Integrated OBF-NN models for extrapolation enhancement in conventional neural networks for nonlinear systems

H. Zabiri, M. Rasamay, T. D. Lemma, and A. Maulud

***Abstract*—In this paper the integration of linear and nonlinear models in parallel for nonlinear system identification is investigated. A residuals-based sequential identification algorithm using parallel integration of linear Orthornormal basis filters (OBF) and a nonlinear feedforward (MLP) NN models is used and applied to the nonlinear Van de Vusse reactor. Results show promising capability of the proposed method with improved extrapolation capability.**

I. Introduction

ONE of the open areas in system identification is the identification of nonlinear models and it is probably the most active area in system identification today [1]. In essence, there are two standard approaches for building mathematical models: (i) the traditional physical modeling (white box) and (ii) the purely empirical modeling (black box) [2]. White box models are based entirely on fundamental principles and as such lead to mathematical representations with clear physical interpretations. However, they tend to be highly complex and generally are difficult and very time consuming to be developed [3-5]. Nonlinear black box models, on the other hand, mainly aim to determine a mathematical model of process dynamics that matches observed input/output data according to some objective matching criteria. Hence, they have certain advantages over the white box models in terms of development time and efforts.

The most common among the different black box model structures are nonlinear auto-regressive with exogenous inputs (NARX) models, Volterra series expansion models, block-oriented models (Hammerstein and Wiener structures), and artificial neural networks [3, 5-8]. From model-based control strategies point of view, algorithms based on neural network (NN) models are preferred [5]. NN are highly efficient in learning data from complex processes with significant nonlinearity [9-11, 5, 12]. Training of NN requires no technical knowledge about the process and the corresponding models usually have a small number of parameters with simple structure [5]. The two most common types of NN that are widely employed in process control applications are recurrent neural networks (RNN) and feed-forward neural networks or also known as Multi-Layer Perceptron (MLP) [13].

Even though they are highly efficient, one of the significant drawbacks of NN as a model is that they have poor extrapolation property in regions outside those that are used during training [5, 10]. They are pure black-box models, i.e. model structure and model parameters (the weights and biases) have no physical interpretation [5]. To overcome or weaken this disadvantage, two possible solutions have been cited in literatures; namely (1) the use of dynamic nonlinear models are suggested, and one of the proposed solutions is to utilize Wiener model structure [9-11] with NN as the static nonlinear subsystem, and (2) integration of a linear model in parallel to the nonlinear NN model [14] (parallel linear/nonlinear models).

Wiener model structure is one of the most commonly used block-oriented (BO) models [10] and various applications in industrial nonlinear systems, such as distillation columns and pH processes have been reported [6-7, 15]. Wiener models consist of a dynamic linear part cascaded with a static nonlinear component as shown in Figure 1. Common model classes for the dynamic linear subsystem are FIR filters, input/output models or state space models, and Orthonormal Basis filters (OBF), e.g. Laguerre or Kautz filters [16]. The memoryless requirement for the nonlinear subsystem allows almost unlimited choices for the nonlinear element and the use of NN as the static nonlinear component has been investigated by various authors [5, 8-9, 17-23]. However, majority of these methods, in general, were either more parametric and require the use of efficient optimization methods [3, 5, 20, 24], involve complex structure [22] or necessitate the usage of two types of input designs [9].

In contrast, the parallel integration of linear/nonlinear models approach as suggested in [14] provides an interesting alternative in overcoming the extrapolation weakness of NN. This approach relies on the fact that a nonlinear model may perform worse than the linear one if it is not chosen appropriately. By developing a linear model (obtained by using either input/output data or by first principle strategy) in parallel with the nonlinear one, such that the overall model output is determined by the sum of the linear and the nonlinear parts, the performance of the overall nonlinear model is then ensured to be either equivalent or superior than the linear model (see Figure 2).

Manuscript received May 2, 2001. This work was supported Universiti Teknologi PETRONAS.

H. Zabiri is with the Universiti Teknologi PETRONAS, Bandar Seri Iskandar, Tronoh, 31750 Perak Malaysia (corresponding author to provide phone: 605-638-7625; fax: 605-365-6176; e-mail: haslindazabiri@petronas.com.my).

M. Ramasamy is with the Universiti Teknologi PETRONAS, Bandar Seri Iskandar, Tronoh, 31750 Perak Malaysia (e-mail: marappagounder@petronas.com.my).

T. D. Lemma is with the Universiti Teknologi PETRONAS, Bandar Seri Iskandar, Tronoh, 31750 Perak Malaysia (e-mail: lemma_dendena@petronas.com.my).

A. Maulud is with the Universiti Teknologi PETRONAS, Bandar Seri Iskandar, Tronoh, 31750 Perak Malaysia (e-mail: halims@petronas.com.my).

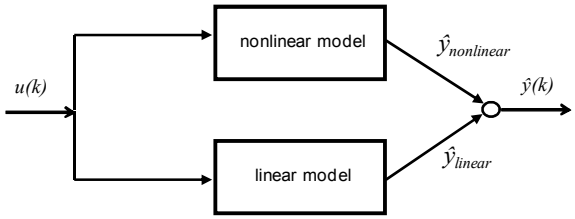

Fig. 2. The nonlinear model is used supplementary in parallel to a linear model [14].

In this paper, residuals-based sequential identification of the parallel linear/nonlinear models approach is investigated. To represent the linear dynamic subsystem, Orthonomal Basis Filters (OBF) (e.g. Laguerre, Kautz filters etc) is selected. OBF models have recently found widespread applications in linear system identification [3, 6, 8, 14, 26]. OBF models have several characteristics that make them very promising for control relevant system identification. Their parameters can be easily estimated using linear least square method. They are consistent in their parameters for most practical open-loop identification problems and time delays can be easily estimated and incorporated into the model. To model the nonlinear subsystem, MLP NN is chosen due to its simpler structure and fewer parameters in comparison to RNN.

II. RESIDUALS-BASED PARALLEL OBF-NN MODEL

A. The idea

In prediction methods, the analysis of residuals is often underestimated [27]. There are occasions where residuals are not due to randomness and may actually inherit the characteristics of the original system. In this paper the contribution of residuals in a sequential nonlinear identification technique using parallel integration of the linear OBF and nonlinear NN models is used.

The proposed configuration may be represented using a block diagram as shown in Figure 4.

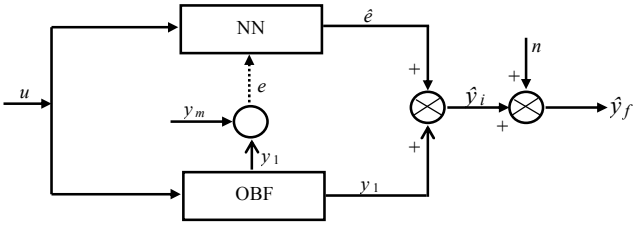

Figure 4. The block diagram of the proposed residuals-based sequential identification of parallel OBF-NN models.

Note that in the proposed once-through sequential method, the linear dynamic OBF model is developed first. The input to the NN is the original input to the process, u, and the residuals (i.e. $e = y_m - y_1$, where y_m is the actual measured output and y_1 is the linear OBF model predicted output) are the output to be predicted by the nonlinear NN. Iterations, based on some convergence criteria of standard NN algorithm, are performed only in the NN block. The overall predicted output is then the summation of the predicted residuals, $\hat{e}$ and the OBF output, y_1.

Hence,

$$\hat{y}_i = y_1 + \hat{e} \tag{1}$$

The measurement noise is considered to be acting on the output and is designated by n. The NN used throughout this paper is a single hidden layer MLP network. The mathematical descriptions of the methods are given below using Laguerre filters to represent the OBF models for ease of discussions.

B. Identification of parallel Laguerre-NN models

First, a parsimonious linear Laguerre model is identified using methods described in [25] and as shown in Figure 5. Given a set of nonlinear data to be identified [$u(k)$,$y_m(k)$], a Laguerre model is developed using crude estimate or arbitrarily chosen poles. Then, one or two of the dominant poles of the system are estimated using the methods proposed in [25]. The estimated dominant poles are used to develop more accurate Laguerre model. A better estimate of the dominant poles is obtained from the new Laguerre model. The process is repeated until a convergence criterion is satisfied.

A SISO linear system modeled by Laguerre filters can generally be represented as follows [3, 6, 20]:

$$y_1(k) = \left(\sum_{i=1}^{N} c_i L_i(k) \right) u(k) \tag{2}$$

where

$$L_i(k) = \sqrt{(1-p^2)T_s}\,\frac{(1-pk)^{i-1}}{(k-p)^i} \tag{3}$$

$L_i(k)$ denotes the ith order Laguerre filter, N the number of Laguerre filters used for model development, p the Laguerre filter parameters (i.e. the dominant poles of the system), T_s the sampling interval, $y_1(k)$ the model output, and $u(k)$ is the process input. Once a satisfactory Laguerre model is obtained, the model output y_1 is used to calculate the residuals. These residuals are then passed to the MLP NN. Considering a SISO example with a single output neuron, the output of an MLP neural network with one hidden layer is given by (with reference to Figure 6)

$$\hat{e}(k) = \gamma(u(k), e(k)) = \beta\left[b^2 + \sum_{i=1}^{K} w_i^2 \varphi(b_i^1 + w_{i,1}^1 u(k)) \right] \tag{4}$$

where $\gamma(\cdot)$ is the nonlinear neural network function approximation, which is trained with data sets consisting of the input or regression vector, $u(k)$ and the residuals, $e(k)$. Also $\varphi, \beta : R \to R$ are the nonlinear activation functions

(e.g. hyperbolic tangent etc.), b are the biases, K is the number of hidden neurons, and the weights of the network are denoted by $w_{i,j}^1, i = 1, \ldots, K$ (with ith neuron and jth input, in this case $j = 1$) for the first layer, and $w_i^2, i = 1, \ldots, K$ for the second layer.

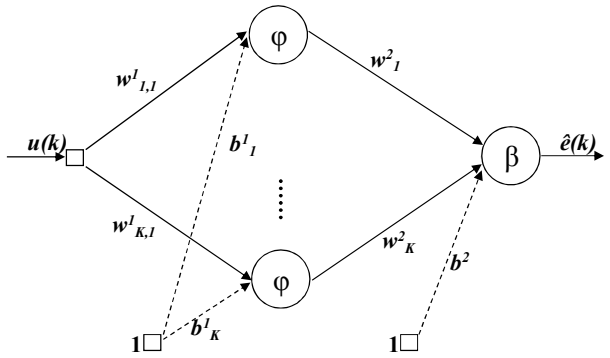

Figure 6. The structure of one hidden layer MLP neural network (a single input single output (SISO) example).

Without loss of generality and for ease of discussions, zero noise is assumed, i.e. $n = 0$, such that $\hat{y}_f = \hat{y}_i$. The overall model output then is the summation of both the linear dynamic model and the predicted residuals $\hat{e}$.

$$\hat{y}_f(k) = y_1(k) + \hat{e}(k) \tag{5}$$

$$\hat{y}_f(k) = y_1(k) + \gamma(u(k), e(k)) \tag{6}$$

Substituting equations (2) and (4),

$$\hat{y}_f(k) = \left(\sum_{j=1}^{N} c_j L_j(k) \right) u(k) + \beta \left[b^2 + \sum_{i=1}^{K} w_i^2 \varphi(b_i^1 + w_{i,1}^1 u(k)) \right] \tag{7}$$

C. Identification algorithm

The proposed identification algorithm is based on the recently developed identification algorithm for parsimonious OBF models by Lemma *et al.* [25] and the standard back-propagation (BP) algorithm for the neural network training.

Given a set of nonlinear data to be identified [$u(k)$,$y_m(k)$], the algorithm can be described as follows:

1. Develop a parsimonious OBF model using methods described by [25] to get y_1.
2. Calculate the residuals using $e = y_m - y_1$.
3. Develop the NN model using standard BP algorithm.

The input data to the NN is segregated as training and validation sets (75% for training, and 25% for validation) as is normally done with any NN modeling. The optimal NN model is determined by varying the activation functions for each layer and the regressors sets. Convergence criteria used is based on Root Mean Square Error (RMSE), defined by

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}} \tag{9}$$

III. Results and Discussions

To demonstrate the feasibility of the proposed residuals-based identification algorithm described in the previous section, Van de Vusse reactor case study is considered. The performance of the proposed method is compared with the standard NN model.

In all cases, the number of Laguerre filters is fixed at six (wherever applicable). For both standard NN and the proposed OBF-NN models, a single-hidden layer standard MLP network is adopted. The determination of the relevant inputs for the nonlinear NN function, $\gamma(\cdot)$, is essentially equivalent to the problem of dynamic order selection. In this work, the dynamic orders are arbitrarily selected, where two sets of regression vectors are considered for the standard NN and parallel OBF-NN models, with increasing complexity or dynamic order from m=1 to 2.

The case study is the Van de Vusse CSTR reactor. It is a highly nonlinear process and frequently used as a benchmark problem for various identification and nonlinear control strategies. In this isothermal CSTR reactor, a reactant A is to be converted to the desired product B, but the product B is degraded to product C. In addition to this consecutive reaction, a high-order parallel reaction occurs by which the reactant A is converted to by-product D.

$$A \xrightarrow{k_1} B \xrightarrow{k_2} C$$

$$2A \xrightarrow{k_3} D$$

The mathematical model of this reactor is described by the following set of ordinary differential equations (ODE):

$$\frac{dc_A}{dt} = \frac{q_r}{V_r}(c_{A0} - c_A) - k_1 c_A - k_2 c_A^2$$

$$\frac{dc_B}{dt} = -\frac{q_r}{V_r} c_B + k_1 c_A - k_2 c_B$$

$$\frac{dT_r}{dt} = \frac{q_r}{V_r}(T_{r0} - T_r) - \frac{\Delta h_r}{p_r c_{pr}} + \frac{A_r U}{V_r p_r c_{p_r}}(T_c - T_r)$$

$$\frac{dT_c}{dt} = \frac{1}{m_c c_{pc}}(Q_c + A_r U(T_r - T_c))$$

The net heat of reaction (Δh_r) for the above reactions is expressed as:

$$\Delta h_r = h_1 k_1 c_A + h_2 k_2 c_B + h_3 k_3 c_A^2$$

where h_i means reaction enthalpies. Nonlinearity can be found in reaction rates (k_j) which are described via the Arrhenius expression:

$$k_j(T_r) = k_{0,j} \exp\left(\frac{-E_j}{RT_r}\right), \text{ for } j = 1, 2, 3$$

where $k_{0,j}$ represents the pre-exponential factors and E_j are activation energies. Fixed parameters of the system are taken from [13]. The nonlinear identification is done for SISO system by considering the dynamic characteristics from the changes of the feed flow rate, F, and the product outlet concentration, C_B.

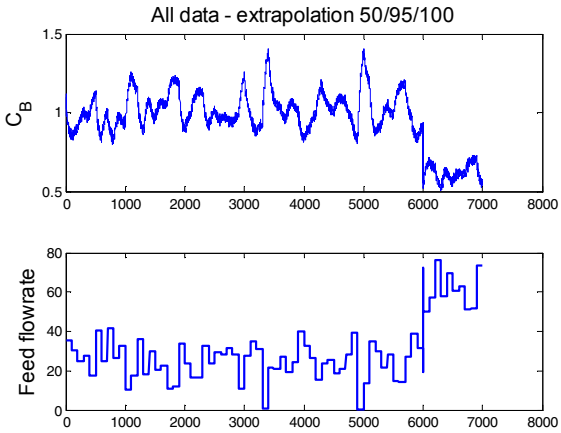

Figure 12. Training and Extrapolation analysis: Input-output data set for the Van de Vusse CSTR reactor case study.

Figure 12 shows the input output data for both the training set and the extrapolation analysis. Only very mild nonlinearity is considered in this paper. Both NN and OBFNN models are initially trained with the same training data set. The resulting final model is then used to predict the extrapolated data. In the extrapolation analysis, the range of the input-output data is very different from those used during training (please see Figure 12).

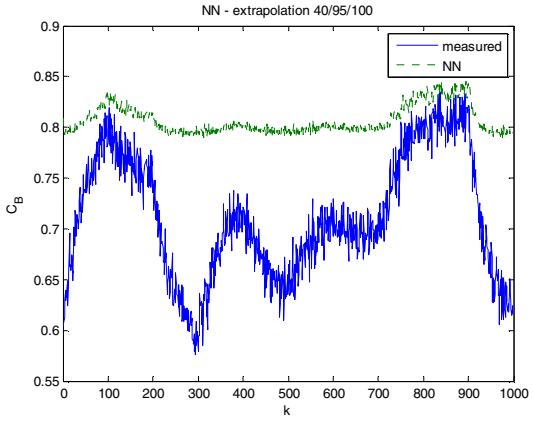

Figure 13. Extrapolation analysis: NN model.

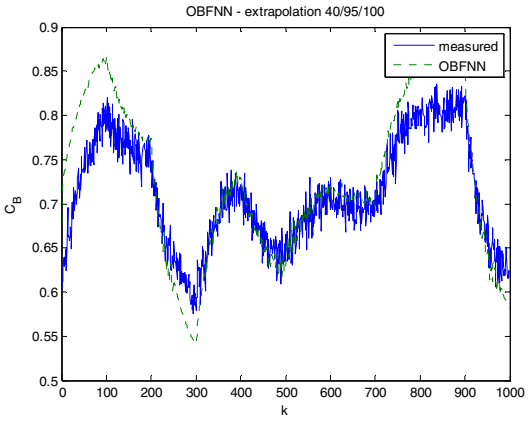

Figure 14. Extrapolation analysis: OBFNN model.

Figures 13-14 show the resulting performance of both NN and OBFNN models when applied to the extrapolated data sets. It is obvious that the enhanced formulation of residuals-based parallel OBF-NN model is able to predict the trend of the new data set more effectively than the standard NN. This example clearly illustrates that the newly proposed method significantly enhanced the extrapolation capability of standard NN. As mentioned by Sjoberg et al. [28], the use of linear models often leads to a robust and reasonable model. By applying the NN on the residuals, it is then ensured that a model at least as good as the linear one can be obtained.

IV. Conclusion

This paper presented a sequential identification approach using parallel integration of linear OBF and nonlinear NN models. Using the nonlinear Van de Vusse reactor case study, results show promising performance in the parallel OBF-NN structure, with better extrapolation capability than conventional MLP NN. Future research for the current work is to compare the efficacy of the method with existing Wiener-based methods.

Acknowledgment

The authors would like to thank Universiti Teknologi PETRONAS for the full financial support given of the project.

References

[1] L. Ljung, "Perspectives on System Identification", Proceedings of 17th IFAC World Congress, pp. 7172-7184, July, 2008.

[2] G. Gregorc ic, and G. Lightbody, "Nonlinear system identification: From multiple-model networks to Gaussian processes", Engineering Applications of Artificial Intelligence 21 (2008) 1035–1055.

[3] Saha, P., Krishnan, S.H., Rao, V.S.R., Patwardhan, S.C., 2004. "Modeling and predictive control of MIMO nonlinear systems using Wiener--Laguerre models". Chemical Engineering Communication 191, 1083--1119.

[4] S. I. Biagiola, J. L. Figueroa, "Wiener and Hammerstein uncertain models identification", Mathematics and Computers in Simulation (2009), 79, 3296-3313.

[5] M. Lawrynczuk, "Computationally efficient nonlinear model predictive control based on neural Wiener models", Neurocomputing (2010), doi:10.1016/j.neucom.2010.03.014.
[6] S. Mahmoodi, J. Poshtan, M. R. Jahed-Motlagh, A. Montazeri, "Nonlinear model predictive control of a pH neutralization process based on Wiener-Laguerre model", Chemical Engineering Journal (2009), 146, 328-337.
[7] G. Shafiee, M. M. Arefi, M. R. Jahed-Motlagh, A. A. Jalali, "Nonlinear predictive control of a polymerization reactor based on piecewise linear Wiener model", Chemical Engineering Journal (2008), 143, 282-292.
[8] P. Aadaleesan, N. Miglan, R. Sharma, P. Saha, "Nonlinear system identification using Wiener-type Laguerre-Wavelet network model", Chemical Engineering Science (2008), 63, 3932-3941.
[9] M. M. Arefi, A. Montazeri, J. Poshtan, M. R. Jahed-Motlagh, "Wiener-neural identification and predictive control of a more realistic plug-flow tubular reactor", Chemical Engineering Journal (2008), 138, 274-282.
[10] S. Piche, B.S. Rodsari, D. Johnson, M. Gerules, Nonlinear model predictive control using neural network, Proc. IEEE Contr. Syst. Mag. (2000), 20 (3), 53–62.
[11] H. Zhao, J. Guiver, R. Neelakantan, L.T. Biegler, A nonlinear industrial model predictive controller using integrated PLS and neural net state-space model, Control Eng. Pract. 9 (2001) 125–133
[12] Kalafatis, N. Arifin, L. Wang, W. R. Cluett, "A new approach to the identification of pH processes based on the Wiener model", Chemical Engineering Science (1995), 50 (23), 3693-3701.
[13] H.-W. Ge, F. Qiana, Y.-C. Liangb, W. Dua, L. Wang, "Identification and control of nonlinear systems by a dissimilation particle swarm optimization-based Elman neural network", Nonlinear Analysis: Real World Applications 9 (2008) 1345 – 1360.
[14] Nelles, O., 2001. Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models. Springer, Berlin.
[15] S. Totterman, H. T. Toivonen, "Support vector method for identification of Wiener models", Journal of Process Control (2009), 19, 1174-1181.
[16] Hagenblad, L. Ljung, A. Wills, "Maximum likelihood identification of Wiener models", Automatica (2008), 44, 2697-2705.
[17] L. R. Castro, O. E. Agamennoni, C. E. D'Attelis, "Rational wavelets in Wiener-like modeling", Mathematical and Computer Modeling (2002), 35, 991-1006.
[18] H. Al-Duwaish, M.N. Karim, V. Chandrasekar, "Use of multilayer feedforward neural networks in identification and control of Wiener model", IEE Proc. Contr. Theor. APpli., 143 (1996), 255-258.
[19] G. Chen, Y. Chen, H. Ogmen, "Identifying chaotic systems via a Wiener-type cascade model", IEEE Contr. Syst. Mag., 17 (1997), 29-36.
[20] Arto, P. Hannu, A. Halme, "Modeling of chromatographic separation process with Wiener-MLP representation", J. of Proc. Contr., 11 (2001), 443-458.
[21] G. B. Sentoni, L. T. Biegler, J. B. Guiver, H. Zhao, "State-space nonlinear process modeling: identification and universality", AIChE J. (1998), 44 (10), 2229-2239.
[22] Y.-L. Hsu, J.-S. Wang, "A Wiener-type recurrent neural network and its control strategy for nonlinear dynamic applications", Journal of Process Control 19 (2009) 942–953.
[23] J. S. Pei, A. W. Smyth, E. B. Kosmatopolous, "Analysis and modification of Volterra/Wiener neural networks for the adaptive identification of nonlinear hysteretic dynamic systems", Journal of Sound and Vibration (2004), 275, 693-718.
[24] Selami Beyhan, Musa Alci , Fuzzy functions based ARX model and new fuzzy basis function models for nonlinear system identification, Applied Soft Computing Volume 10 , Issue 2 (March 2010) pp 439-444 2010 ISSN:1568-4946.
[25] Lemma, D. Tufa, M. Ramasamy, M. Shuhaimi, "Improved method for development of parsimonious orthonormal basis filter models", Journal of Process Control, doi:10.1016/j.procont.2010.10.001, 2010.
[26] M. Ardalani-Farsa, S. Zolfaghari, "Chaotic time series prediction with residual analysis method using hybrid Elman-NARX neural networks", Neurocomputing (2010), doi: 10.1016/j.neucom.2010.06.004.
[27] Y. Liu, H. Wang, J. Yu, P. Li, "Selective recursive kernel learning for online identification of nonlinear systems with NARX form", Journal of Process Control, Volume 20, Issue 2, February 2010, pp 181-194.
[28] J. Sjoberg, Q. Zhang, L. Ljung, A. Benveniste, B. Deylon, P. Glorennec, H. Hjalmarsson, A. Juditsky, "Nonlinear black-box modeling in system identification: a unified overview", Automatica, 31 (1995), 1691-1724.