



Selective combination in multiple neural networks prediction using independent component regression approach

See Lee Foon¹, Nazira Anisa Rahim², Ahmad Zainal³, Zhang Jie⁴

^{1,3}School of Chemical Engineering, Engineering Campus, Universiti Sains Malaysia, Seri Ampangan, 14300 Nibong Tebal, Seberang Perai Selatan, Pulau Pinang, Malaysia. ²River Basin Research Centre, National Hydraulic Research Institute of Malaysia, Lot 5377, Jalan Putra Permai, 43300 Seri Kembangan, Selangor, Malaysia. ⁴School of Chemical Engineering and Advanced Materials, University of Newcastle, Newcastle upon Tyne NE1 7RU, UK. *Article Info: Submitted on March 20, 2017, Accepted on June 20, 2017.*

Abstract: Biological processes are highly nonlinear in nature and difficult to represent accurately by simple mathematical models. However, this problem can be solved by using neural network. Neural network is a prominent modeling tool especially when it comes to intricate process such as biological process. In this paper, a multiple single hidden layer with ten hidden neurons Feedforward Artificial Neural Network (FANN) was used to model the complex and dynamic relationships between the input (dilution rate, D) and outputs (conversion, y and dimensionless temperature value, θ) for the reactive biological process. Levenberg-Marquardt Backpropagation training method was used. The multiple neural networks predicted outputs were then combined through three different methods which are simple averaging, Principal Component Regression (PCR) and Independent Component Regression (ICR). Multiple neural networks which were created by the bootstrap approach help improved single neural network performance as well as the model robustness for nonlinear process modeling. Comparison was made between the three methods. The result showed that ICR is slightly superior between the three methods especially in noise level 1, 2 and 3, however ICR slightly suffer in noise level 4 and 5. This is due to the independent component regression used the latent factors and non-Gaussian distribution of y and θ values for the combination.

Keywords: Independent component regression; Multiple neural network; Principle component regression; Reactive biological process.

Introduction

Biological processes are nondeterministic systems which mean that they are both time variant and nonlinear in nature. This may be due to the fact that the microbial continuously undergoing physiological and morphological changes in the bioreactor and this makes it become a complex model which is difficult to validate. The nonlinear nature of biological processes may be due to the kinetics of cellular growth and product formation, thermodynamic limitations of the process, mass transfer and heat transfer effects, and the likes. Besides, biological processes in real situations are difficult to describe accurately by simple mathematical models. The term "simple" here is important because a model that follows the complex intra- and inter-cellular processes sufficiently accurately under different conditions may itself be too complex for easy computation and use in automation. Therefore, it is important to generate models based on on-line measured variables, as these can be used in process control and on-line optimization of biological systems. In order to avoid hardware limitations, Artificial Neural Network (ANN) is a prominent class of software that is very suitable to use. The neural

networks, which offer a data-driven modeling approach, are a viable modeling paradigm that can be used for that purpose, having good generalization and prediction capabilities.^{1,2}

As we all know, Single Neural Network (SNN) has been dominating the neural network "world". Even though there are lots of advantages of SNN but there are also some limitations for SNN as well as some problems that can deteriorate neural network performance. These include poor generalization of the networks because of fitting the noise in the training data, the trained network may not minimize the error on the training data set because of the presence of uncontrolled and excess dynamics or because the training data themselves are corrupted with noise. Last but not least, SNN is small and may be restricted to less complicated process.³ Therefore, a new technique, which is Multiple Neural Networks (MNN) that this paper is going to focus on, is implemented to improve the neural networks performance as well as the model robustness for nonlinear process modeling.^{4,5} One of the common methods to develop MNN is by employing combination method. Principal Component

Analysis (PCA) and Independent Component Analysis (ICA) both are data driven tools that had been applied to different benchmarks as process monitoring methods.⁶ Recent studies⁷⁻¹² shown that ICA has drawn significant attention on the prediction modeling and found potential application as a further development of principle component analysis. Motivated by the appealing role of ICA, independent component regression (ICR) was naturally developed as a combination of a first-step ICA implementation on process measurement and a second step quantitative regression modeling.¹³ While Principal Component Regression (PCR) is combination methods that used to determine the best process outputs as well as to improve the prediction accuracy of any process outputs.¹⁴

The main difference between PCA and ICA is that the components of PCA that is PCs are still dependent while components in ICA that is ICs are no orthogonality constraint and totally independent. Furthermore, PCA monitors process often supposed that PCs is Gaussian distribution whereas ICs in ICA are non-Gaussian distribution. In term of algorithm, ICA algorithm is different with the PCA algorithm at which ICA utilizes Eigen Value decomposition while PCA is singular value decomposition. When it comes to the regression, ICR algorithm is very similar to that of PCR. The only difference is to use ICs and I coefficient matrix obtained by ICA in the regression operation, instead of PCs and score matrix obtained by PCA.¹⁵ Hence, this work focus on solving the exothermic bio catalytic reaction using ICR combination method and the results are compared with PCR and simple averaging combination methods

Materials and Methods

Case study: Exothermic biocatalyics reaction

Bio catalytic reaction is a type of reaction which uses enzyme or whole-cell as a (bio)-catalyst to achieve a desired conversion, under controlled conditions in a bioreactor. The exothermic bio catalytic reaction takes place in a continuous-stirred-tank reactor (CSTR). Substrate A was converted into the product B through an enzymatic reaction. In this case, one substrate one product is considered. Therefore, the dynamic equations (non-linear differential equations) which are derived from the Michaelis-Menten and Arrhenius equations, by performing the mass and energy balances on the CSTR reactor can be described as shown in¹⁶. The model was developed based on the steady state condition before step test were introduce to generate the data for modeling.

Feedforward artificial neural network (FANN) model development

Most common biological models are constructed from first principles by applying theories of mass and energy conservation, and incorporating into the framework of thermodynamic, kinetic, heat, and mass transfer correlations.¹ Process variables of this study case are dilution rate, D , dimensionless temperature value, θ , conversion, y , and the likes. After developing the mechanistic model, the value of the independent variable (input) that is dilution rate, D , is varied from $D=10$ to $90s^{-1}$. Besides, some noises are added to mimic the real situation of the biological processes which are not ideal and perfect in nature. In order words, noises are added to mimic the measurement noise in on-line sampling process. The noise was label from level 1 to level 5 as shown in Table 1. The higher the noise power the higher the amplitude as well as the deviation. In addition to that, to show the complex correlation while increasing the noise, Table 2 shows the reduction of correlation coefficient between dilution rate and dimensionless conversion, y and it's more difficult to model.

Table 1: Noise level in dilution rate, D .

Noise level	Noise Power
1	0.2
2	0.4
3	0.6
4	1.0
5	1.5

Table 2: Correlation coefficient between the dilution rate, D and dimensionless conversion, y .

Noise level	Correlation coefficient
1	0.9919
2	0.9918
3	0.9908
4	0.9902
5	0.9886

Among all the neural networks paradigms available, a feedforward multilayer perceptron with a single hidden layer is considered to be the best choice for this thesis. The activation functions chosen are the log-sigmoid transfer function in the hidden layer and the linear function in the output layer. The log-sigmoid transfer function is commonly used in multilayer networks that are trained using the backpropagation algorithm, in part because this function is differentiable. While the linear function in the output layer is chose because; when a regression problem is tackled by a neural network all the neurons in the output layer must have a linear activation function. This allows the range of the network outputs to be outside the range $[0; +1]$ or $[-1;$

+1] of the sigmoid function. Moreover, there is an advantage by using sigmoidal-type transfer functions in the hidden layer and linear transfer functions in the output layer. It is obvious when it is necessary to extrapolate beyond the range of training data.¹⁷ Levenberg-Marquardt Backpropagation training method is used in this paper to train the network. This FANN work as single input single output (SISO) model for two networks where the input is the dilution rate while the output are dimensionless temperature value, θ , and the other network is dilution rate as the input and the output is the conversion, y .

Multiple neural networks (MNN)

Basically single neural network model exhibits lot of disadvantages. These include lack of robustness, under fitting and over fitting of the model as mention as before. Therefore, many users are now switching to use multiple neural networks. There are two major types of multiple neural networks. The first category is multiple models neural networks in which the training data are totally different in building the individual networks which can be built using different inputs in different regions of operation. The second category is to creating multiple models using the same training data which are re-samples or partitioned using specific algorithms. There are three main algorithms being used to resample or partition the training data, which are bagging or bootstrap re-sampling, adaboost and randomization. The motivation for creating those different inputs or partitions is to create the effective networks ensembles.³ The diagram of combination of multiple neural networks is shown in Fig. 1. As mentioned earlier, biological processes are often associated with complicated process. Some of them may require a long reaction time.

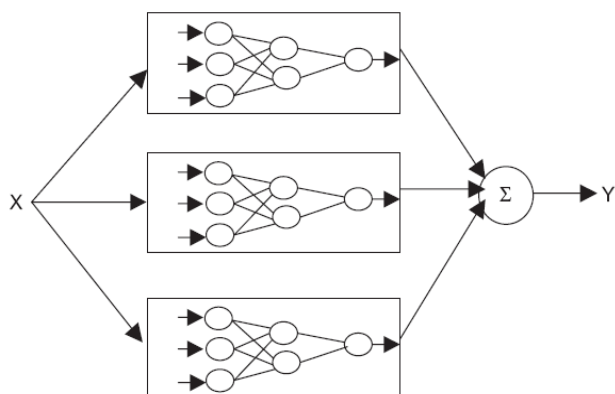


Figure 1: Multiple neural network.¹⁸

With time constraint, lack of data can be one of the possible issues in modelling biological processes. Rather than solely depended on multiple neural network model to deliver its best results, another way to improve the model is to resample the original. This can be done by applying bootstrap re-sampling technique to resample the original data. Bootstrap re-sampling method is the best way to cope with problems such as lack of data as it can re-sample the original data into multiple sets of data without comprising with the integrity of the data.

In neural network, bootstrap basically refers to replication of a training data set through re-sampling the original training data set. Some of the data samples may occur several times, and others may not occur in the sample at all. The individual training sets are independent and the neural networks can be trained in parallel. Combining multiple neural networks trained on bootstrap re-sampled data does actually increase the robustness of the model. Bootstrap technique also can generate diverse networks when the base learning algorithm is unstable in those small changes in the training data set which will cause large changes in the learned classifiers while boosting can result in less instability. The analogy of bootstrap re-sampling technique can be found in³.

Combination of multiple neural networks (MNN)

After bootstrapping the model, combination of the multiple neural networks is next step in the model development procedures. There is linear and non-linear combination of the multiple neural networks. Besides, all users can either selectively chooses to combine the useful neural networks only or even can just combine all the neural networks.⁵ In this paper, the focus is on selective combination of multiple neural networks by using PCR, ICR as well as simple averaging method. Combining of the networks improves the generalization capability of the neural networks models in such a way that it guards against the failure of individual components networks. This may due to certain neural networks fail to deliver the result or output prediction because of limited training data set or over fitting of the training data. In other words, combining a set of imperfect estimators (networks) can be thought of as a way of managing the recognized limitation of the individual estimators, each component is known to make errors, but they are combined in such a way as to minimize the effect of this errors.^{3,4}

Independent component regression (ICR) combination

ICA is a novel feature extraction technique that aims at recovering independent sources from their mixtures, without knowing the mixing procedure or any specific

knowledge of the sources.¹⁹ The independent sources, called independent components (ICs), are hidden information of the observable data. ICA model can be classified as linear ICA (LICA) and nonlinear ICA (NLICA) models.

$$X=AS$$

ICR is the main combination method that is focusing on this work. This combination method is not very popular among neural network practitioner if compare to PCR and simple averaging. This is probably due to the fact that ICR needs additional software that is embedded into Matlab™ such as FastICA in order to implement it. In ICR, The mixed signals, X (training data outputs from the 20 predicted networks) can be described by its coefficient matrix or mixing matrix, A, the multiple linear regression equation between A and y (conversion or dimensionless temperature value; for the moment conversion, y is taken for example) matrix can be written as following like PCR:

$$y=AB$$

The whole ICR algorithm is very similar to that of PCR. The only difference is to use ICs and their mixing matrix obtained by the FastICA package in Matlab™ in the regression operation, instead of PCs and score matrix obtained by PCA.¹⁵ The details of the ICR calculation can be summarized as the following five steps:

- i. Input the training data outputs from the 20 predicted networks, X to the FastICA package in Matlab™.
- ii. Obtain the mixing matrix, A and independent components, ICs. (IC same as S in equation X=AS)
- iii. Calculate the regression coefficients, B in equation y=AB with A+y, where A+ is the pseudo inverse of A.
 $A+ = [ATA]^{-1}AT$
- iv. Calculate the mixing matrix, Ap of the prediction set with XICT, and then predict the conversion, yp with ApB.
- v. Finally, calculate the SSE and R values.

Both ICA and PCA are process monitoring methods based on data that have been utilized in different level of processes and systems. The difference between the two methods is that the components of PCA are still dependent while ICA has no orthogonality constraint and its latent variables are independent. Most of the studies, the independent ICA components are also functionally independent.²⁰ As mentioned earlier, the

implementation of ICA needs additional software called FastICA. This software package was created in Laboratory of Computer and Information Science (CIS), Department of Computer Science and Engineering at the Helsinki University of Technology. In this work, multiple neural networks have been utilized with three methods of combination, namely simple averaging, PCR and ICR methods. To cope with issues such as lack of data for training and testing processes due to intricate process and time constraint, bootstrap re-sampling method has been employed to the original data to increase the number of data.

Results and Discussion

One of the aspects to obtain the “best” performance networks in ANN modeling is through the optimum number of hidden neurons. In order to determine the effect of the number of hidden neurons towards network performance, a particular network approach was used which is conversion, y, predicted with noise power of one, using single hidden layer and PCR combination. Therefore, in this paper, sum square error (SSE) values were used as a tool to indicate the performance of a network whereby larger SSE values signify bad performance and vice versa. Fig. 2 shows the graph of SSE values for the validation data versus the number of hidden neurons, varying from 5 hidden neurons to 30 hidden neurons in a single network. From the figure, we know that there was no trend pointed the exact number of hidden neurons in order to get the minimum value. Hence, the best number of hidden neurons used is about 30 which generating the smallest SSE value that is 7.6867. However, it is computational burden as the value of hidden neurons considered high. Other than that, the SSE value for 10 hidden neurons is closed to the above discussed hidden neurons, which are 7.8378. Therefore, there is no significant change of SSE while varying the numbers of hidden neurons in the hidden layer.

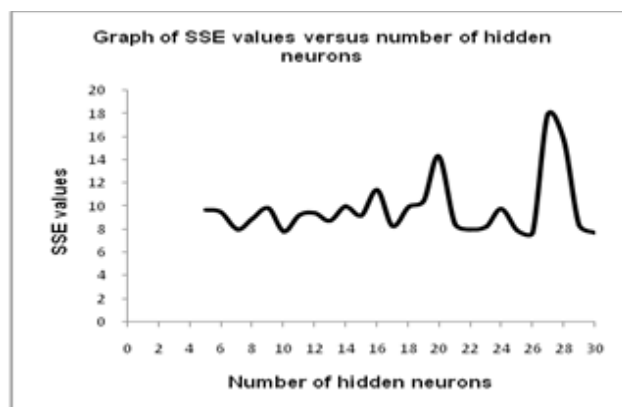


Figure 2: Graph of SSE values versus number of hidden neurons in a hidden layer.

Fig. 3 shows the graph of SSE values for the validation data versus number of hidden layers used in a single network. It is noticed that even though the number of hidden layer have been increased, the SSE values does not change much, but it takes longer time to perform. Indeed, the complexity of the network is directly proportional to the number of hidden layer. Consequently, a single hidden layer is the best to use for the exothermic bio catalytic Reaction prediction.

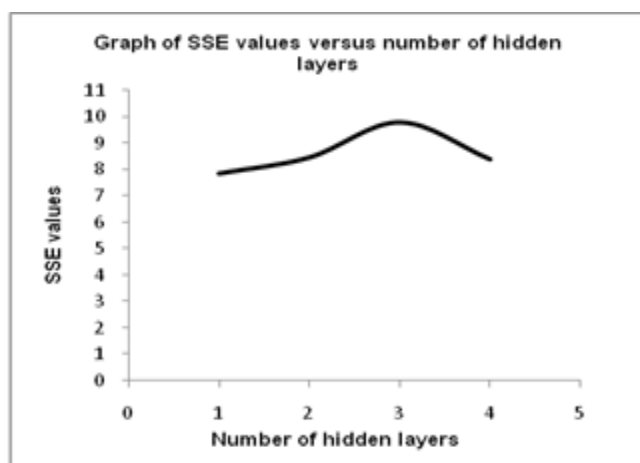


Figure 3: Graph of SSE values versus number of hidden layers in a single network.

Table 3: Best single neural network for dimensionless temperature values, θ .

Noise level	m	n	R^2
1	10.6361	8	0.9840
2	14.6223	12	0.9781
3	19.0730	15	0.9704
4	24.0523	6	0.9625
5	41.5826	6	0.9338

Table 4: Best single neural network for dimensionless conversion, y .

Noise level	m	n	R^2
1	3.7220	3	0.9945
2	5.0891	3	0.9924
3	6.3296	4	0.9905
4	8.4303	15	0.9875
5	12.1150	8	0.9817

While developing the mechanistic model, noises were added to the dilution rate, D in order to mimic the real situation of the exothermic bio catalytic reaction which is nonlinear in nature. The input variable that is

dilution rate, D was corrupted with normally distributed random noise with zero mean and a standard deviation of 0.2, 0.4, 0.6, 1.0 and 1.5. The noise level was varied while the number of hidden neurons and hidden layer were fixed to 10 and single layer, respectively. Table 3 and 4 shows the best single neural network among the 20 predicted networks which had been trained by using the single hidden layer with 10 hidden neurons for conversion, y and dimensionless temperature values, θ . The SSE values for the validation data as well as the corresponding R^2 value are well tabulated in Table 3 and 4. R^2 value represents the normalized measure of the strength of linear relationship between variables. With m is the lowest SSE values, n is the i^{th} neural network with the lowest SSE value, (best single neural network) ($i=1, 2, 3, \dots, 20$) and R^2 is correlation coefficient value corresponding to the best single neural network. Higher R^2 value (normally range from 0.95 to 1.00) shows best relationship between the predicted output values and the original output values. Increasing the noise level deviate the input values from the nominal values, and so does happen to the output values. Therefore, increasing the noise level will increase the SSE values and affect the linear relationship between the predicted output values and original output values. As the conclusion, there is no consistent for the best single network at which different noise level it will end up with different number of network (n that is i^{th}) as the best neural network.

Table 5: Overall results for dimensionless conversion, y .

Noise level	PCR			ICR		
	SSE	N	R^2	SSE	N	R^2
1	4.4721	3	0.9933	4.2792	1	0.9936
2	5.5209	11	0.9917	5.2483	2	0.9921
3	6.6491	19	0.9900	6.3964	6	0.9904
4	7.8378	16	0.9884	8.1888	5	0.9877
5	13.0240	6	0.9803	13.0533	1	0.9803

Table 5 and 6 shows the SSE for the validation data, R^2 values and the number of networks that have been combined (N) for the final prediction for conversion, y , and dimensionless temperature value, θ for PCR and ICR prediction methods while varying the noise level. The SSE values are indirectly proportional to the R^2 values. Furthermore, it can conclude that the results obtained for the prediction of conversion, y , and dimensionless temperature value, θ , by using both the PCR and ICR methods were almost the same. This can be proved by the SSE and R^2 values obtained for both of the methods. Furthermore, comparison between simple averaging method and PCR as well as ICR methods were made. Simple averaging method is just a

method that finding the mean values for the combination of 20 networks predicted output values. After that, SSE and R^2 values for this method were calculated.

Table 6: Overall results for dimensionless temperature, θ .

Noise level	PCR			ICR		
	SSE	N	R ²	SSE	N	R ²
1	10.8657	9	0.9834	10.5227	13	0.9841
2	16.0021	20	0.9754	15.6695	2	0.9763
3	18.9240	10	0.9708	18.2589	5	0.9723
4	22.9755	13	0.9640	24.3460	5	0.9630
5	44.9360	13	0.9295	44.2564	14	0.9327

Table 7 and 8 shows the results obtained for the ICR and simple averaging methods. From the results obtained, it seems that the simple averaging combination method gave lower SSE values and higher R^2 values compared to PCR and ICR combination methods. Therefore, it can be concluded that simple averaging combination method is slightly better and more suitable to use for combination of the 20 neural networks predicted outputs as compared to PCR and ICR using the selective combination method. Fig. 4 and 5 shows the graphs for SSE values for validation data and R^2 values for prediction for conversion, y , and dimensionless temperature value, θ , by using the PCR, ICR and simple averaging combination methods respectively.

Table 7: Overall results between ICR and simple averaging for dimensionless conversion, y .

Noise level	ICR			Simple averaging	
	SSE	N	R ²	SSE	R ²
1	4.2792	1	0.9936	4.2698	0.9936
2	5.2483	2	0.9921	5.2127	0.9922
3	6.3964	6	0.9904	6.3917	0.9904
4	8.1888	5	0.9877	8.1618	0.9877
5	13.0533	1	0.9803	12.9526	0.9804

Table 8: Overall results between ICR and simple averaging for dimensionless temperature, θ .

Noise level	ICR			Simple averaging	
	SSE	N	R ²	SSE	R ²
1	10.5227	13	0.9841	10.4006	0.9841
2	15.6695	2	0.9763	14.3061	0.9779
3	18.2589	5	0.9723	17.8205	0.9723
4	24.3460	5	0.9630	23.6179	0.9629
5	44.2564	14	0.9327	42.2503	0.9327

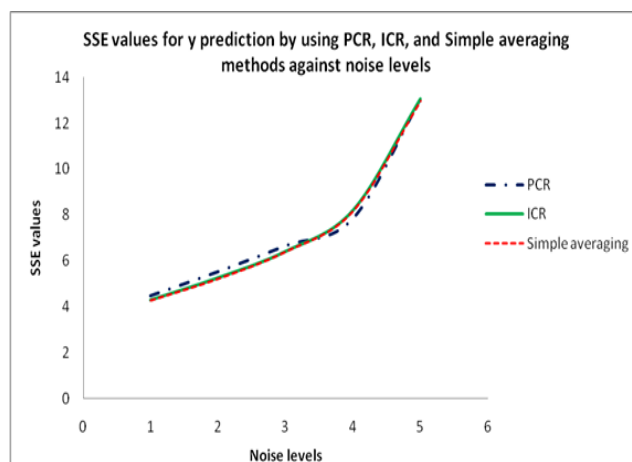


Figure 4: SSE values for PCR, ICR and simple averaging versus noise levels for dimensionless conversion, y .

Fig. 6 and 7 shows the actual, PCR predicted, and ICR predicted output y values for noise level 1 and noise level 5. While Fig. 8 and 9 shows the actual, PCR predicted, and ICR predicted output θ values for noise level 1 and noise level 5 as well. It is noticed that increasing the noise level will increase the deviation and this end up with the increasing in the amplitude of the output y and θ values. The trend is quite obvious from Fig. 9. Hence, it can be concluded that increased the noise level disturbed the order of the system and therefore caused the SSE values for validation data to increase and decreased in the R^2 values.

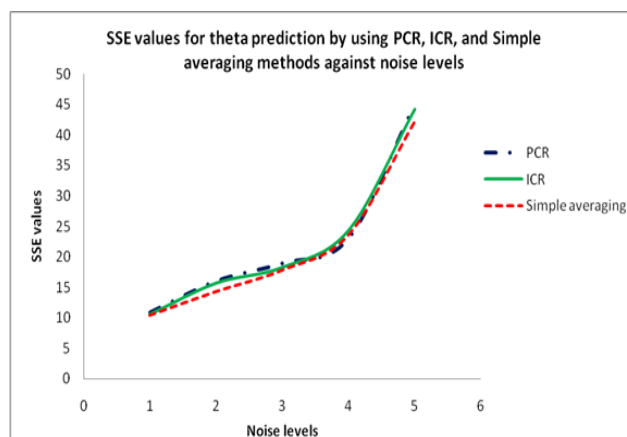


Figure 5: SSE values for PCR, ICR and simple averaging versus noise levels for dimensionless temperature, θ .

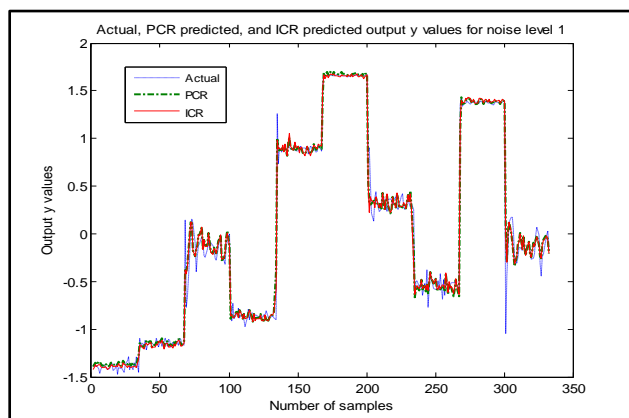


Figure 6: Normalize Actual, PCR and ICR prediction output for dimensionless conversion, y for noise level 1.

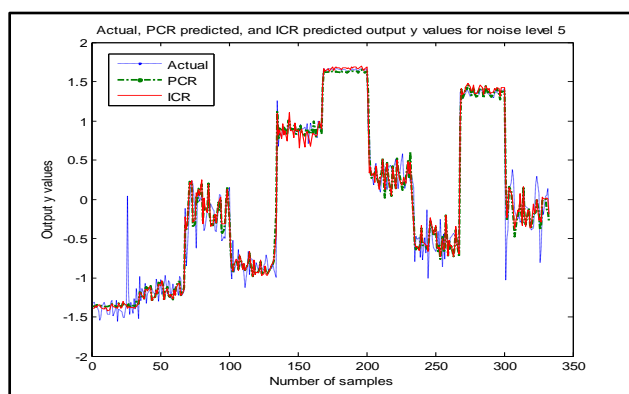


Figure 7: Normalize actual, PCR and ICR prediction output for dimensionless conversion, y for noise level 5.

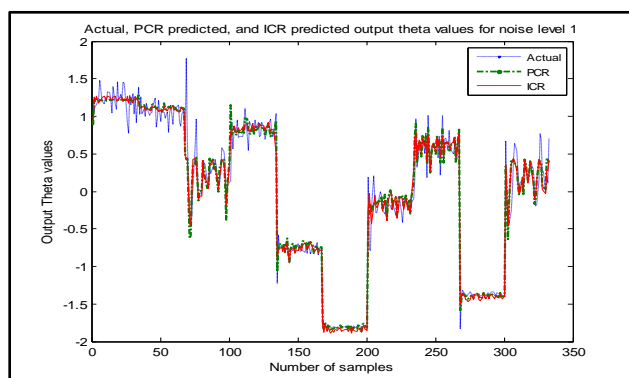


Figure 8: Normalize actual, PCR and ICR prediction output for dimensionless temperature, θ for noise level 1.

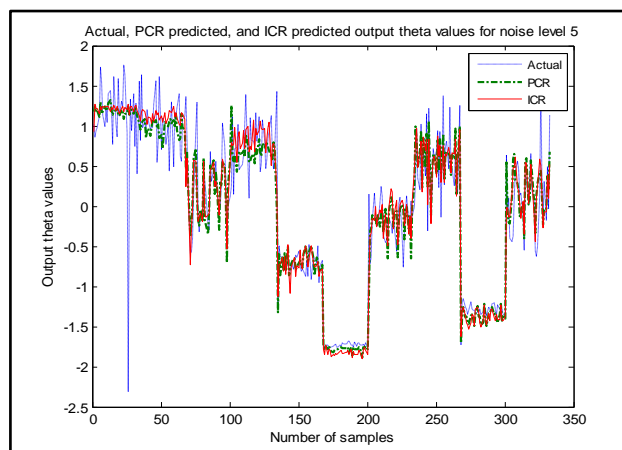


Figure 9: Normalize actual, PCR and ICR prediction output for dimensionless temperature, θ for noise level 5.

Conclusion

Feedforward Artificial Neural Network (FANN) with single hidden layer of 10 hidden nodes, log-sigmoid neurons followed by an output layer of linear neuron was then applied to develop the prediction model of the exothermic bio catalytic reaction. Then, bootstrap re-sampling approach was used to create multiple neural networks. In this study, the combinations methods used were focusing on ICR and compare with the two other conventional combination approaches names simple averaging and PCR. ICR combination method was not commonly used; this might due to the complexity to obtain the mixing matrix and ICs. The advantages of using ICs instead of PCs are it was obtained under a high order statistic. However, in this study, it showed some significant different between the PCR and ICR combination method. It is observed that by applying combination through ICR, multiple neural network performance is further improved especially in noise level 1, 2 and 3 however for noise level 4 and 5 PCR is a bit superior as well as simple averaging.

Acknowledgement

This work was supported by Universiti Sains Malaysia (USM) through Graduate on Time (GOT) grant 8023006.

References

1. M.N. Karim, T.Yoshida, S.R. Rivera, V. M. Saucedo, B. Eikens and O. Gyu-Seop, "Review Global and Local Neural Network Models in Biotechnology: Application to Different Cultivation Processes," *Journal of Fermentation and Bioengineering*, vol.83, pp. 1-11, 1997.
2. P.R. Patnaik, "Applications of Neural Networks to Recovery of Biological Products. *Biotechnology Advances*," vol. 17, pp. 477-488, 1999.

3. Z. Ahmad, R.A. Mat Noor and J. Zhang, "Multiple Neural Networks Modeling Techniques in Process Control: A Review," *Asia-Pacific Journal of Chemical Engineering*, vol. 4, pp. 403-419, 2009.
4. Z. Ahmad, N. A. Rahim, A. Bahadori and J. Zhang, "Improving Water Quality Index Prediction in Perak River Basin Malaysia through the Combination of Multiple Neural Networks," *International Journal of River Basin Management*, 2019. DOI: 10.1080/15715124.2016.1256297
5. Z. Ahmad, N. A. Rahim, A. Bahadori and J. Zhang, J, "Air Pollution Index Prediction using Multiple Neural Network," *IJUM Journal*, in press, 2017.
6. L. Xiong, J. Liang and J. Qian, "Multivariate Statistical Process Monitoring of an Industrial Polypropylene Catalyzer Reactor with Component Analysis and Kernel Density Estimation," *Chinese Journal of Chemical Engineering*, vol.15, pp. 524-532, 2007.
7. A.Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural Networks*, vol.13, pp. 411-430, 2002.
8. A. Jutten and J.Herault, "Independent Component Analysis versus PCA," In the Proceeding of European Signal Processing Conference, pp: 287-314, 1998.
9. P. Comon, "Independent component analysis — a new concept?" *Signal Processing*, vol. 36, pp. 287-314, 1994.
10. E. Oja, 1997. The nonlinear PCA learning rule in independent component analysis. *Neurocomputing*, vol. 17, pp. 25-46,1997.
11. J.M. Lee, C.K. Yoo and I.B Lee, "Statistical process monitoring with independent component analysis," *Journal of Process Control*, vol. 14, pp. 467-485, 2004.
12. M. Chen, C. Hsu, M. Bharat, and K. T. Manoj, "An efficient ICA-DW-SVDD fault detection and diagnosis method for non-Gaussian processes," *International Journal of Production Research*, vol. 54, no. 17, pp. 5208-5218,2016.
13. Y. Zhang and Y. Zhang, "Optimized independent components for parameter regression," *Chemometrics and Intelligent Laboratory Systems*, vol. 104, pp. 214-222, 2010.
14. S.M. Al-Alawi, S.A. Abdul-Wahad and C.C. Bakheit, "Combining Principal Component Regression and Artificial Neural Networks for More Accurate Predictions of Ground-Level Ozone," *Environmental Modelling & Software*, vol. 23, pp. 396-403, 2008.
15. X.G. Shao, W. Wang, Z.Y. Hou and W.S. Cai, "A New Regression Method Based On Independent Component Analysis," *Talanta*, vol. 69, pp. 676-680, 2006.
16. M.R. Mohd. Radzi and M.H. Uzir, "Stability Study of an Exothermic Bio catalytic Reaction and its Application in Bioprocess Systems," *Pertanika Journal of Science & Technology*, vol. 17, pp. 95-115, 2009.
17. A.J. Adeloye and A. De Munari, "Artificial Neural Network Based Generalized Storage-yield-reliability Models Using the Levenberg-Marquardt Algorithm," *Journal of Hydrology*, vol. 362, pp. 215-230, 2006.
18. Z. Ahmad and J. Zhang, "Selective Combination of Multiple Neural Networks For Improving Model Predictive In Nonlinear Systems Modeling Through Forward Selection and Backward Elimination," *Neurocomputing*, vol. 72, pp. 1198-1204, 2009.
19. W. Dai, J.Y. Wuc and C.J. Lu, "Combining nonlinear independent component analysis and neural network for the prediction of Asian stock market indexes," *Expert Systems with Applications*, vol. 39, pp. 4444-4452, 2012.
20. M.P.S Chawla, H.K. Verma and K. Vinod, "Artifacts and Noise Removal in Electrocardiograms using Independent Component Analysis," *International Journal of Cardiology*, vol. 129, pp. 278-281, 2008.

Available online at <http://www.banglajol.info/index.php/CERB>