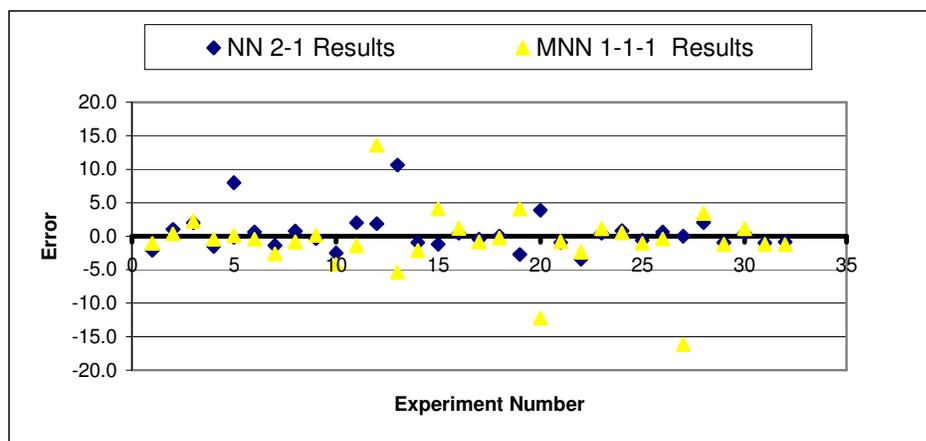


Figure 4.23. Residual analysis for NN 2-1 and MNN 1-1-1 models

4.1.8. Modeling of CO Conversion from Design and Reaction Studies Together with Measured Metal Content

Finally the CO conversion was modeled using catalyst preparation conditions (pH and precipitation temperature), measured metal contents of catalysts and reaction conditions were modeled together using an NN of 3-1 structure and MNN of 2-2-1 structure. Although, modular neural network has one more neuron than the artificial one, it has six less connections due to its modular nature. The schematic representation of the MNN 2-2-1 model is given in Figure 4.24. The results obtained from these models are given in Table 4.13.

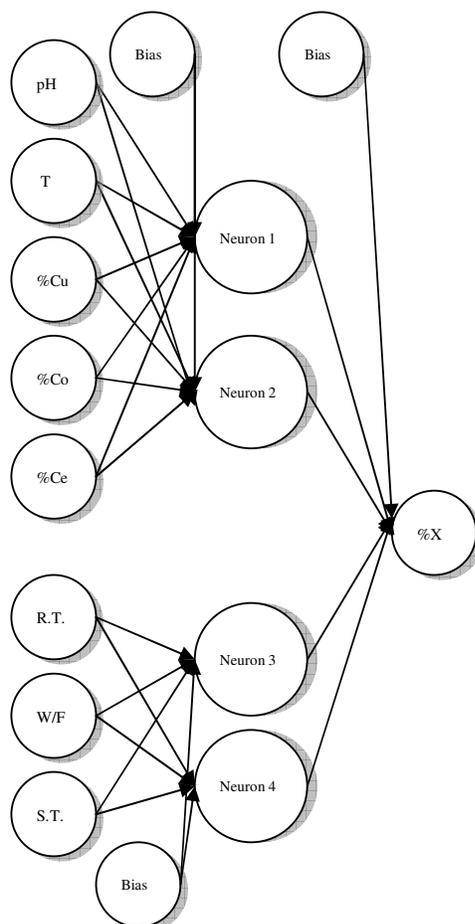


Figure 4.24. Schematic representation of the MNN 2-2-1 model

R^2 values are reasonable high for both models while MNN 2-2-1 seems slightly better. The fitness of both models are also evident from calculated versus experimental CO conversion plot (Figure 4.25) and residual analysis in Figure 4.26.

Table 4.13. Results for the NN 3-1 and MNN 2-2-1 models

	Neural Network Results (3-1)			Modular Neural Network Results (2-2-1)		
%X Exp.	X Cal.	Error	Error %	%X Cal.	Error	Error %
28.2	23.0	-5.2	18.4	27.3	-0.9	3.0
32.4	25.6	-6.8	20.9	38.1	5.7	17.7
25.2	19.7	-5.5	21.8	25.2	0.0	0.0
27.7	40.7	13.0	46.9	31.2	3.5	12.7
32.9	22.1	-10.8	32.7	35.0	2.1	6.5
32.9	33.2	0.3	0.9	36.3	3.4	10.4
36.6	33.4	-3.2	8.7	35.8	-0.8	2.1
28.9	23.2	-5.7	19.7	27.4	-1.5	5.2
27.1	40.8	13.7	50.4	24.3	-2.8	10.2
31.6	28.8	-2.8	9.0	33.7	2.1	6.8
29.1	24.1	-5.0	17.3	30.2	1.1	3.9
27.1	12.6	-14.5	53.3	33.0	5.9	21.7
46.3	48.1	1.8	4.0	44.3	-2.0	4.3
78.2	67.3	-10.9	14.0	80.2	2.0	2.5
71.4	75.4	4.0	5.6	80.7	9.3	13.0
100.0	99.1	-0.9	0.9	100.6	0.6	0.6
100.0	98.7	-1.3	1.3	92.1	-7.9	7.9
99.4	89.5	-9.9	10.0	90.4	-9.0	9.1
36.6	25.6	-11.0	30.2	33.6	-3.0	8.2
53.1	59.4	6.3	11.9	53.3	0.2	0.3
75.7	77.4	1.7	2.2	84.9	9.2	12.2
76.7	77.4	0.7	0.9	82.6	5.9	7.7
100.0	100.1	0.1	0.1	101.1	1.1	1.1
98.5	100.8	2.3	2.4	93.1	-5.4	5.5
100.0	90.9	-9.1	9.1	91.1	-8.9	8.9
41.1	36.8	-4.3	10.5	35.7	-5.4	13.0
57.0	69.0	12.0	21.1	65.1	8.1	14.2
70.3	85.9	15.6	22.2	86.5	16.2	23.0
74.3	79.3	5.0	6.8	83.3	9.0	12.1
100.0	101.0	1.0	1.0	101.6	1.6	1.6
100.0	102.5	2.5	2.5	93.6	-6.4	6.4
100.0	92.3	-7.7	7.7	91.5	-8.5	8.5
1.90	11.5	9.6	504.3	3.6	1.7	90.8
100.00	73.5	-26.5	26.5	73.4	-26.6	26.6
87.62	81.3	-6.3	7.2	78.4	-9.3	10.6
84.97	69.6	-15.3	18.0	71.5	-13.5	15.9
71.67	83.2	11.5	16.1	84.3	12.6	17.6
2.73	6.7	3.9	143.8	6.8	4.0	147.5
2.05	7.6	5.5	270.4	12.4	10.3	503.9
0.70	11.5	10.8	1540.3	3.6	2.9	417.9
6.49	3.9	-2.6	40.5	6.9	0.4	6.3
10.75	7.7	-3.0	27.9	12.1	1.4	12.9
0.23	-5.0	-5.2	2257.0	2.1	1.8	797.4
8.41	3.9	-4.5	54.1	6.9	-1.5	18.0
86.87	72.9	-14.0	16.1	74.9	-12.0	13.8
4.05	-5.0	-9.0	222.5	2.1	-2.0	49.0
R²	0.938			R²	0.955	
R²_{adjusted}	0.920			R²_{adjusted}	0.930	
RMSE	18.0			RMSE	12.3	

It should be also noted that the results are quite similar to those obtained (section 4.1.6) when the targeted metal loadings were used instead of measured ones while keeping all the other parameters the same. Apparently both targeted and measured metal loadings are sufficient to represent the catalytic performance.

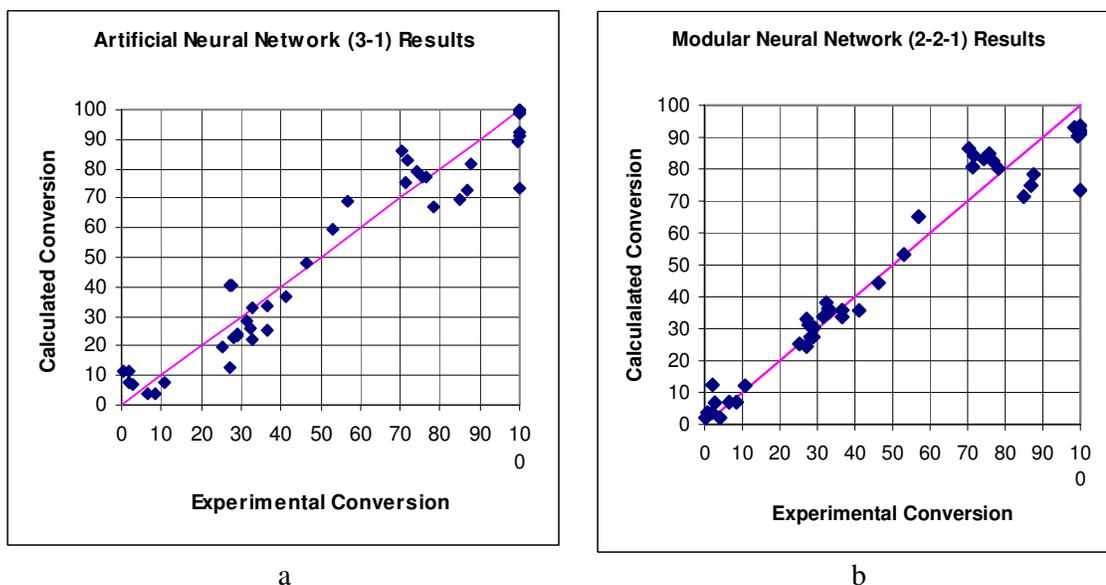


Figure 4.25. Comparison of the (a) NN 3-1 and (b) MNN 2-2-1 results with experimental data

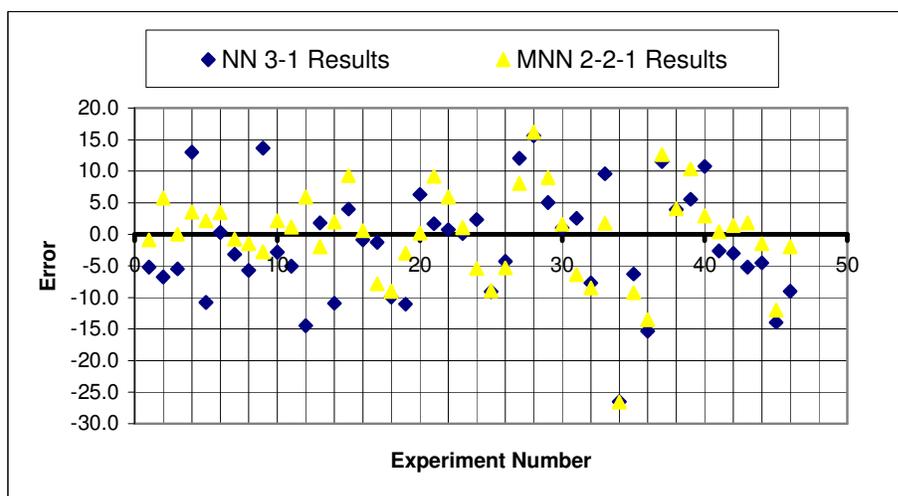


Figure 4.26. Residual Analysis for the NN 3-1 and MNN 2-2-1 models

4.2. Optimization of Artificial Neural Networks

Kibar (2005) tried to determine optimum precipitation pH and temperature at the constant reaction temperature of 120°C and W/F ratio of 1.0 mg min/ml using quadratic multiple regression equation in his work. Özdemir (2006) also studied the optimum metal loading at the reaction temperature of 150°C and W/F ratio of 1.0 mg min/ml. The optimization calculations were repeated using the modular neural networks constructed in sections 4.1.1, 4.1.3 and 4.1.6 and constrained optimization by MATLAB® Optimization Toolbox. Then the results are compared with the optimization results of the previous studies in the following sections.

4.2.1. Optimization of Precipitation pH and Temperature

The optimum values for pH and precipitation temperature are given in Table 4.14 together with the results obtained by Kibar (2005) using quadratic regression equation. Although optimum pH value of 11.5 could be considered close to (pH = 10.0) obtained value by Kibar (2005), the optimum temperature found in this case is different than 37.5 °C found by Kibar (2005). This difference could be attributed to experimental as well as modeling errors from both techniques. Predicted CO conversion value of 34.06 is the same for both model.

Table 4.14. Optimization results of the NN 2-1 for precipitation pH and temperature

Parameter	Optimum Value with NN 2-1	Optimum value with quadratic equation (Kibar, 2005)
pH	11.5	10.0
Temperature (°C)	50.8	37.5
%X	34.1	34.1

4.2.2. Optimization of Target Metal Content

The neural network model (2 neurons in the hidden layer) for the target metal content of the catalyst was also optimized using constrained optimization using Quasi-Newton method. The optimum metal loadings values are given in Table 4.15. The values calculated by Özdemir (2006) were also added for comparison, the results are in good approximation.

Table 4.15: Optimization results of the NN 2-1 for target metal contents

Parameter	Optimized Values	
	NN 2-1	Quadratic Equation (Özdemir, 2006)
wt% Cu	47.9	50.0
wt% Co	1.2	0.0
wt% Ce	50.9	50.0
%X	87.7	88.2

4.3. Input Significance Analysis of the Neural Models

The input significant analysis for both preparation and reaction conditions were performed using the “change of root mean square test” method; one input column is extracted each time and the deterioration of outputs are observed by calculating the new RMSE (Sung, 1998). Since, the significance of the input parameters can differ from network to network, the analysis was carried out for three representative networks: NN 2-1 for preparation parameters, MNN 1-1-1 for preparation and reaction conditions together and MNN 3-3-1 for all parameters together.

4.3.1. Input Significance of the Preparation Conditions

Input significances for precipitation pH and temperature were calculated with neural network 2-1. When pH column was extracted from the model, the RMSE error was calculated as 1.31, similarly for temperature column, RMSE was calculated as 2.04 indicating that the precipitation temperature has a higher impact on the results (Table 4.16). It should be noted that, the number of parameters (two) is small to have a definite conclusion about the input significance hence the results should be treated carefully.

Table 4.16. Input significance analysis for the NN 2-1 model

Extracted Input	RMSE
pH	1.31
Temperature (°C)	2.04
Average	1.68

4.3.2. Input Significance of the Preparation and Reaction Conditions

Next, the input significances of the preparation and reaction conditions, when they are used together, were also analyzed using MNN 1-1-1 network structure. The results of the significance analysis were given in Table 4.17. It is interesting to note that the parameters having similar nature have close values of significance with each other. The preparation temperature and pH have almost the same significance while the reaction

temperature and W/F ratio have higher. Although the time on stream is considered as a reaction parameter, it has a different nature than reaction temperature and W/F ratio.

Table 4.17. Input significance analysis for the MNN 1-1-1 model

Extracted Input	RMSE
pH	6.1
Temperature (°C)	6.5
Reaction Temperature (°C)	26.1
W/F	15.1
Time on Stream (h)	6.0
Average	12.0

4.3.3. Input Significance of the Design Studies and Reaction Conditions

Finally, input significance analysis covers all input parameters were done by the MNN 3-3-1 model in section 4.1.5. The results were summarized in Table 4.18.

Table 4.18. Input significance analysis for the MNN 3-3-1 model

Extracted Input	RMSE
pH	14.9
Temperature (°C)	14.3
%Cu	113.7
%Co	92.7
%Ce	82.6
Reaction Temperature (°C)	62.8
W/F	29.9
Sampling Time (h)	14.4
Average	53.2

The same group of parameters was found to have similar level significance again. Apparently the metal loadings have highest significance. This is a reasonable result since

these are the factors that have the main responsibility for the catalytic activity. The difference in the significance of these factors as well as the difference among the reaction parameters' significance can not be further due to the errors in model and experiments.

4.4. Analysis of Factor Effects

The effects of preparation and reaction parameters were analyzed using the modular neural network of 1-1-1 structure for the catalyst containing target metal content of 20 wt% of Cu, 20 wt% Co and 60 wt% Ce. The effects of metal loadings were not studied since the neural network model for this case (4.1.3) was not as successful as the others.

The change of CO conversion was calculated by changing one variable while holding the other variables constant. For example the effects of the precipitation pH was studied at the constant precipitation temperature of 37.5 °C, the reaction temperature of 120. °C, W/F ratio of 1.0 and time on stream of 3 hours (Figure 4.27.a) while the same parameters were used with the pH of 10 for the effects of precipitation temperature. (Figure 4.27.b)

For the effects of reaction temperature (Figure 4.27.c), however, W/F ratio was held constant at 2.5 mg min /ml and time on stream at 2 hours while the reaction temperature was varied between 140 - 160 °C. In the fourth graph (Figure 4.27.d), the reaction temperature was held constant at 160 °C and time on stream at 2 hours, and W/F ratio was varied between 1.0 and 2.5 mg min/ml. In the fifth graph (Figure 4.27.e) the time on stream was changed between 1 and 3 hours while the reaction temperature was kept constant at 160 °C, and W/F ratio at 2.5 mg min/ml,

The predicted factor effects are given in Figure 4.27 as solid line while the experimental data were also presented as symbols if they were available. It is clear that the model predicts the experimental data points reasonable well.

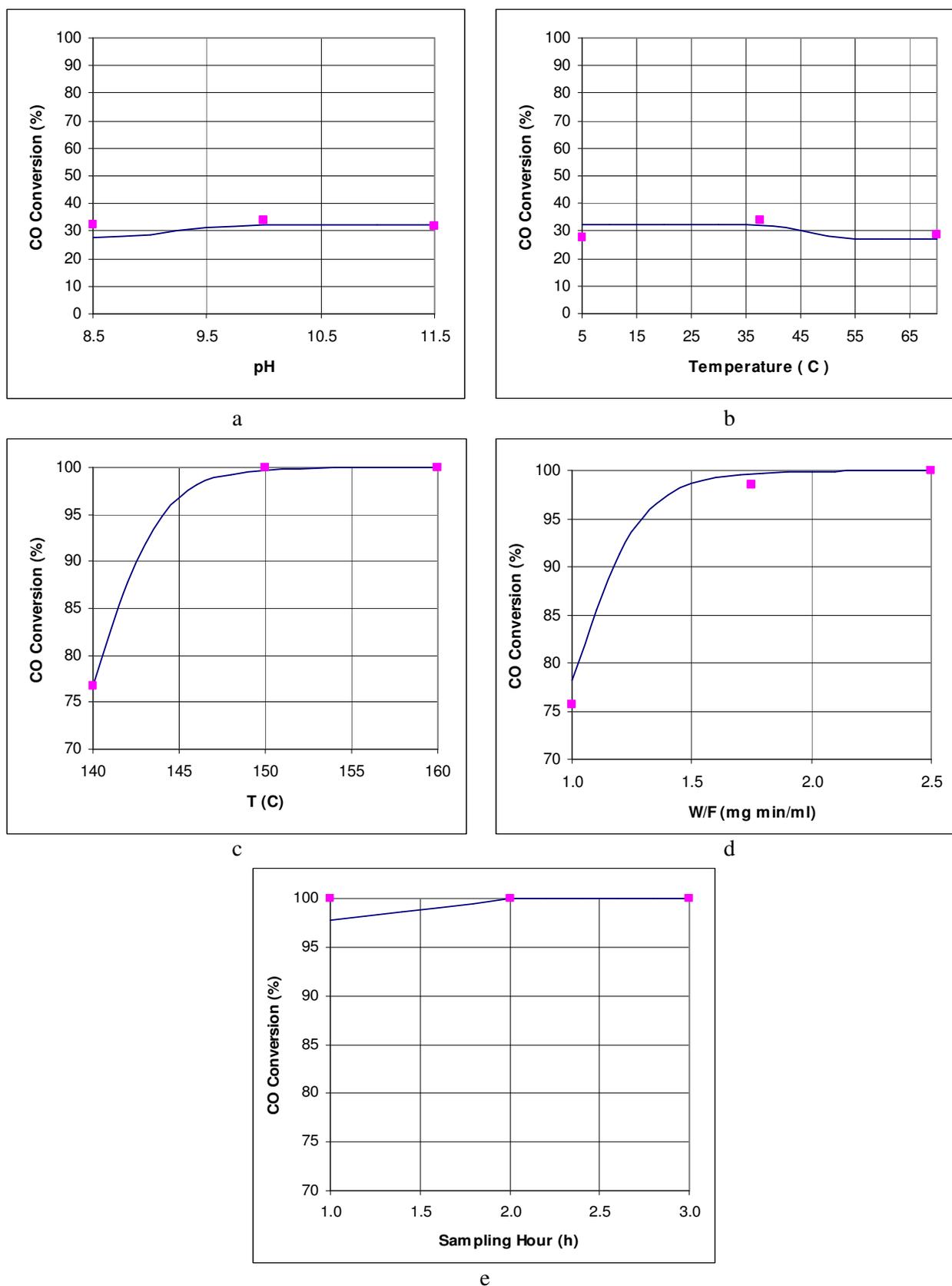


Figure 4.27. Factor effects graphs for the preparation and reaction conditions

5. CONCLUSIONS AND RECOMMENDATIONS

5.1. Conclusions

The preparation and reaction conditions of $\text{CuO}_x\text{-CoO}_x\text{-CeO}_2$ catalyst were modeled with neural and modular neural networks. Each data group was also modeled with quadratic multiple regression for comparison. Both neural and modular neural networks fit the experimental data well and much better than quadratic multiple regression. Also, in cases where there are two different groups of parameters namely, preparation and reaction conditions, the modular neural networks performed better than the neural ones. Furthermore, modular neural networks had the advantage of splitting the input parameters, hence decreasing the number of connections between the neurons.

First, a small neural and modular neural network was constructed for modeling, then the network was enlarged until a sufficient representation of experimental data obtained without over-fitting the data. The correlation parameters of R^2 , adjusted R^2 and root mean square error were used for statistical analysis. The cross validation analysis was applied to preparation and reaction conditions modeling to ensure that the models represented the entire data set successfully.

Then the preparation conditions, pH, temperature and target metal loadings were optimized using constrained optimization for maximum conversion with Quasi-Newton method. Finally the significance of the input parameters and their effects on CO conversion were analyzed.

5.1.1. Conclusions for Modeling of Preparation Conditions

The preparation conditions were modeled with a neural network of 2 neurons in the hidden layer. A modular neural network did not constructed since the parameters belonged to the same set. The R^2 obtained from the model was 0.851 which was much better than quadratic multiple regression with a value of 0.573. If the conversion belonging to

experiment 7 was extracted from the data a R^2 0.97 is obtained, which suggest that an experimental error could be associated with that point.

k- fold cross validation was applied to the model which resulted with an average root mean square error of 2.23. There was no significant deviation from the average RMSE except for the 7th experimental data point which was mentioned as a possible experimental error previously.

5.1.2. Conclusions for Modeling of Reaction and Preparation Conditions Together

The catalyst preparation conditions were modeled with the reaction parameters using a neural network of 3-1 and modular neural networks of 2-3-1 and 1-1-1 structures. The R^2 of the neural model is 0.981, the modular neural networks has a R^2 of 0.989 for 2-3-1 structure and 0.971 for 1-1-1 structure.

The success of MNN 1-1-1 network was remarkable although its structure is quite simple and number of connection (only 10 for the data set consisting 30 experimental points) is quite low indicating that modular neural network is indeed better for modeling of CO oxidation.

All the data subsets gave similar RMSE values to the average RMSE of 6.48 in k-fold cross validation of MNN 1-1-1 model.

5.1.3. Conclusions for Modeling of Catalyst Metal Content

The target metal content was modeled by a neural network of 2 neurons in the hidden layer. A R^2 of 0.825 obtained from the neural model and 0.803 from the quadratic multiple regression model. Both low correlation factors suggest that only target metal content were not sufficient to model conversion properly.

The effects of the measured metal loadings and total surface were also modeled using both neural network (NN 2-1) and modular neural network (MNN 2-1-1), which gave quite good results with R^2 of 0.988 and 0.979 respectively.

5.1.4. Conclusions for Modeling of Reaction Conditions

A neural network with 4 neurons in the hidden layer and modular neural network of 2-1-2-1 structure used to model conversion from reaction temperature, W/F ratio and sampling time. The regression parameter for NN was 0.983 and for MNN was 0.981. Since the inputs are from the same group of parameters modular approach did not provide any improvement although it as more complex structure.

5.1.5. Conclusions for Modeling of Design and Reaction Studies Together

All preparation and reaction parameters together were modeled with neural models of 3 and 4 neurons in their hidden layers, and modular neural models of 4-3-1, 3-3-1 and 2-3-1 structures. The R^2 of MNN 3-3-1 and NN 3-1 are 0.963 and 0.949 respectively. Although the modular model has 3 more neurons, it has only 6 more connections with a better fitting capability to the network.

NN 4-1 has a R^2 of 0.896, (lower than 0.949 for NN 3-1) indicating that using more neurons does not always result with a better fit. Hence, starting with a small network is the best choice.

When the measured metal contents were used in the model, a neural network with 3 neurons in the hidden layer gave a R^2 of 0.938 while the modular neural network with 2-2-1 structure had a R^2 value of 0.955. The measured metal loadings increased the performance of modular neural network by decreasing the required neuron number in the hidden layers.

5.1.6. Conclusions for Optimization of Neural Networks

The NN 2-1 model was optimized using constrained optimization with Quasi-Newton method. The optimum precipitation pH was calculated as 11.5 and temperature as 50.8 °C.

Target metal loadings of the $\text{CuO}_x\text{-CoO}_x\text{-CeO}_2$ catalyst were optimized. Optimum metal loadings for maximum conversion were calculated as 47.9 % for Cu, 1.2 % for Co and 50.9 % for Ce.

5.1.7. Conclusions for Input Significance and Factor Effects Analysis

The Input significance analysis for the precipitation parameters indicated that the temperature seems to be slightly more significant. However the number of parameters (only 2) was too small to have any definitive conclusion.

The input significance analysis for all parameters together revealed that similar parameters such as metal content, preparation parameters and reaction parameters have similar significance on the CO conversion results.

The factor effect analysis was also carried for the catalyst preparation and reaction parameters using MNN 1-1-1. The model approximated the experimental data points with good precision.

5.2. Recommendations

The recommendations for the future work on catalyst design using neural networks are stated below.

There are various algorithms as learning algorithms for backpropagation. Other algorithms such as batch training, conjugate gradient and Quasi-Newton algorithms might be used in modeling.

The data points in this thesis have been taken once an hour, the time interval might be decreased and models that have high number data points could be constructed.

Since the neural network model can relate any data points, various catalysts might be used in the same network to enlarge its structure.

Always, one should start with a small network to avoid using unnecessary neurons and prevent over-fitting.

When studying with modular neural networks, different activation functions might be used in each hidden layer. Since every function have different aspects, one could benefit from the advantage of using various functions in a network together.

REFERENCES

- Avgouropoulos G., T. Ionnides, C. Papadopoulos, J. Batista, J. Hocevar, H. Matralis, 2002, "A comparative study of Pt/ γ -Al₂O₃, Au/ α -Fe₂O₃ and CuO-CeO₂ catalysts for the selective oxidation of carbon monoxide in excess hydrogen", *Catalysis Today*, Vol. 75, pp. 157-167.
- Callan R., 1999, "The Essence of Neural Networks", Prentice Hall Europe, New York.
- Chen, Y.-Z., B.-J. Liaw, and H.-C. Chen, 2006, "Selective Oxidation of CO in Excess Hydrogen over CuO/CeZr_{1-x}O₂ Catalysts", *International Journal of Hydrogen Energy*, Vol. 31, pp. 427-435.
- Günay, M.E., 2002, *Neural Network Modeling of Low Temperature CO Oxidation Over Pt Based Catalysts*, M.S. Thesis, Boğaziçi University, İstanbul.
- Günay E., and R. Yıldırım, 2007, "Neural Network Aided Design of Pt-Co-Ce/Al₂O₃ Catalyst for Selective CO Oxidation in hydrogen-rich streams", *Chemical Engineering Journal*, doi: 10.1016/j.cej.2007.09.047.
- Haykin S., 1994, "Neural Networks: A Comprehensive Foundation", Macmillan College Publishing Company, Inc., USA.
- Hecht-Nielsen, R., 1987. *Kolmogorov's mapping neural network existence theorem*. First IEEE International Joint Conference on Neural Networks. *Institute of Electrical and Electronic Engineering San Diego, CA*, pp. 11-14.
- Igarashi, H., H. Uchida, M. Suzuki, Y. Sasaki, M. Watanabe, 1997, "Removal of Carbon Monoxide from Hydrogen Rich Fuels by Selective Oxidation over Platinum Catalyst Supported on Zeolite", *Applied Catalysis A: General*, Vol. 159, pp. 159-169.

- İnce, T., G. Uysal, A. N. Akın, R. Yıldırım, 2005, "Selective Low-Temperature CO Oxidation over Pt-Co-Ce/Al₂O₃ in Hydrogen-Rich Streams", *Applied Catalysis A: General*, Vol. 292, pp.171-176.
- Kibar, E., 2005, Investigation of Preparation Conditions of CoO_x-CuO_x-Ce-CeO₂ Catalysts for Low Temperature – Selective CO Oxidation by Experimental Design, M.S. Thesis, Kocaeli University, Kocaeli.*
- Korotkikh O., and R. Frarauto, 2000, "Selective Catalytic Oxidation of CO in H₂: Fuel Cell Applications, *Catalysis Today*, Vol. 62, pp. 249-254.
- Manasilp A., and E. Gulari, 2002, "Selective Catalytic Oxidation over Pt/alumina Catalysts for Fuel Cell Applications", *Applied Catalysis B: Environmental*, Vol. 37, pp. 17-25.
- Melin P., A. Mancilla, M. Lopez, O. Mendoza, 2007, "A Hybrid Modular Neural Network Architecture with Fuzzy Sugeno Integration for Time Series Forecasting", *Applied Soft Computing*, Vol. 7, pp. 1217-1226.
- Nandi, S., Y. Badhe, Y. Lonari, U. Sridevi, B. S. Rao, S. S. Tambe, and B. D. Kulkarni, 2004, "Hybrid Process Modeling and Optimization Strategies Integrating Neural Networks/Support Vector Regression and Genetic Algorithms: Study of benzene Isopropylation on Hbeta Catalyst", *Chemical Engineering Journal*, Vol. 97, pp. 115-129.
- Oh, S.H., R.M. Sinkevitch, 2000, "Carbon Monoxide Removal from Hydrogen – Rich Fuel Cell Feed Streams by Selective Catalytic Oxidation", *Journal of Catalysis*, Vol. 62, pp. 249-254.
- Omata K., Y. Kobayashi, and M. Yamada, 2005, "Artificial Neural Network-Aided Development of Supported Co Catalyst for Preferential Oxidation of CO in Excess Hydrogen", *Catalysis Communications*, Vol. 6, pp. 563-567.

- Özdemir, E., 2006, *Investigation of Catalytic Activity and Selectivity on CuO_x-CeO₂ Composite Catalyst Promoted with CoO_x*, M.S. Thesis, Kocaeli University, Kocaeli.
- Panzer, G., V. Modafferi, S. Candamano, A. Donato, F. Frusteri, P. L. Antonucci, 2004, "CO Selective Oxidation on Ceria-Suported Au Catalysts for Fuel Cell Applications", *J. Power Sources*, Vol. 135, pp. 177-183.
- Park J.-W., J.-H. Jeong, W.-L. Yoon, H. Jung, H.-T. Lee, D.-K. Lee, Y.-K. Park and Y.-W. Rhee, 2004, "Activity and Characterization of the CO-promoted CuO-CeO₂/γ-Al₂O₃ Catalyst for the Selective Oxidation of CO in Excess Hydrogen", *Applied Catalysis A: General*, Vol. 274, pp. 25-32.
- Perry R. H., D.W. Green, 1997, "Perry's Chemical Engineers' Handbook", McGraw-Hill, USA.
- Ren S., and X. Hong, 2007, "CO Selective Oxidation in Hydrogen-Rich Gas over Platinum Catalysts", *Fuel Processing Technology*, Vol. 88, pp. 383-386.
- Schmitz PJ, RK Usman, CR Peters, GW Graham and RW McCabe, 1993, "Effect of calcination temperature on Al₂O₃ – supported CeO₂ – complementary results from XRD and XPS", *Applied Surface Science*, Vol. 72, pp. 181-187.
- Schneider J., 1997, *Cross Validation*, <http://www.cs.cmu.edu/~schneide/tut5/node42.html>.
- Serra J. M., A. Corma, A. Chica, E. Argento, and V. Botti, 2003, "Can Artificial Neural Networks Help the Experimentation in Catalysis?", *Catalysis Today*, Vol. 81, pp. 393-403.
- Sung A.H., 1998, "Ranking Importance of input parameters of neural networks", *Expert Ststems Applications*, Vol. 15, pp. 405.

Şimşek E., Ş. Özkara, A.E. Aksoylu, and Z.İ. Önsan, 2007, “Preferential CO oxidation over activated carbon supported catalysts in H₂-rich gas streams containing CO₂ and H₂O”, *Applied Catalysis A: General*, Vol. 316, pp. 169-174.

Zou H., X. Dong, and W. Lin, 2006, “Selective CO Oxidation in Hydrogen-Rich Gas over CuO/CeO₂ Catalysts”, *Applied Surface Science*, Vol. 253, pp. 2893-2898.

Chen Y.-Z., B.-J. Liaw, H.C. Chen, 2006, Selective Oxidation of CO in Excess Hydrogen over CuO/Ce_xZr_{1-x}O₂ Catalysts”, *International Journal of Hydrogen Energy*, Vol. 31, pp. 427-435.