Journal of Natural Gas Science and Engineering 3 (2011) 518–529

Contents lists available at ScienceDirect

Journal of Natural Gas Science and Engineering

journal homepage: www.elsevier.com/locate/jngse

Application of artificial neural networks for simulation of experimental CO_2 absorption data in a packed column

A. Shahsavand*, F. Derakhshan Fard, F. Sotoudeh

Department of Chemical Engineering, Faculty of Engineering, Ferdowsi University of Mashhad, Mashhad, Islamic Republic of Iran

ARTICLE INFO

Article history:
Received 11 May 2010
Received in revised form
12 March 2011
Accepted 26 May 2011
Available online 8 July 2001

Keywords:
Absorption
Unit operations
Optimization
Packed bed
RBF
MLP

ABSTRACT

The generalization performances of the Back Propagation Multi-Layer Perceptron (BPMLP) and the Radial Basis Function (RBF) neural networks were compared together by resorting to several sets of experimental data collected from a pilot scale packed absorption column. The experimental data were obtained from an 11 cm diameter packed tower filled with 1.8 m ¼ inch ceramic Rashig rings. The column was used for separation of carbon dioxide from air using various concentrations and flow rates of Di-Ethanol Amine (DEA) and Methyl Di-Ethanol Amine (MDEA) solutions. Two in-house efficient algorithms were employed for optimal training of both neural networks. The simulation results indicated that the RBF networks can perform more adequately than the MLP networks for filtering the noise (measurement errors) and capturing the true underlying trend which is essential for a reliable generalization performance.

1. Introduction

Separation of carbon dioxide from air and various industrial gases is essential from both operational and environmental views. For example, CO_2 must be separated from natural gas to increase its heating value or carbon dioxide is usually extracted from various flue gases in beverage industry. To reduce global warming, CO_2 should be also removed from industrial flue gases before exhausting them to atmosphere.

In many practical applications, the natural gas contains around five percent carbon dioxide (e.g. Iranian Mozduran sour natural gas contains 6.5% CO_2). On the other hand, flue gases resulting from complete combustion of almost pure natural gases (e.g. Iranian Mozduran sweetened natural gas contains 98.4% CH_4) usually contain around 9% CO_2 when natural gas is burnt with 5% or more excess air. The experimental CO_2 concentrations (1–5%) were selected to cover approximately the above specifications while meeting the CO_2 flow-meter restrictions.

Although several adsorption and membrane processes are recently used for CO_2 separation purposes (Xu et al., 2005; Fauth et al., 2005; Gray et al., 2005; Park et al., 1997; Datta and Sen, 2006) absorption processes are still more popular in this area (Vaidya and Kenig, 2007; Yeh et al., 2001; Huttenhuis et al., 2007). Alkanolamines (such as DEA or MDEA) are usually used for efficient separation of carbon dioxide from various industrial gases. Packed towers usually provide higher mass transfer areas and lower pressure drops when compared to tray towers.

Lin and Shyu (1999) investigated the absorption of carbon dioxide from nitrogen using MEA and MDEA solutions in a packed column under various operating conditions. A two parameter theoretical model was presented for describing the CO_2 absorption behavior. The proposed model was validated using test data. They concluded that "an increase in the absorption load due to increased inlet CO_2 concentration or gas flow rate leads to a much shorter breakthrough time. However, an increase in the amine concentration significantly enhances the CO_2 absorption".

Sultan et al. (2002) presented a theoretical model to investigate the effect of various operational parameters on the performance of a regeneration packed column. The experimental data were then correlated to estimate the water evaporation rate from desiccant ($CaCl_2$) at various operating conditions. They concluded that "the water evaporation rate increases with increase of air and solution inlet parameters, namely, flow rate and temperature".

Brettschneider et al. (2004) used a non-equilibrium heat and mass transfer model to describe the chemical absorption of ammonia, carbon dioxide and hydrogen sulfide in an aqueous solution containing sodium hydroxide, MEA and MDEA. The chemical reaction and its influence on mass transfer in the

* Corresponding author. Tel.: +98 511 8810739; fax: +98 511 8816840.
E-mail address: shahsavand@um.ac.ir (A. Shahsavand).

1875-5100/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jngse.2011.05.001

electrolyte system were accounted by enhancement factors. The calculation of the hydraulic parameters was based on standard correlations. The predictions of the mass transfer model were validated using experimental data.

Sharma et al. (2004) employed back propagation artificial neural networks to investigate the fault diagnosis in an ammonia–water packed distillation column. The network was reported to perform satisfactorily on detection of the designated faults. The relative importance of various input variables on the output parameters was calculated by partitioning the connecting weights. The simulation results indicated that "bottoms temperature, overhead composition and overhead temperature are not much affected by the disturbances in feed rate, feed composition and vapor rate in the given range".

Liua et al. (2006) proposed a complex computational mass transfer model (CMT) for modeling of chemical absorption processes in packed columns. The model was able to consider heat effect for prediction of the concentration and temperature profiles as well as the velocity distributions. The presented model coupled the computational fluid dynamics (CFD) with computational heat transfer (CHT). The model was successfully validated using borrowed experimental data collected from a 0.1 m ID and 7 m height pilot scale tower randomly packed with ½ inch ceramic Berl saddles. The column was used for chemical absorption of CO_2 from air by aqueous monoethanolamine (MEA) solution at total pressure of 103.15 kpa. Other sets of literature data collected from an industrial-scale packed column (1.9 m ID and 26.6 m height) randomly filled with 2 inch stainless steel Pall rings were also used for validation purposes. Chemical absorption of CO_2 from natural gas was conducted using aqueous MEA solution. They argued that "the common viewpoint of assuming constant turbulent mass transfer diffusivity (D_t) throughout the entire column is questionable, even for the small size packed columns", since D_t varies along both axial and radial directions.

In this article, the generalization performances of Multi Layer Perceptron (MLP) and Radial Basis Function (RBF) artificial neural networks are compared together using a pilot-scale packed absorption column. The training data were collected by conducting several experiments on absorption of CO_2 from air using various DEA and MDEA solutions with different concentrations. Two different points will be emphasized in this article: a) Superior performance of RBF networks (when equipped with proper regularization level) to MLP networks and conventional software's for empirical modeling of the absorption process. b) Providing experimental data for absorption of CO_2 using a packed tower, which is quite limited and is essential for better analysis of the entire process.

Finally, a brief overview of neural networks and the two in-house algorithms for training RBF and MLP networks have been presented in Appendices A–C.

2. Experimental setup

A pilot scale packed column was used to collect the required experimental data. Fig. 1 represents the schematic diagram of the packed tower and the corresponding auxiliary equipments. Two separate glass columns (ID = 4.5 inches), one mounted over the other, and each one of them packed with 90 cm of ¼ inches ceramic Rashig rings were used. A liquid re-distributor was assembled between the two packed sections. Sampling points were fitted at both ends of the column for collection of pressure drop data and concentration analysis purposes.

An 80 L stainless steel solvent storage tank was used. The solvent and air rates were measured with separate glass rotameters prior to entering the tower. Large cylinders were used to provide

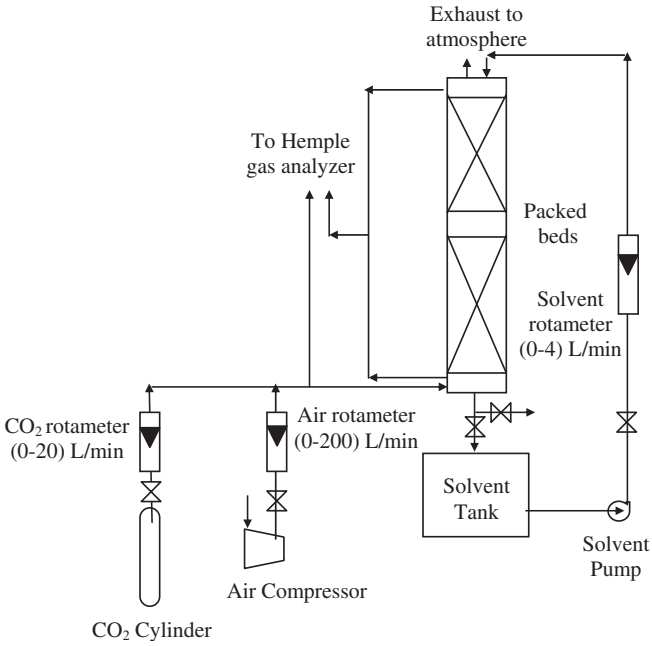

Fig. 1. Schematic diagram of the pilot scale packed absorption system.

the required carbon dioxide. The CO_2 stream was also measured via a small rotameter before mixing with air and entering the tower.

The Hempl apparatus (as shown in Fig. 2) was used to measure the carbon dioxide concentrations at inlet and outlet air streams. These measured concentrations were occasionally verified (and corrected) by titrating both inlet and outlet solvent streams and using the component mass balance around the packed sections.

3. Experimental data

After calibration of the Hempl gas analyzer apparatus, various measurements of inlet and outlet gas concentrations were performed. The following operating variables were varied during the experiments:

- Type and solvent (DEA, MDEA and pure water),
- Gas and liquid flow rates,
- Concentrations of both solvents and gas streams.

The small temperature fluctuations during each experiment were ignored and the average temperature was calculated using the initial and final conditions. The barometric pressure was close to

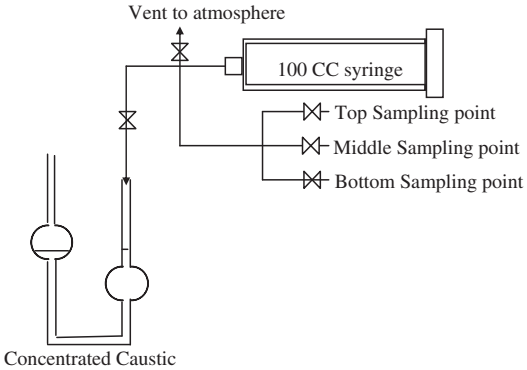

Fig. 2. Schematic diagram of the Hempl gas analyzer.

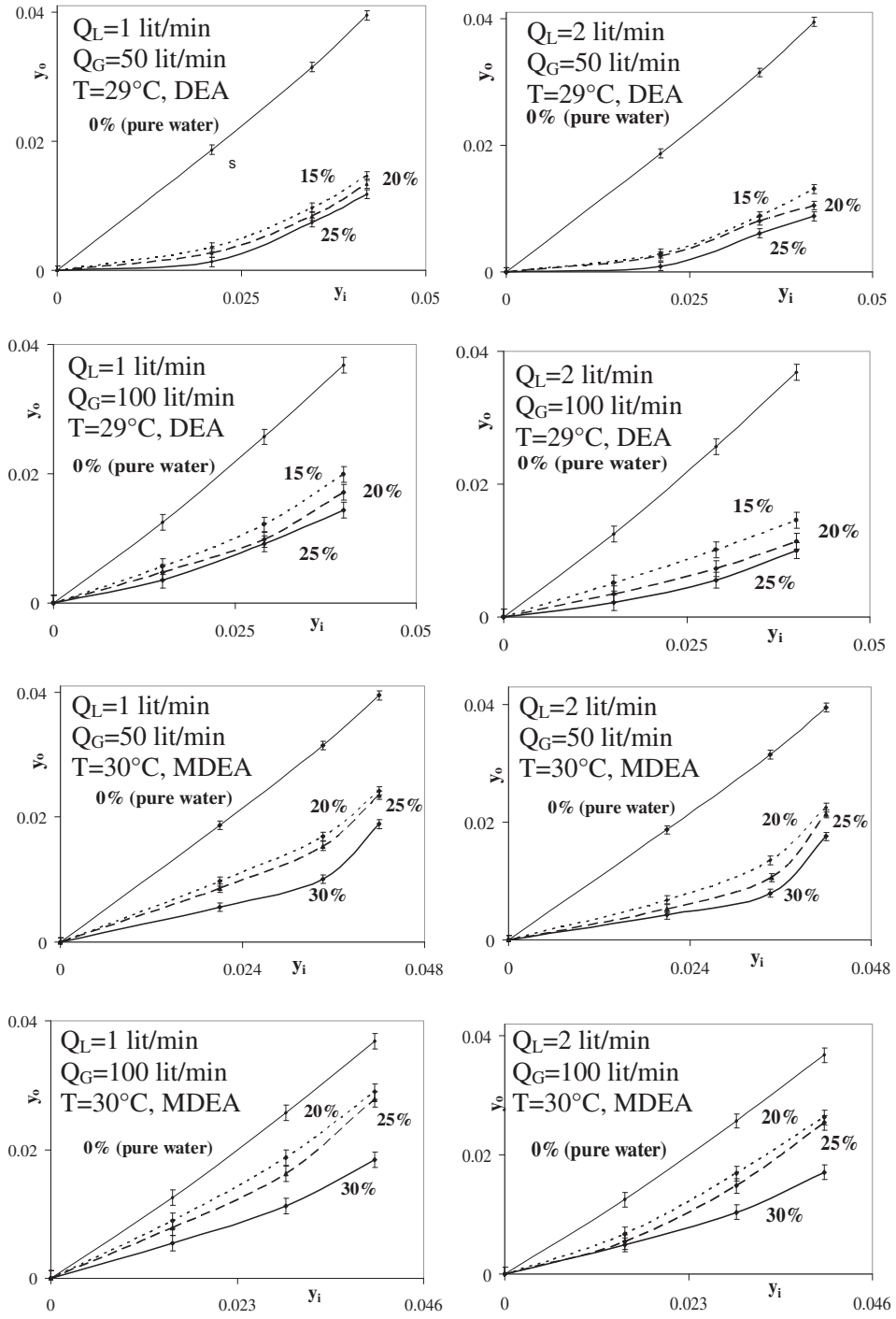

Fig. 3. Measured output gas concentrations versus mole fractions for DEA and MDEA solvents at corresponding operating conditions.

one standard atmosphere for all experiments. Figs. 3 and 4 represent the collected data for various operating conditions as specified in the corresponding figures. The error bars were computed from the following equations using multiple measurements at selected points.[1]

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \tag{1}$$

$$\sigma^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1} \tag{2}$$

For small CO_2 concentrations of entering air streams and assuming constant total gas flow rate across the entire column, the percent CO_2 absorbed from air was computed for both figures as following:

[1] The $(n-1)$ was used in the denominator of the standard deviation equation, since one degree of freedom is already used for computation of the expectation (mean) value.

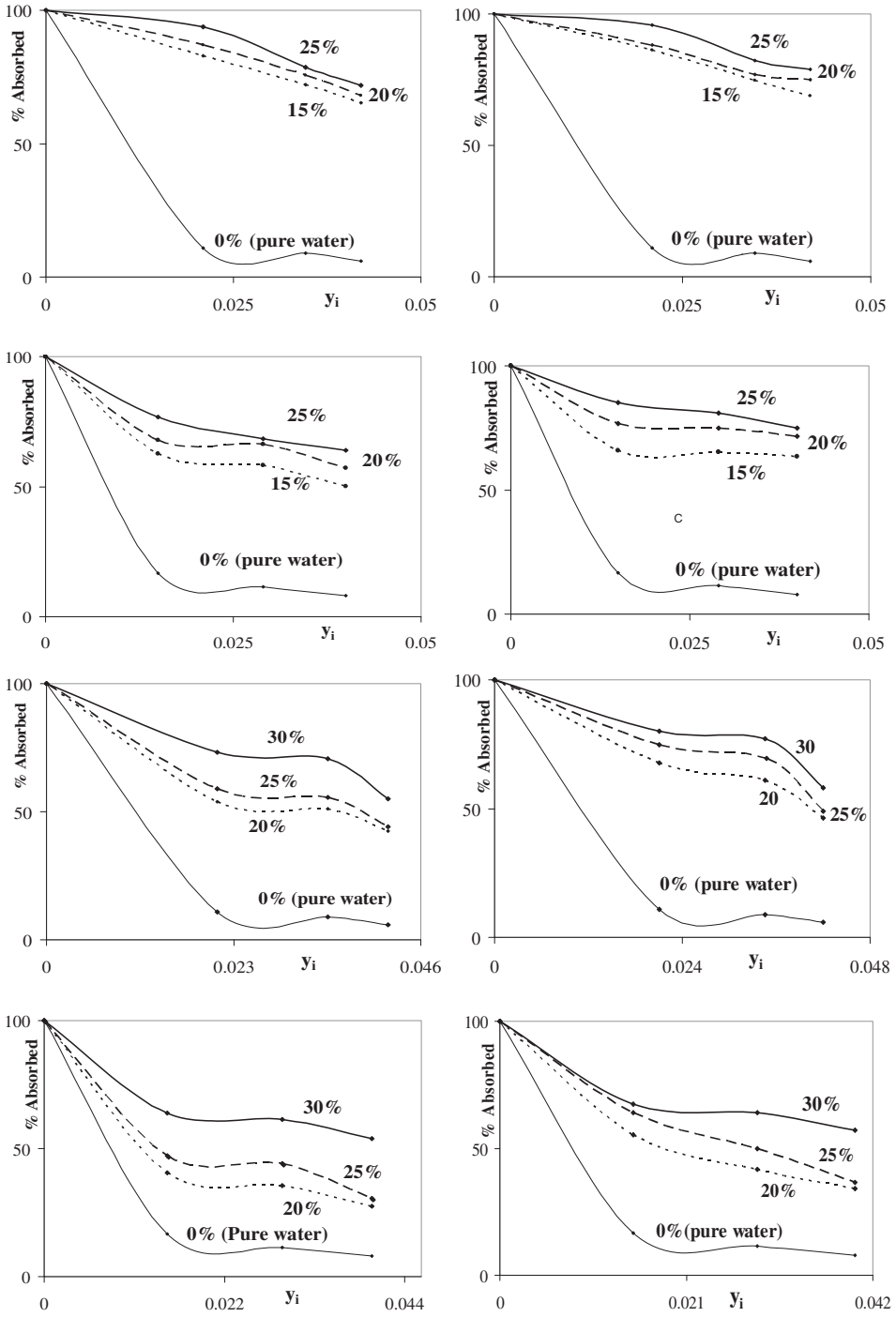

Fig. 4. Calculated values for percents of absorbed CO_2 versus input concentrations for various solvents and operating conditions mentioned in Fig. 5 respectively.

$$\%(CO_2)\text{Absorbed} = \frac{y_{\text{inlet}} - y_{\text{outlet}}}{y_{\text{inlet}}} \times 100 \quad (3)$$

Evidently, the outlet gas concentration increases by increasing the inlet CO_2 mole fraction at constant solvent concentration as shown in Fig. 3. On the other hand, increasing the CO_2 mole fraction of entering air (provided that all other conditions are kept constant) although increases the absorption rate (due to relatively larger driving forces), however it may not necessarily lead to higher percents of CO_2 absorption. Actually, in most cases, the increase in mass transfer rate is much smaller than the increase in inlet concentration due to limitations of equilibrium concentrations[2] (or limited solvent capacity). In such cases, the percent of CO_2 absorption will decrease with increase in the CO_2 mole fraction of entering air.

Evidently, the absorption capacity of the solvents increases by increasing the DEA or MDEA concentrations of the entering

[2] Note that the equilibrium constant is actually a function of solute concentration in both phases as well as the operating temperature and pressure.

solutions. Furthermore, since DEA is much stronger alkali compared to MDEA, hence, DEA solutions are able to absorb more carbon dioxide than MDEA solutions (at constant solvent concentrations).

4. HYSYS and Aspen simulation results

The pilot scale process illustrated in Fig. 1 was simulated using HYSYS and Aspen technical software. Various built in thermodynamic packages were tried during the simulations. Fig. 5 shows two predicted simulated concentration profiles across the packed section computed by HYSYS and Aspen. Table 1 provides a comparison between the measured outlet concentrations and the nearest simulation results using the most appropriate thermodynamic package for each case. Although both softwares failed to predict proper outlet concentrations, however, Aspen program provides relatively more realistic results.

Traditional simulation methods such as HYSYS or Aspen software's always use predefined models for their property estimations and they can be modified usually by introducing adequate interactions parameters for a fixed set of components. The selection of proper interaction parameters from mere experimental data can be achieved by resorting to a trial and error procedure which is very time consuming and in most cases it is almost impractical. On the other hand, the empirical models such as RBF or MLP networks are designed to use directly the measured data and provide the required non-linear multivariate interpolation as the output or response.

5. Neural network simulation results

The entire collection of experimental data presented in Figs. 3 and 4 for CO_2 outlet concentrations and percentage of absorbed CO_2 were used to train both MLP and RBF networks.

By definition, the RBF regularization network employs the same number of neurons as the data points. For better comparison of both networks performances, the number neurons of MLP networks were kept equal to the number of training exemplars.

The regularization network was completely self-sufficient and did not require any initial values for its linear and non-linear parameters. As it was mentioned in the theoretical section, these networks use all the input data as their centers. Furthermore, our in-house optimization technique (Shahsavand, 2000; Shahsavand, 2009; Shahsavand and Pourafshari Chenar, 2007; Haykin, 1999; Golub et al., 1979) was used to select the optimal value of isotropic spreads at each case, separately. The leave one out cross validation (LOOCV or CV in brief) criterion was also employed to provide the optimum level of regularization for a given set of centers and the optimal isotropic spread.

Table 1
Comparison of HYSYS and Aspen simulation results with measured CO_2 outlet concentrations for sample DEA and MDEA solvents.

L (lit/min)	G (lit/min)	Y_i	$Y_{o\ experimental}$	$Y_{o\ HYSYS}$	$\%E_{HYSYS}$	$Y_{o\ Aspen}$	$\%E_{Aspen}$
25 wt % DEA in water							
1	50	0	0	0	*0*	0	*0*
1	50	0.021	0.0013	0.0201	*1446*	0.01397	*975*
1	50	0.0345	0.0074	0.0326	*341*	0.0231	*212*
1	50	0.042	0.0118	0.0402	*241*	0.0281	*138*
1	100	0	0	0	*0*	0	*0*
1	100	0.015	0.0035	0.0144	*311*	0.01222	*249*
1	100	0.029	0.0092	0.0279	*203*	0.0237	*158*
1	100	0.040	0.0144	0.0385	*167*	0.03275	*127*
2	50	0	0	0	*0*	0	*0*
2	50	0.021	0.0009	0.0198	*2100*	0.00719	*699*
2	50	0.0345	0.0061	0.0325	*433*	0.0119	*95*
2	50	0.042	0.0088	0.0396	*350*	0.0146	*66*
2	100	0	0	0	*0*	0	*0*
2	100	0.015	0.0022	0.0143	*550*	0.00993	*351*
2	100	0.029	0.0055	0.0277	*404*	0.0193	*251*
2	100	0.040	0.010	0.0383	*283*	0.0267	*167*
30 wt % MDEA in water							
1	50	0	0	0	*0*	0	*0*
1	50	0.021	0.0056	0.0201	*259*	0.01397	*149*
1	50	0.0345	0.0101	0.033	*227*	0.0231	*129*
1	50	0.042	0.0189	0.0402	*113*	0.0281	*49*
1	100	0	0	0	*0*	0	*0*
1	100	0.015	0.0054	0.0145	*169*	0.01222	*126*
1	100	0.029	0.0112	0.0280	*150*	0.0237	*112*
1	100	0.040	0.0184	0.0386	*110*	0.03275	*78*
2	50	0	0	0	*0*	0	*0*
2	50	0.021	0.0042	0.0198	*371*	0.00719	*71*
2	50	0.0345	0.0079	0.0326	*313*	0.0119	*51*
2	50	0.042	0.0176	0.0397	*126*	0.0146	*−17*
2	100	0	0	0	*0*	0	*0*
2	100	0.015	0.0049	0.0144	*194*	0.00993	*103*
2	100	0.029	0.0104	0.0278	*167*	0.0193	*86*
2	100	0.040	0.0171	0.0383	*124*	0.0267	*56*

Italic numbers are used to distinguish between computed mole fractions and the corresponding percent errors.

In contrast, the MLP networks performances completely depended on the set of initial values selected for their synaptic weights. In this work, various initial synaptic values were tried for each training data set and then the best values were selected based on visual considerations. Evidently, the above procedure for training of MLP networks is much more time demanding than the previously described straightforward method for training of Regularization networks.

Figs. 6 and 7 compare the recall and generalization performances of optimally regularized RBF networks with the best

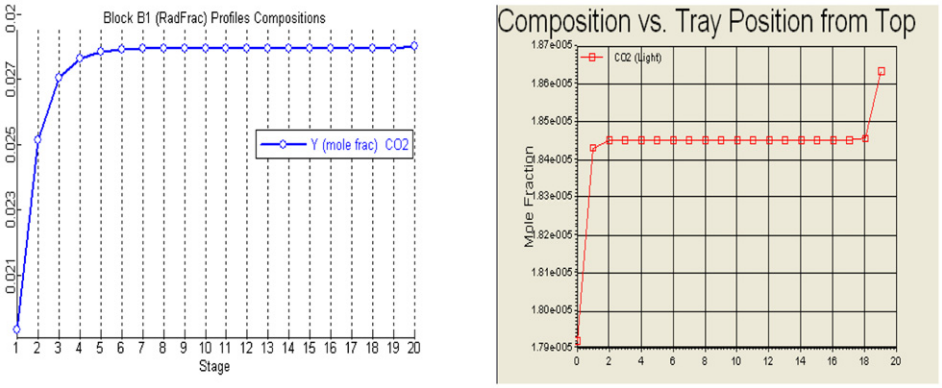

Fig. 5. Sample simulation results for packed tower concentration profile (left: Aspen, right: HYSYS).

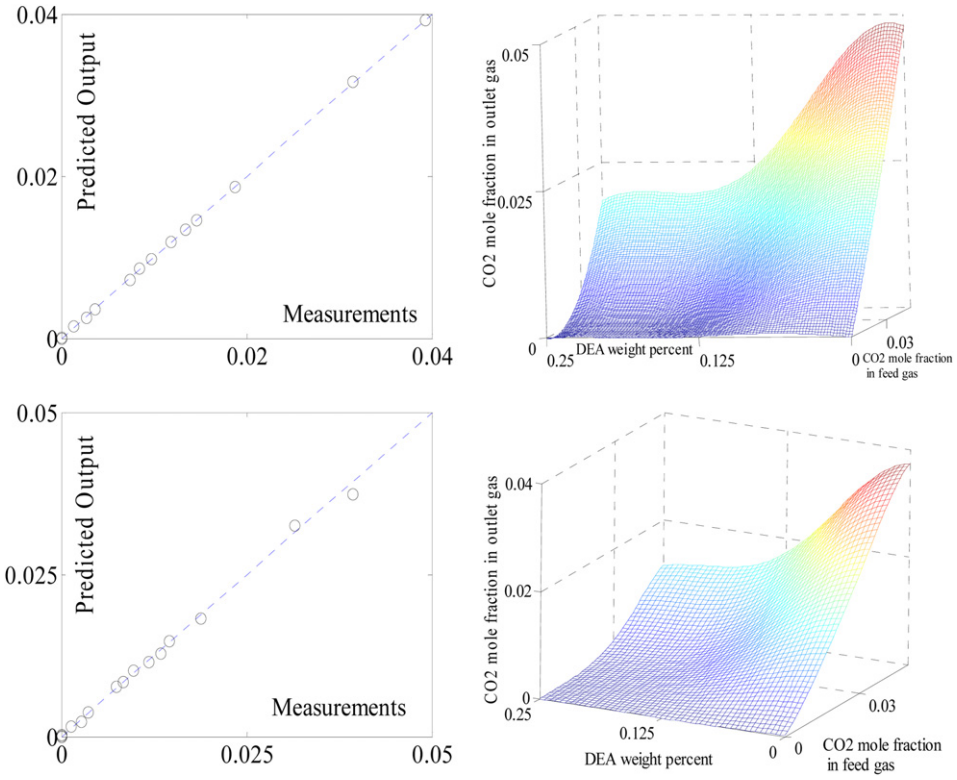

Fig. 6. Recall and generalization performances of MLP (top) and RBF (bottom) networks for gas outlet concentration using DEA solvent (Q_L = 1 lit/min, Q_G = 50 lit/min).

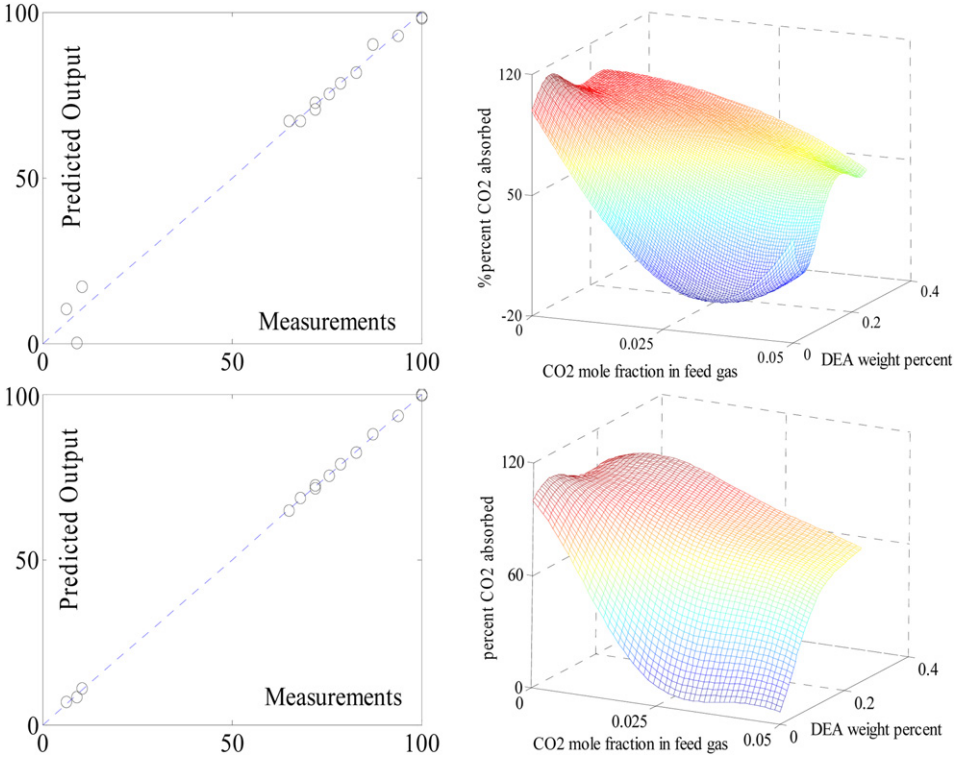

Fig. 7. Recall and generalization performances of MLP (top) and RBF (bottom) networks for percent CO_2 absorbed using DEA solvent (Q_L = 1 lit/min, Q_G = 50 lit/min).

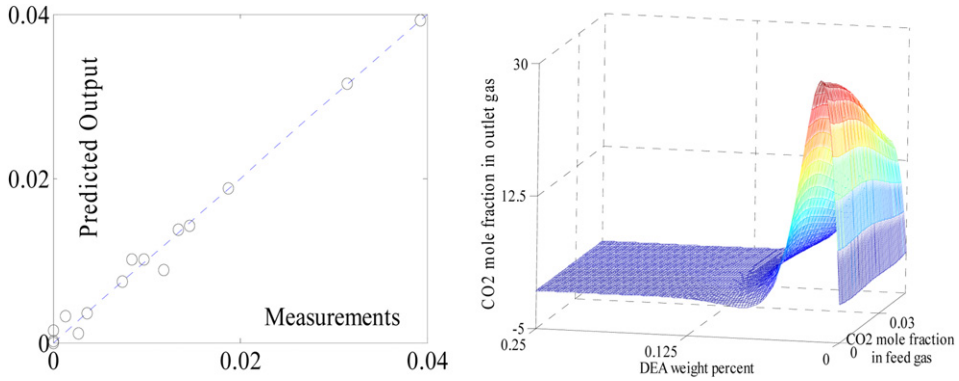

Fig. 8. Recall and generalization performances of MLP network with improper initial weights trained for gas outlet concentration using DEA solvent ($Q_L = 1$, $Q_G = 50$).

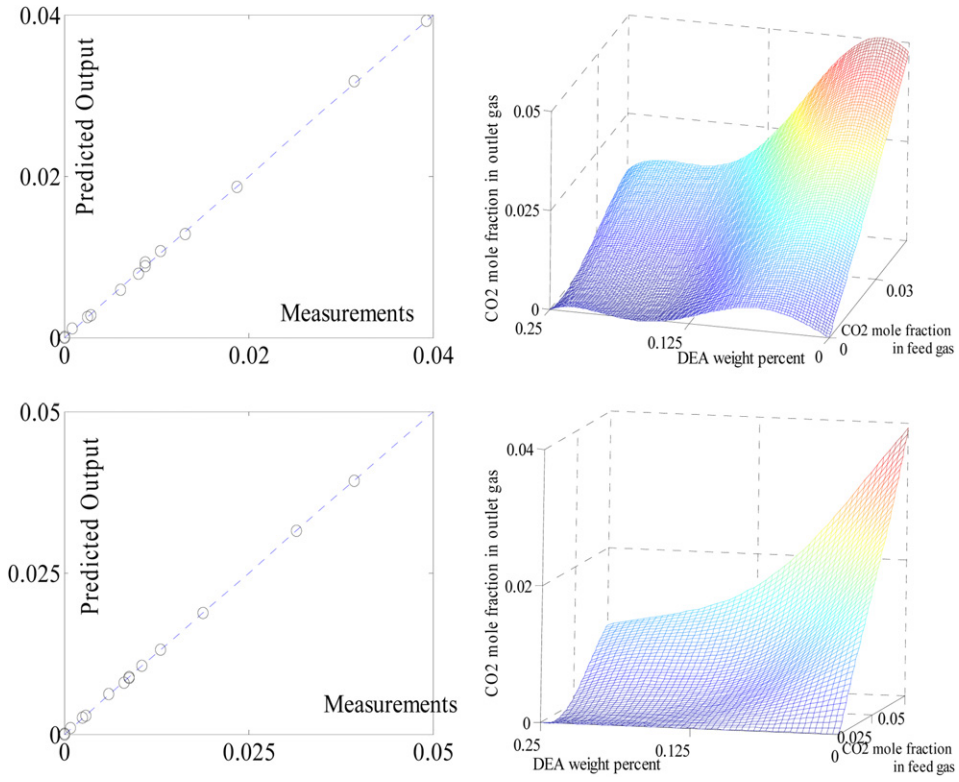

Fig. 9. Recall and generalization performances of MLP (top) and RBF (bottom) networks for gas outlet concentration using DEA solvent ($Q_L = 2$ lit/min, $Q_G = 50$ lit/min).

selected MLP network for prediction of CO_2 outlet concentrations and percents of CO_2 absorption versus CO_2 mole fractions of the entering air streams and DEA weight fractions in the entering solvents. Although both networks performs very close to each other, however, the RBF network provides better generalization performances with less oscillations due to its noise filtering capability. Furthermore, as it was mentioned above, the appropriate choice of initial synaptic weights was crucial for the proper generalization performance of MLP network. Fig. 8 shows the generalization performance of the MLP network using exactly the same training data as Fig. 7 but starting with different set of non optimal initial synaptic weights. The failed performance[3] of Fig. 8 is due to the extreme flexibility of MLP model which usually leads to the well known but undesirable over fitting phenomenon.

Figs. 9–13 compare typical recall and generalization performances of optimally trained RBF network with the best selected MLP networks (trained with proper initial weights). The experimental data collected for DEA and MDEA solvents at different liquid and gas flow rates (Figs. 3 and 4) were used for training both networks.

The above simulation results clearly demonstrate that the RBF networks provide much more reasonable and also more reliable hyper-surfaces than the corresponding MLP networks. In some cases (as in Figs. 11–13), the latter network fails drastically and leads to oscillatory generalization performances which are due to excess flexibility of the network and the so called “over-fitting” phenomenon. The superior generalization performance of the RBF network is the result of its strong theoretical backbone due to the

[3] Note the scale of the vertical axis in the 3D plot.

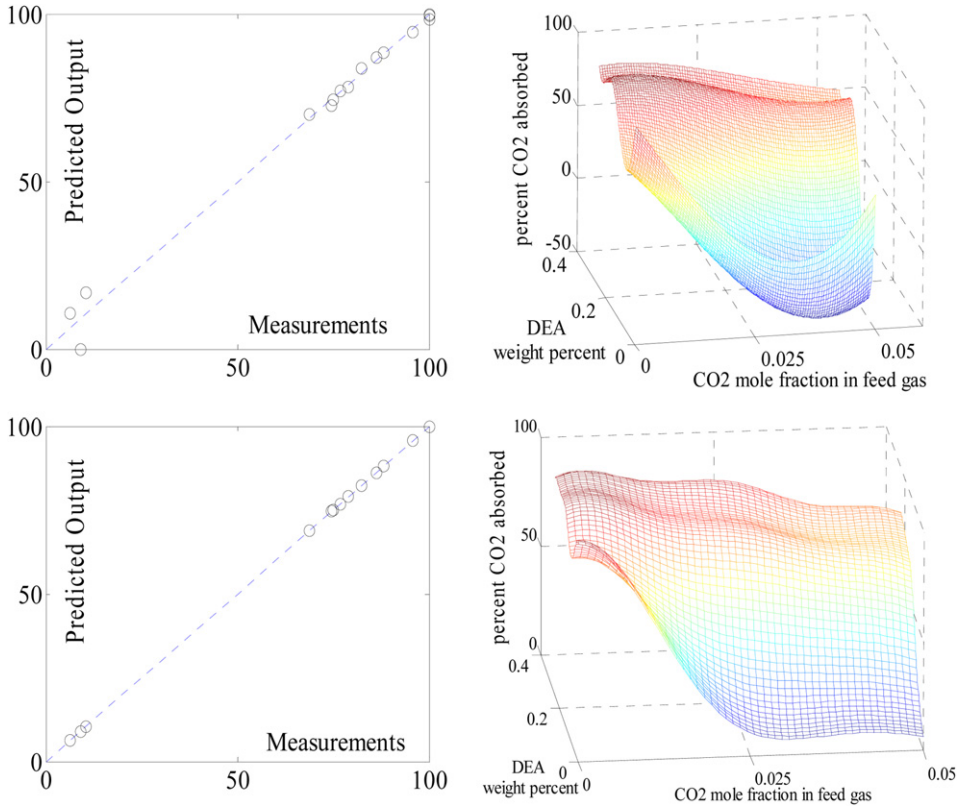

Fig. 10. Recall and generalization performances of MLP (top) and RBF (bottom) networks for percent CO_2 absorbed using DEA solvent ($Q_L = 2$ lit/min, $Q_G = 50$ lit/min).

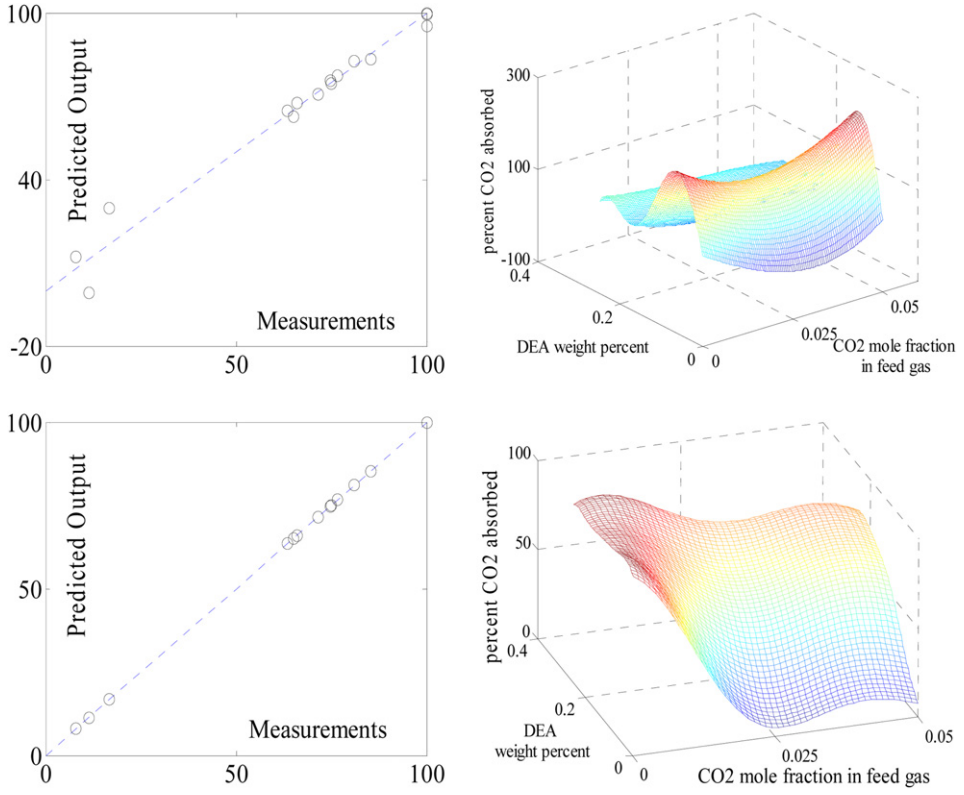

Fig. 11. Recall and generalization performances of MLP (top) and RBF (bottom) networks for percent CO_2 absorbed using DEA solvent ($Q_L = 2$ lit/min, $Q_G = 100$ lit/min).

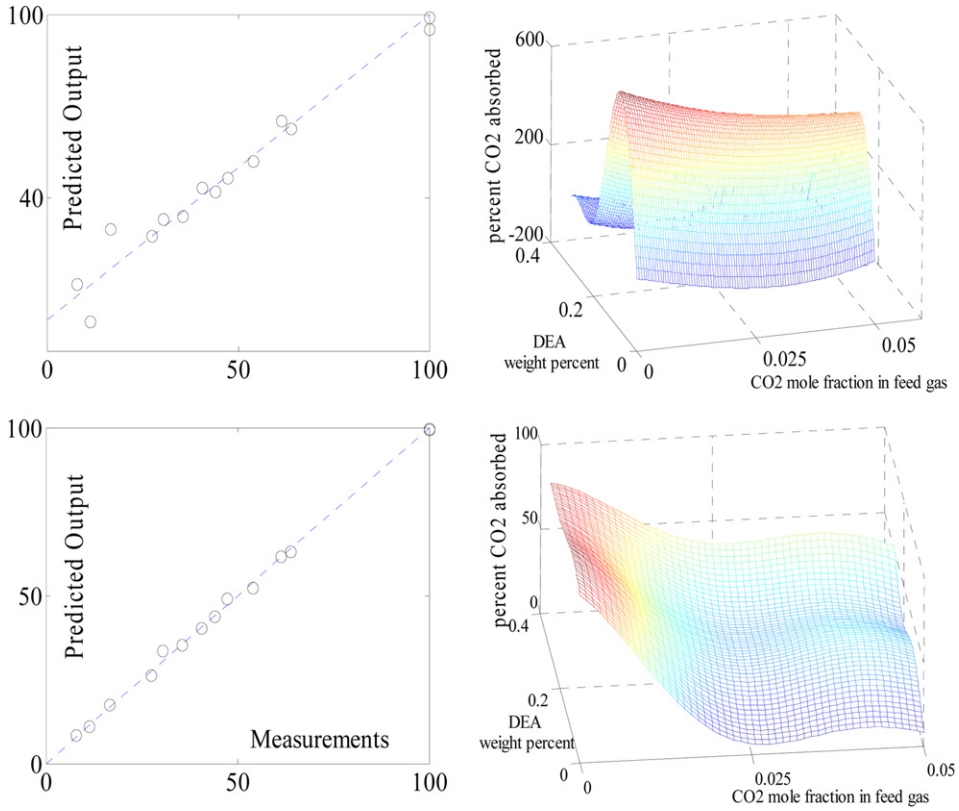

Fig. 12. Recall and generalization performances of MLP (top) and RBF (bottom) networks for percent CO_2 absorbed using MDEA solvent (Q_L = 1 lit/min, Q_G = 100 lit/min).

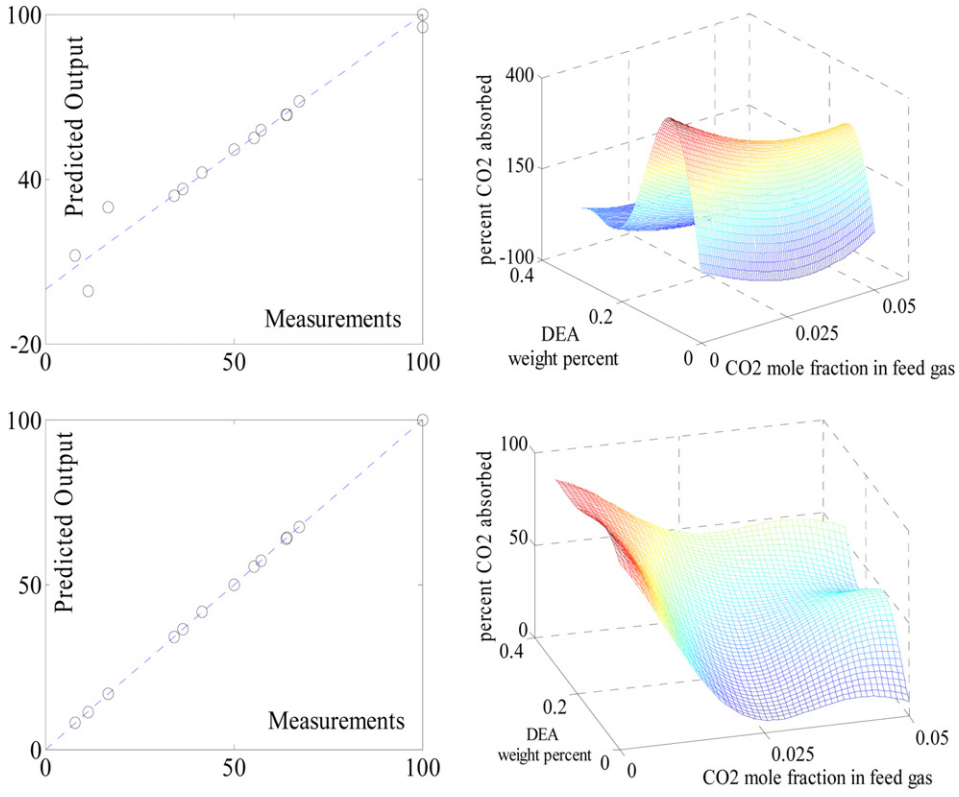

Fig. 13. Recall and generalization performances of MLP (top) and RBF (bottom) networks for percent CO_2 absorbed using MDEA solvent (Q_L = 2 lit/min, Q_G = 100 lit/min).

powerful multivariate regularization theory coupled with the efficient technique of leave one out cross validation (CV) criterion.

6. Conclusion

The experimental data collected in this article was used as an influential tool to improve our understandings of the packed absorption processes especially the absorption of CO_2 from air by various alkanolamine solutions. The simulation results presented here show that both HYSYS and Aspen software may perform inadequately for prediction of CO_2 concentrations profiles across the packed column. This failure emphasizes the complex behavior of the process.

As an alternative approach, the performances of two well known classes of artificial neural networks were compared together using the collected experimental data. It was clearly illustrated that the Regularization networks provide more reliable generalization performances compared to MLP networks. The superior performance of RBF networks was due to their solid theoretical foundations and their strong noise filtering capabilities.

Acknowledgment

The authors wish to acknowledge the valuable contribution of Khangiran gas refinery officials for providing DEA and MDEA solutions and Ferdowsi university of Mashhad for the financial support. We also appreciate the kind assistance of Mr. Mirmehrabi.

Appendix A. A brief overview of neural networks

Neural networks generally consist of several interconnected neurons in one or more *hidden* layers. They can be classified from different points of views such as the type of input transformation, their structural architecture and the type of learning algorithm. Neural networks may employ either projection or kernel based transformations to account for correlation among the inputs.

In the first transformation the inputs are projected on a single axis, the projection may be linear or non-linear. The McCulloch-Pitt neuron, Perceptron and Adaline are examples of linear projections (Park et al., 1997). In the second transformation, the norm (usually Euclidean) of the input vector with respect to a fixed point (center) is used. Radial basis function networks are the most popular examples of the kernel based input transformations (Shahsavand, 2009).

Neural networks are commonly classified based on their direction of signal flow into feed-forward and recurrent networks. In a feed-forward network, signals flow from the input layer to hidden layer(s) and then to the output layer via unidirectional connections. The neurons are connected from one layer to the next but not within the same layer. These networks can most naturally perform static mappings between input and output spaces. In other words, the output of a feed-forward network at a given instant is only a function of its input at the same instant. In general, two different classes of *single* and *multiple* layer feed-forward network architectures can be identified (Shahsavand, 2000).

In its simplest form, a feed-forward network is constructed from an *input layer* of source nodes that are projected onto an *output layer* of computation nodes via synaptic weights. The "single-layer" designation refers to the output layer containing feed-forward computation nodes (neurons). A linear associative memory is an example of a single-layer neural network. The network associates an output pattern (vector) with an input pattern (vector), and the information is stored in the network by virtue of the modifications made to the synaptic weights of the network.

Multi-layer feed-forward neural networks contain one or more *hidden layer*(*s*), whose corresponding computational nodes are called *hidden neurons* or *hidden units*. The function of the hidden neurons is to intervene between the external input and the network response. By adding one or more hidden layers, the network is enabled to extract higher order statistics (more information) by virtue of the extra set of synaptic connections and increased neural interactions. The ability of the hidden neurons to extract higher order statistics is crucial for large input dimensions. The neurons of each layer may be either *partially* or *fully* connected to the neighboring layers.

In a recurrent network, the outputs of some neurons are fedback to the neurons of the same layer or to the nodes of the preceding layers. The signals can therefore flow in both forward and backward directions. Recurrent networks have dynamic memories, that is their outputs at a given instant reflect both the current input as well as past inputs and outputs. Due to their dynamic memory, recurrent neural networks are particularly suited for control applications and dynamic simulations.

The learning algorithm of a neural network deals with the adjustment of the network parameters and usually settles to solving an unconstrained or constrained optimization problem. The model representing the neural network may be linear, non-linear or a combination of both with respect to the network parameters. The merit function characterizing the network performance may depend on the inputs alone or both the inputs and the outputs.

The former objective function leads to learning without a teacher (*unsupervised learning*) while the latter objective requires a teacher to direct the learning (*supervised learning*). The learning process of an unsupervised network does not require target (measured) output(s). Only input patterns are presented to the network during training. For a fixed architecture, the network automatically adapts the network parameters to cluster the input pattern into groups of similar features.

In many engineering applications we are concerned with the estimation of an underlying trend (or function) from a limited number of input–output data points with little or no knowledge of the form of the true function (truth). This problem is sometimes referred to as non-parametric regression, function approximation, system identification or inductive learning. In neural network parlance, it is usually called *supervised learning*. The underlying function is *learned* from the exemplars which a *teacher* supplies.

The set of examples (the *training set*) contains elements that consist of paired values of the independent (inputs) and the dependent (outputs) variables. A supervised learning algorithm adjusts the network parameters according to the differences between the measured response $y(\underline{x}_i)$ and the network outputs $\hat{y}(\underline{x}_i)$ corresponding to a given input $\underline{x}_i$. Supervised learning requires a supervisor, to provide the target signals.

Appendix B. Training of the MLP network

The basic element of a Multi Layer Perceptron (MLP) neural network is the artificial neuron which performs a simple mathematical operation on its inputs. The input of the neuron consists of the variables $x_1, x_2, \ldots, x_p$ and a threshold (or bias) term. Each of the input values is multiplied by a weight, w_i, after which the results are added with the bias term to produce z. Finally, a known activation function, φ, performs a pre-specified (non-linear) mathematical operation on the projected inputs. Various activation functions such as sigmoid or hyperbolic tangent are traditionally used for this purpose (Shahsavand and Pourafshari Chenar, 2007).

MLP networks may consist of many neurons ordered in layers. The neurons in the hidden layers perform the actual processing, while the neurons in the input and output layer merely distribute and collect the signals. Although, many hidden layers can be used, however, one hidden layer networks are more popular for practical applications due to their simple structures. Using multi layer hidden neurons usually leads to unnecessarily large degrees of freedom.

The MLP network is trained by adapting the synaptic weights using a back-propagation technique or any other optimization procedure. During training phase, the network output is compared with a desired output. The error between these two signals is used to adapt the weights. This rate of adaptation may be controlled by a learning rate. A high learning rate will make the network adapt its weights quickly, but will make it potentially unstable (Haykin, 1999). Setting the learning rate to zero, will make the network keep its weights constant. The steepest-descent optimization technique with constant step length parameters (η) was employed in this article.

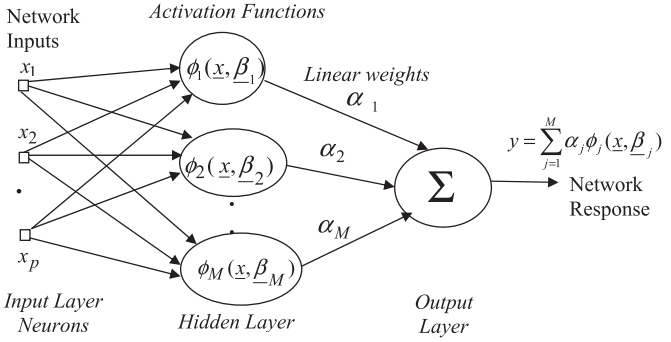

Fig. B.1. Schematic representation of a feed-forward neural network.

Additional linear weights (α's, as shown in Fig. B.1) were used in this work to accelerate the network convergence. The optimal values of these linear parameters were updated after each back-propagation iteration using the following set of linear equations:

$$\left(\Phi^T\Phi\right)\underline{\alpha} = \Phi^T\underline{y} \tag{B.1}$$

where $\Phi_{i,j} = \varphi(z_{i,j}), i = 1,\ldots,N \& j = 1,\ldots,M$ and y is the $N \times 1$ vector of measured values. The parameters N and M represent number of training data and number of neurons respectively. The training flow chart of such MLP network is given in our previous articles (Shahsavand and Pourafshari Chenar, 2007).

Similarly, other quadratic methods (e.g. Newton like techniques) may be used to compute the optimal performances of the MLP networks. The so called "line search" technique is usually used to predict the optimum learning rate compared to Newton step length[4] ($\Delta x = -G^{-1}g$). Evidently, the quadratic methods converge faster near optimal point and not necessarily far from it. This is one of the reasons why quasi Newton techniques such as Levenberg–Marquadt or Gauss–Newton are used. Finally, the steepest-descent methods (such as back-propagation) are more robust than the quadratic techniques when a proper step length control is used. In practice, almost never Newton like optimization methods are used for efficient neural network training because they can easily trap into sub-optimal solutions.

Appendix C. Training of the RBF network

The training of a projection based network (such as MLP) always reduces to the solution of a large-scale non-linear optimization problem. Such problems are usually very time demanding and often encounter severe convergence problems. In contrast, the training of RBF networks with pre-specified non-linearities (centers and spreads) reduces to the solution of an over-determined set of linear equations which can be solved by a variety of highly stable techniques.

[4] G is the Hessian matrix and g is the gradient vector.

Poggio and Girosi (1990a,b) illustrated that the solution of multivariate regularization problem can be represented as:

$$(G + \lambda I_N)\underline{w}_\lambda = \underline{y} \tag{C.1}$$

G is the $N \times N$ *symmetric* Green's matrix with elements $G_{ij} = G(\underline{x}_i, \underline{x}_j)$, and λ is the regularization parameter. In practice, we may always choose λ sufficiently large to ensure that the matrix $(G + \lambda I_N)$ is positive definite and hence invertible. Equation (C.1) can be symbolized as the *Regularization network* shown in Fig. C.1. The network consists of a single hidden layer with N neurons[5] and the activation function of the jth hidden neuron is a Green's function $G(\underline{x}, \underline{x}_j)$ centered at a particular data point $\underline{x}_j$. The influence of the regularization parameter λ is embedded in the unknown synaptic weights $w_j's$.

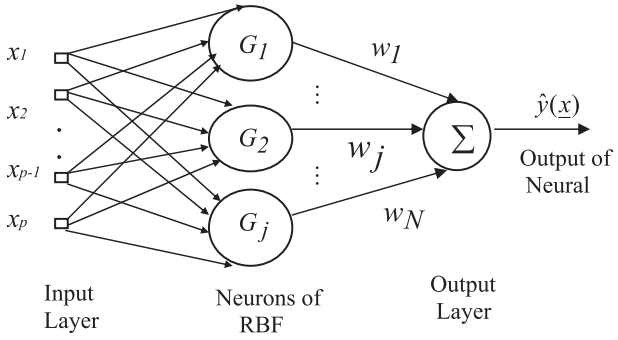

Fig. C.1. The regularization network.

Poggio and Girosi (1990a,b) also pointed out that for a special choice of stabilizing operator, the Green's function reduces to a multidimensional factorizable isotropic Gaussian basis function which is both translationally and rotationally invariant and has infinite number of continuous derivatives (Haykin, 1999).

$$G\left(\underline{x},\underline{x}_j\right) = \exp\left[-\frac{\|\underline{x}-\underline{x}_j\|^2}{2\sigma_j^2}\right] = \prod_{k=1}^{p}\exp\left[-\frac{\left(x_k - x_{j,k}\right)^2}{2\sigma_j^2}\right] \tag{C.2}$$

The σ_j appearing in equation (3) denotes the *isotropic* spread of the jth Green's function which is assumed identical for all input dimensions. The performance of the Regularization network strongly depends both on the appropriate choice of the isotropic spread and the proper level of regularization (Sugiyama and Ogawa, 2002). The leave one out cross validation criterion (Golub and Van Loan, 1996) is used here for efficient computation of the optimum regularization parameter λ^* for a given spread (σ).

$$\mathrm{CV}(\lambda) = \frac{1}{N}\sum_{k=1}^{N}\left[\frac{\underline{e}_k^T(I_N - H(\lambda))\underline{y}}{\underline{e}_k^T(I_N - H(\lambda))\underline{e}_k}\right]^2 \tag{C.3}$$

where, N represents both the number of training exemplars and the number of Regularization network neurons, $\underline{e}_k$ is the kth unit vector of size N, I_N is the $N \times N$ unit matrix and $H(\lambda)$ will be the smoother matrix originally defined by Hastie and Tibshirani[6] (Hastie and Tibshirani, 1990).

By definition, the Regularization network is a linear smoother as developed by Poggio and Girosi (1990a,b). The generalization

[5] As before, N is the number of measurements (or training exemplars).

[6] "If we focus on the fit at the observed data points $x_1, x_2, \ldots, x_N$, a linear smoother can be written as $f = Sy$; where $S = S_{i,j}$ is a $N \times N$ smoother matrix." (Hastie and Tibshirani, 1990).

performance of such network can be simply computed from $f = G\underline{w}_\lambda$. Replacing $\underline{w}_\lambda$ from equation (C.1) results:

$$f = G(G + \lambda I)^{-1}\,\underline{y} \tag{C.4}$$

Therefore, as defined in footnote 2, the smoother matrix of the Regularization network $H(\lambda)$ can be computed as:

$$H(\lambda) = S = G(G + \lambda I)^{-1} \tag{C.5}$$

The effective number of parameters or the approximate degrees of freedom (df) of a linear smoother is defined as the trace[7] of the smoother matrix. $df = \mathrm{tr}(H(\lambda))$. Evidently, the number of degrees of freedom is a function of the span[8] and the predictor values in the data set. It is not a function of the response (y) (Hastie and Tibshirani, 1990; Shahsavand and Ahmadpour, 2005).

The evaluation of $H(\lambda)$, and hence CV(λ), at each trial value of λ requires the inversion of the $N \times N$ matrix $(G + \lambda I)$ and may prove too time consuming. This can be avoided by resorting to the similarity transformation technique (Shahsavand, 2000). The basic idea was initially presented by Golub et al. (1979) for ridge regression.

Equations (C.2)–(C.4) show that CV(λ) is a complex function of both λ and σ. Therefore, the optimal value of the regularization parameter, λ^* (which minimizes CV(λ)), is highly correlated with the value of isotropic spread, σ. In other words, the appropriate value of λ^* greatly depends on σ for a specific data set with fixed level of noise.

The strong correlation between these two parameters (λ^* and σ) is extremely complicated and may not be described directly in analytical form. A simple indirect procedure is proposed in our previous articles (Shahsavand, 2000; Shahsavand, 2009; Shahsavand and Pourafshari Chenar, 2007; Haykin, 1999; Golub et al., 1979) for decoupling of such powerful dependency and finding the optimum value of spread and the corresponding optimal level of regularization for the given noisy data set. The motivation of such decoupling procedure is to train the best optimal network for reconstructing the true hyper-surface embedded in a bunch of noisy data set. As it was clearly shown, such an optimum trained network can successfully filter out the noise and provide a reliable generalization performance.

References

Brettschneider, O., Thiele, R., Faber, R., Thielert, H., Woznya, G., 2004. Separation and Purification Technology 39, 139–159.

Datta, A.K., Sen, P.K., 2006. Optimization of membrane unit for removing carbon dioxide from natural gas. Journal of Membrane Science 283, 291–300.

Fauth, D.J., Frommell, E.A., Hoffman, J.S., Reasbeck, R.P., Pennline, H.W., 2005. Eutectic salt promoted lithium zirconate: novel high temperature sorbent for CO_2 capture. Fuel Processing Technology 86, 1503–1521.

Golub, G.H., Van Loan, C.G., 1996. Matrix Computations, third ed. Johns Hopkins University Press, Baltimore.

Golub, G.H., Heath, M., Wahba, G., 1979. Generalized cross validation as a method for choosing a good ridge parameter. Techno Metrics 21 (2), 215.

Gray, M.L., Soong, Y., Champagne, K.J., Pennline, H., Baltrus, J.P., Stevens Jr., R.W., Khatri, R., Chuang, S.S.C., Filburn, T., 2005. Improved immobilized carbon dioxide capture sorbents. Fuel Processing Technology 86, 1449–1455.

Hastie, T.J., Tibshirani, R.J., 1990. Generalized Additive Models, first ed. Chapman and Hall, London.

Haykin, S., 1999. Neural Networks: A Comprehensive Foundation, second ed. Prentice Hall, New Jersey.

Huttenhuis, P.J.G., Agrawal, N.J., Hogendoorn, J.A., Versteeg, G.F., 2007. Gas solubility of H_2S and CO_2 in aqueous solution of N-methyldiethanolamine. Journal of Petroleum Science and Engineering 55, 122–134.

Lin, Sh.H., Shyu, Ch.T., 1999. Performance characteristics and modeling of carbon dioxide absorption by amines in a packed column. Waste Management 19, 255–262.

Liua, G.B., Yua, K.T., Yuana, X.G., Liua, C.J., Guob, Q.C., 2006. Simulations of chemical absorption in pilot-scale and industrial-scale packed columns by computational mass transfer. Chemical Engineering Science 61, 6511–6529.

Park, S.W., Suh, D.S., Hwang, K.S., Kumazawa, H., 1997. Gas absorption of carbon dioxide in a hollow fiber contained liquid membrane absorber. Korean Journal of Chemical Engineering 14 (4), 285–291.

Poggio, T., Girosi, F., 1990a. Regularization algorithms for learning that are equivalent to multilayer networks. Science 247, 978.

Poggio, T., Girosi, F., 1990b. Networks for approximation and learning. Proceedings of the IEEE 78, 1481.

Shahsavand, A., Ahmadpour, A., 2005. Application of optimal RBF neural networks for optimization and characterization of porous materials. Computers and Chemical Engineering 29, 2134–2143.

Shahsavand, A., Pourafshari Chenar, M., 2007. Neural networks modeling of hollow fiber membrane processes. Journal of Membrane Science 297, 59–73.

Shahsavand, A., 2000. Optimal and adaptive radial basis function neural networks, Ph.D. Thesis, University of Surrey, UK.

Shahsavand, A., 2009. An optimal radial basis function (RBF) neural network for hyper-surface reconstruction. Scintia Iranica 16 (1), 41–53.

Sharma, R., Singh, K., Singhal, D., Ghoshd, R., 2004. Neural network applications for detecting process faults in packed towers. Chemical Engineering and Processing 43, 841–847.

Sugiyama, M., Ogawa, H., 2002. Optimal design of regularization term and regularization parameter by subspace information criterion. Neural Networks 15 (3), 349.

Sultan, G.I., Hamed Ahmed, M., Sultan, A.A., 2002. The effect of inlet parameters on the performance of packed tower-regenerator. Renewable Energy 26, 271–283.

Vaidya, P.D., Kenig, E.Y., 2007. Absorption of CO_2 in to aqueous blends of alkanolamines prepared from renewable resources. Chemical Engineering Science 62, 7344–7350.

Xu, X., Song, Ch., Miller, B.G., Scaroni, A.W., Xu, 2005. Adsorption separation of carbon dioxide from flue gas of natural gas-fired boiler by a novel nanoporous 'molecular basket' adsorbent. Fuel Processing Technology 86 (14–15), 285.

Yeh, J.T., Pennline, H.W., Resnik, K.P., 2001. Study of CO_2 absorption and desorption in a packed column. Energy and Fuels 15, 274–278.

[7] The trace of a matrix equals to the sum of its eigenvalues or sum of its diagonal elements.

[8] Domain of inputs.