2009 American Control Conference
Hyatt Regency Riverfront, St. Louis, MO, USA
June 10-12, 2009

Recursive estimation on spatially rearranged data for dynamic order determination and nonlinear component detection

Ming Su and R. Russell Rhinehart

Abstract—**Determination of dynamic order of variables is the first step in system identification. Order determination is in general difficult for nonlinear system identification due to the interaction of system structure (unknown orders) and unknown nonlinearity. If the attenuation of unknown nonlinearity is possible, different system structures could then be fairly compared. Guided by this concept, this work uses a recursive estimation to reduce the effect of the underlying nonlinearity on parameter variation, and proposes a sequential nearest neighbor rearrangement to enhance the reduction. The "best" dynamic order will minimize final prediction error with the consideration of the locality of the model parameters. In addition to determining dynamic orders, the sequential nearest neighbor rearrangement is also extended to detect nonlinear components.**

I. Introduction

The first step in system identification is to determine the dynamic order of the model. Equation (1) is a generic expression of a single input and single output dynamic system with dynamic orders *ny*, *nu*, pure time delay *d*, and an additive disturbance *e* (t).

$$y(t)=f\begin{pmatrix} y(t-1),\cdots,y(t-ny), \\ u(t-d),\cdots,u(t-nu-d) \end{pmatrix}+e(t) \tag{1}$$

where, the estimation of *ny*, *nu* and *d* are expected before model coefficients values are estimated. It is also desired to know if *f* is linear or nonlinear with respect to its input arguments.

Dynamic order determination is well developed for linear systems where a preliminary analysis using autocorrelation or partial autocorrelation [1] is able to estimate dynamic orders. For static linear systems, subset selection methods [2] are often used to find influential regressors. Analysis of variance (ANOVA) can also be used for regressor analysis [3].

For nonlinear dynamic systems with unknown nonlinearities, there is no such general method; and order analysis falls into two categories. One accepts either known or assumed nonlinear structures. There are various choices of nonlinear structures such as bilinear structures, Wiener, Hammerstein structures, or their combinations. With known nonlinear structures, analysis could be conducted rigorously.

This work was supported in part by the Edward E. and Helen Turner Bartlett Foundation.

M. Su is with the School of Chemical Engineering, Oklahoma State University, Stillwater, OK 74078-5021 USA (ming.su@okstate.edu).

R. Russell Rhinehart is with the School of Chemical Engineering, Oklahoma State University, Stillwater, OK 74078-5021 USA (rrr@okstate.edu).

Another approach does not depend on a predefined nonlinear structure. The geometric method [4], False Nearest Neighbor [5], and Lipschitz Quotient [6] all belong to this category. These methods can be roughly argued upon the first-order Taylor expansion. One common problem in these generic methods is their sensitivity to noise.

In this work, a generic method is proposed to determine the dynamic orders of stochastic dynamic systems from data. In addition, the order determination technique can be extended to detect the nonlinear components, which provides insights on the underlying structure of the model.

II. Recursive estimation on spatially ordered data

The aim of this work is to detangle nonlinearity and order. In this work, the *f* in Equation (1) is assumed to be nonlinear; but its functional form is unknown, and adaptive linear models are used to treat nonlinearity.

A. Recursive estimation

The model in Equation (1) is described by the following model in Equation (2) in a linear format.

$$\begin{aligned} y(t)=c(t)+a_1(t)y(t-1)+\cdots+a_{ny}(t)y(t-ny)+ \\ b_0(t)u(t-d)+\cdots+b_{nu}(t)u(t-nu-d)+e(t) \end{aligned} \tag{2}$$

where model parameters $c(t)$, $a_i(t)$ and $b_i(t)$ are not constants. They could be considered as either functions of time or viewed as functions of states. Equation (2) is represented in a compact form as:

$$y(t)=\mathbf{x}^T(t)\boldsymbol{\theta}(t)+e(t) \tag{3}$$

with

$$\mathbf{x}(t)=\left[1,y(t-1),\cdots,y(t-ny),u(t-d),\cdots,u(t-nu-d)\right]^T$$

$$\boldsymbol{\theta}(t)=\left[c(t),a_1(t),\cdots,a_{ny}(t),b_0(t),\cdots,b_{nu}(t)\right]^T$$

where $\mathbf{x}(t)$ is the regressor vector and $\boldsymbol{\theta}(t)$ is the parameter vector. The estimation of $\boldsymbol{\theta}(t)$ can be approached in two ways. Windowing or weighting [7] techniques can estimate model parameters locally. Another approach is to model the parameters stochastically (such as random walk) and update them using a discrete Kalman filter [8]. There are discussions on high-order stochastic models with time varying parameters [9]. If the covariance matrix of model parameters is known a prior, the stochastic modeling is preferred. Otherwise, a windowing or weighting technique would be a proper choice.

The equations [7] for estimating $\boldsymbol{\theta}(t)$ by an exponential weighting are:

$$\begin{aligned}
&\hat{y}(t) = \mathbf{x}^T(t)\hat{\boldsymbol{\theta}}(t-1)\\
&\mathbf{K}(t) = \mathbf{P}(t-1)\mathbf{x}(t)\left(\mathbf{x}^T(t)\mathbf{P}(t-1)\mathbf{x}(t)+\alpha\right)^{-1}\\
&\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1)-\mathbf{K}(t)\left(\hat{\mathbf{y}}(t)-\mathbf{y}(t)\right)\\
&\mathbf{P}(t) = \frac{1}{\alpha}\left(\mathbf{P}(t-1)-\mathbf{K}(t)\mathbf{x}^T(t)\mathbf{P}(t-1)\right)
\end{aligned} \tag{4}$$

where $\hat{\boldsymbol{\theta}}(t-1)$ is the vector of parameter estimates at t-1. $\hat{y}(t)$ is the prediction of y(t) at time t using $\hat{\boldsymbol{\theta}}(t-1)$. $\mathbf{K}$(t) is the gain used to correct $\hat{\boldsymbol{\theta}}(t-1)$ to $\hat{\boldsymbol{\theta}}(t)$ based on the prediction error. $\mathbf{P}$(t) records the covariance of $\hat{\boldsymbol{\theta}}(t)$.

The tuning factor is the 'forgetting factor', α, which has to be chosen for a balanced performance for nonlinearity adaptation speed and parameter estimation precision.

A. Spatial ordering

In recursive estimation precision improves if α is increased, but also if the nonlinearity is reduced. Unfortunately, the nonlinearity is inherited in the data and solely determined by the nature of the system, which users are unable to change at all. However, the nonlinearity is in fact not really the difficulty; but the source of the difficulty: parameter variation. Therefore, it is desired to minimize the parameter variation directly.

The time-varying parameters in Equation (2) could be defined as functions of states as

$$a_i(t) = a_i\begin{pmatrix} y(t-1),\cdots,y(t-ny), \\ u(t-d),\cdots,u(t-nu-d)\end{pmatrix} \tag{5}$$

the exact functional form of a_i is unknown. If its continuity and differentiability are assumed and its high order derivatives are negligible, the difference between $a_i(k\text{-}1)$ and $a_i(k)$ could be approximated by:

$$\begin{aligned}
&a_i(k)-a_i(k-1) \approx \sum_{l=1}^{ny}\frac{\partial a_i}{\partial y(t-l)}\Big|_{k-1}\left(y(k-l)-y(k-1-l)\right)\\
&+\sum_{j=0}^{nu}\frac{\partial a_i}{\partial u(t-j-d)}\Big|_{k-1}\left(u(k-j-d)-u(k-1-j-d)\right)
\end{aligned} \tag{6}$$

If the first-order derivative is bounded by a constant Ga_i, the minimum of $\|a_i(k\text{-}1) - a_i(k)\|$ is bounded by:

$$\min\left\|a_i(k)-a_i(k-1)\right\| \le Ga_i \min\left\|\mathbf{x}(k)-\mathbf{x}(k-1)\right\| \tag{7}$$

If $\mathbf{x}(k\text{-}1)$ is known, $\mathbf{x}(k)$ is chosen to be the nearest neighbor to $\mathbf{x}(k\text{-}1)$. The selection procedure is then termed as Sequential Nearest Neighbor Rearrangement (SNNR). The resultant regressor and output are labeled as $\mathbf{x}_{snnr}$ and y_{snnr}. The rearrangement starts letting $\mathbf{x}_{snnr}(1) = \mathbf{x}(1)$ and $y_{snnr}(1) = y(1)$. If the nearest neighbor of $\mathbf{x}_{snnr}(1)$ is found (at time t)), and $\mathbf{x}$(t) and y(t) are then added to the rearranged data set by letting $\mathbf{x}_{snnr}(2) = \mathbf{x}(t)$ and $y_{snnr}(2) = y(t)$. Then the nearest neighbor of $\mathbf{x}_{snnr}(2)$ is found and added to the rearranged data set. The procedure continues until the $\mathbf{x}_{snnr}(N)$ is found.

By conducting the SNNR, the raw data in the time-ordered sequence is reorganized by state or "spatial" order, in order to reduce the parameter variation, which enables the choice of a larger forgetting factor, α, and in turn improves the parameter estimates. The impact of the improved parameter estimates on order determination will be demonstrated in the later section.

III. Dynamic Order Determination

A. Model comparison criterion

The methodology for determination of dynamic orders is to try different sets of *ny*, *nu* and *d* and find the best values. Given a set of *ny*, *nu* and *d*, regressors are determined, first on the original time-sequenced data, $\mathbf{x}$(t). A SNNR is then conducted on $\mathbf{x}$(t) and y(t) producing $\mathbf{x}_{snnr}$(t) and y_{snnr}(t), to which an exponential weighting recursive estimation will be applied. The quality of the hypothesized *ny*, *nu* and *d* will then be evaluated by a criterion considering both fitting and generalization performance. In this work, the final prediction error (FPE) metric is chosen and modified. The original FPE [10] is defined for a linear model with N samples:

$$\frac{N+np}{N-np}\frac{1}{N}\sum_{t=1}^{N}\varepsilon^2\left(t,\hat{\boldsymbol{\theta}}_N\right) \tag{8}$$

Equation (8) can be interpreted as a weighted mean squared error where the weighting is determined by N, the size of data set as well as the model complexity, *np*, the number of parameters. The *FPE* criterion results from the performance index:

$$V_N = \sum_{t=1}^{N}\varepsilon^2\left(t,\hat{\boldsymbol{\theta}}_N\right) \tag{9}$$

In exponential weighting recursive estimation, the definition of FPE should be modified according the exponentially weighted performance index

$$V_k = \sum_{t=1}^{k}\alpha^{k-t}\varepsilon^2\left(t,\hat{\boldsymbol{\theta}}_k\right) \tag{10}$$

where V_k is varying and progressively includes more data. The weighting factor, α^{k-t} would become very small on its way back to t=1 that makes the remote error inconsequential in estimating $\boldsymbol{\theta}_k$. A critical number L is hence introduced to decompose V_k :

$$\begin{aligned}
V_k &= \sum_{t=1}^{k-L}\alpha^{k-t}\varepsilon^2\left(t,\hat{\boldsymbol{\theta}}_k\right)+\sum_{t=k-L+1}^{k}\alpha^{k-t}\varepsilon^2\left(t,\hat{\boldsymbol{\theta}}_k\right)\\
&\approx \sum_{t=k-L+1}^{k}\alpha^{k-t}\varepsilon^2\left(t,\hat{\boldsymbol{\theta}}_k\right)
\end{aligned} \tag{11}$$

where, V_k is approximated by its recent portion. By this approximation, the number of data involved in V_k is a constant, L. Subsequently, the FPE based on V_k is redefined

$$FPE(k) = \frac{L+np}{L-np}\frac{1}{L}\sum_{t=k-L+1}^{k}\alpha^{k-t}\varepsilon^2\left(t,\hat{\boldsymbol{\theta}}_k\right) \tag{12}$$

where, the implicit constraints on t by k-L+1≥1 and k≤N bound k between L and N. The average of FPE(k) over all k is then defined as the performance criterion for the Equation (2) recursively estimated by Equations (4).

$$\begin{aligned}\overline{FPE} &= \frac{1}{N-L+1}\sum_{k=L}^{N}FPE(k) \\ &= \frac{L}{N-L+1}\frac{L+np}{L-np}\sum_{k=L}^{N}\sum_{t=k-L+1}^{k}\alpha^{k-t}\varepsilon^2\left(t,\hat{\boldsymbol{\theta}}_k\right)\end{aligned} \tag{13}$$

where, the double sum is decomposed into three parts after being switched

$$\begin{aligned}\sum_{k=L}^{N}\sum_{t=k-L+1}^{k}\alpha^{k-t}\varepsilon^2\left(t,\hat{\boldsymbol{\theta}}_k\right) &= \sum_{t=1}^{L-1}\sum_{k=L}^{t+L-1}\alpha^{k-t}\varepsilon^2\left(t,\hat{\boldsymbol{\theta}}_k\right) \\ &+ \sum_{t=L}^{N-L+1}\sum_{k=t}^{t+L-1}\alpha^{k-t}\varepsilon^2\left(t,\hat{\boldsymbol{\theta}}_k\right) + \sum_{t=N-L+2}^{N}\sum_{k=t}^{N}\alpha^{k-t}\varepsilon^2\left(t,\hat{\boldsymbol{\theta}}_k\right)\end{aligned} \tag{14}$$

The recursive estimation works well if parameter variation within a local range is assumed to be small.

$$\hat{\boldsymbol{\theta}}_{t+L-1} \approx \hat{\boldsymbol{\theta}}_{t+L-2} \approx \cdots \approx \hat{\boldsymbol{\theta}}_t \tag{15}$$

which in turn results in the following approximation

$$\varepsilon^2\left(t,\hat{\boldsymbol{\theta}}_{t+L-1}\right) \approx \varepsilon^2\left(t,\hat{\boldsymbol{\theta}}_{t+L-2}\right) \approx \cdots \approx \varepsilon^2\left(t,\hat{\boldsymbol{\theta}}_t\right) \tag{16}$$

The double sum is then simplified to

$$\begin{aligned}\sum_{k=L}^{N}\sum_{t=k-L+1}^{k}\alpha^{k-t}\varepsilon^2(t) &= \sum_{t=1}^{L-1}\varepsilon^2\left(t,\hat{\boldsymbol{\theta}}_t\right)\sum_{k=L}^{t+L-1}\alpha^{k-t} \\ &+ \sum_{t=L}^{N-L+1}\varepsilon^2\left(t,\hat{\boldsymbol{\theta}}_t\right)\sum_{k=t}^{t+L-1}\alpha^{k-t} + \sum_{t=N-L+2}^{N}\varepsilon^2\left(t,\hat{\boldsymbol{\theta}}_t\right)\sum_{k=t}^{N}\alpha^{k-t}\end{aligned} \tag{17}$$

If N is large. the second part dominates, which results in the further simplified average FPE as:

$$\overline{FPE} \approx \frac{L\sum_{k=0}^{L-1}\alpha^k}{N-L+1}\frac{L+np}{L-np}\sum_{t=L}^{N-L+1}\varepsilon^2\left(t,\hat{\boldsymbol{\theta}}_t\right) \tag{18}$$

The FPE in Equation (18) is similar to the original one in Equation (9), and has the same interpretation as a weighted prediction error, except that the weighting is different. In fact, only the term including np makes a difference. Hence, a simplified criterion termed as pseudo FPE in Equation (19) is used in this work. In the following elaboration, the pseudo FPE will continue to be denoted as FPE

$$FPE = \frac{L+np}{L-np}\sum_{t=L}^{N-L+1}\varepsilon^2\left(t,\hat{\boldsymbol{\theta}}_t\right) \tag{19}$$

The value of L is related to the decomposition (11) and determined by considering α^L small enough to be negligible. In this work, L is determined as below

$$L = \frac{4}{1-\alpha} \tag{20}$$

where $(1-\alpha)^{-1}$ is termed as *memory time constant* [10]. Interpreted as a time-constant in a first order process, L, the window length could then be viewed as the settling time.

B. Order determination procedure

Given several sets of *ny*, *nu* and *d*, their FPEs could be evaluated. The set with the minimum FPE is then reported as the determined order. The determination procedure could be conducted in an exhaustive approach for all possible combinations of different *ny*, *nu* and *d* given pre-defined *max_ny*, *max_nu* and *max_d* are pre-defined for maximum *ny*, *nu* and *d*.

The exhaustive approach is expensive. Suboptimal methods using subset selection procedure exists for linear systems. A subset selection procedure is always combined with orthogonalization to remove the redundant components among regressors. In a nonlinear system with unknown nonlinearity, orthogonalization is not possible. However, it does not mean that the subset selection can not be implemented. In this project, a forward selection procedure combing the above mentioned recursive estimation on spatially ordered data is proposed to find important regressors.

The procedure starts with user's input *max_ny*, *max_nu* and *max_d* as in an exhaustive approach. Then, a number of candidate regressors are generated and denoted as [x_1 x_2 x_3… x_m x_{random}]. $\mathbf{x}_{random}$ is a random regressor that presumably contains no meaningful information to predict output. At first, m+1 FPEs are computed for (y,[x_1]), (y, [x_2]), … , (y, [x_m]), (y, [x_{random}]), where y is the output and x_i in bracket is the regressor in consideration. The regressor with the minimum FPE is selected. If x_2, for instance, is the first selected regressor, there will be other m FPEs to be evaluated for (y, [x_2, x_1]), (y, [x_2, x_3]), …, (y, [x_2, x_m]), (y, [x_2, x_{random}]). Each bracket contains a combination of x_2 (first selected) with the rest of unselected ones. The regressor combination with the minimum FPE is then kept. The selection continues until a minimum FPE goes up or the x_{random} is selected. The injection of a random regressor is mentioned in [2] as a stopping criterion. The selection of x_{random} signifies that the rest of candidates are less influential on y than a presumably irrelevant one.

The forward selection does not determine *ny*, *nu* and *d*

directly. Sometimes, it creates absences between regressors. For instance, a forward selection could end up with a set of regressors such as [y(t-1) y(t-4) u(t-1) u(t-3)]. In this example, the missing regressors, y(t-2), y(t-3) and u(t-2), should not be included, because they provide redundant information, and their inclusion increases the model complexity. However, in this work, for programming convenience, absent regressors will be included using maximum *ny*, *nu* and *d* values due to the forward selection.

IV. Nonlinear component detection

This section extends the above mentioned technique to obtain more information on the system to be modeled, especially to discover the regressors that are affecting the output nonlinearly. The following example serves to demonstrate the concept of nonlinear components

$$y(t)=y(t-1)\,y(t-2)\big(y(t-1)+2.5\big)+u(t-1)+e(t) \quad (21)$$

where, u(t-1) has a linear affect on $y(t)$ while $y(t$-1) and $y(t$-2) affect $y(t)$ nonlinearly. In general, if Equation (3) could be represented by a linear model plus a nonlinear model as below, then the regressors, $x_{k+1},..,\ x_m$, in function g are considered as nonlinear components

$$f\left(x_1(t),\cdots,x_m(t)\right)=\sum_{i=1}^{k}a_i x_i(t)+g\left(x_{k+1}(t),\cdots,x_m(t)\right) \quad (22)$$

Such a decomposition of f is useful. Not only are some insights on the model structure gained, it also directly results in the reduction of modeling effort. For instance, if a neural network model is used, users will get a smaller network for g plus a linear model, rather than a big network model for f.

The proposed method to approach this problem is based on the result from the order determination to find out what regressors are included in the function g in Equation (22). In order to do that, the SNNR will be used again but in a subtle way. As discussed above, the purpose of conducting SNNR is to reduce parameter variation so that the recursive estimation is able to capture the variation better, which in turn results in a smaller FPE. The SNNR mentioned above rearranges data based on all the regressors in order to compare different sets of *ny*, *nu* and *d*. Given the decomposed structure in Equation (22), the corresponding time varying linear format for f is then described as:

$$f\left(x_1(t),\cdots,x_m(t)\right)=\sum_{i=1}^{k}a_i x_i(t)+\sum_{j=k+1}^{m}a_j(t)\,x_j(t) \quad (23)$$

where $a_i(i=1,\ldots,k)$ is constant while $a_j(j=k+1,\ldots,m)$ is time varying. Further more, the variation of a_j is determined by regressor x_j ($j=k+1,\ldots,m$) only. Therefore, in Equation (23), the nonlinear regressor x_j is responsible for parameter variation. It then is expected that a SNNR on $[x_{k+1},\ldots,x_m]$ would result in the minimum FPE, while a SNNR on $[x_1,\ldots,x_k]$ should have no positive impact on FPE. In this work, the detection of nonlinear regressors simply exhausts all possible combinations of a set of identified regressors. The combination with minimum FPE is reported as including nonlinear components.

V. Testing and Discussion

A. Testing models

Nine models are used in this work to test the proposed order determination and nonlinear component detection methods.

Model 1 [11]:

$$y(t)=0.3y(t-1)+0.6y(t-2)+0.6\sin\big(\pi u(t-1)\big)+0.3\sin\big(3\pi u(t-1)\big)+0.1\sin\big(5\pi u(t-1)\big)+e(t)$$

Model 2 [11]:

$$y(t)=\frac{y(t-1)\,y(t-2)\big(y(t-1)+2.5\big)}{1+y(t-1)^2+y(t-2)^2}+u(t-1)+e(t)$$

Model 3 [11]:

$$y(t)=\frac{y(t-1)\,y(t-2)\,u(t-2)\big(y(t-1)-1\big)+u(t-1)}{1+y(t-2)^2+y(t-3)^2}+e(t)$$

Model 4 [11]:

$$y(t)=0.8y(t-1)+\big(u(t-1)-0.8\big)\big(u(t-1)+0.5\big)+e(t)$$

Model 5:

$$y(t)=0.8y(t-1)+0.6y(t-2)+0.4u(t-1)$$

Model 6 [4]:

$$y(t)=4y(t-1)\big(1-y(t-1)\big)$$

Model 7:

$$y(t)=4y(t-1)\big(1-y(t-1)\big)+e(t)$$

Model 8 [4]:

$$y(t)=1-1.4y(t-1)^2+0.3y(t-2)$$

Model 9:

$$y(t)=1-1.4y(t-1)^2+0.3y(t-2)+e(t)$$

The first four models are nonlinear. Model 5 is a linear model used to reveal some interesting observations. Models 6 and 8 are deterministic nonlinear models mentioned in [4]. Models 7 and 9 are the stochastic version of Models 6 and 8.

B. Testing on the impact of SNNR

Assuming the order is known, this set of tests is conducted to reveal the impact of SNNR on recursive estimation compared to using of the original data in time order. The comparison results are summarized in Table 1, where further

comparisons on different norms are also provided. The numbers in Table 1 is the mean squared error (MSE).

TABLE I
MSE DUE TO TIME AND SPATIAL ORDERIING

Model	Time Order	Spatial Order		
		1-norm	2-norm	∞ norm
1	0.0692	0.0728	0.0751	0.0748
2	0.0740	0.0351	0.0365	0.0348
3	0.0052	0.0024	0.0026	0.0024
4	0.0244	0.0164	0.0164	0.0158
5	0.0242	0.0236	0.0240	0.0238

As observed in Table 1, SNNR is able to reduce the MSE significantly in Model 2, 3 and 4 tests. A slight increase of MSE is however observed in the Model 1 test. In the Model 5 test, the impact of SNNR is trivial and it could be argued that there is no statistically significant difference. It is evident that the strength of the influence of SNNR depends on the level of nonlinearity. It then brings forth one possible usage of the SNNR to tell if a given system is linear or nonlinear.

The improvements observed in Model 2, 3 and 4 tests substantiate the hypothesis that parameter variation slows down.

The slight increase of MSE in the Model test 1 is interesting. Regardless what norm is used, the SNNR consistently produces a slightly higher MSE. The consistent impact of SNNR suggests the existence of nonlinearity. The small magnitude of the increase MSE implies a low level of nonlinearity. The nonlinearity in Model 1 is merely due to the Sine functions that are not terribly nonlinear when the input is small. The slight increase in MSE is actually due to the inclusion of y(t-1) and y(t-2) when SNNR is operated. An ideal SNNR should be based on u(t-1) since it is the only nonlinear regressor. It then offers the opportunity in nonlinear component detection to discover which regressor is nonlinear.

It could also be concluded that that the choice of norm is arbitrary. However, it is still beneficial to try different norms. It appears that the model should be linear if different norms have random-like effects as shown in Model 5 test, where the MSE due to different norms wandering around the one based on time-ordered data. If results due to different norms deviate from the 'Time Order' MSE consistently, the nonlinearity should be suspected as shown in first four tests.

C. Testing on dynamic order determination

The following testing is devised to investigate the dynamic order determination. The detail on the determination procedure will be provided for the testing on Model 1. The maximum possible values for *ny*, *nu* and *d* are set to 5, 4 and 1. It then generates 11 candidate regressors [$y(t\text{-}1)$, …, $y(t\text{-}5)$,$u(t\text{-}1)$, …, $u(t\text{-}5)$, random] to be involved in the forward selection. The selection procedure is summarized in Table 2, where the four regressors are selected and $u(t\text{-}4)$ is discarded since the FPE starts increasing.

TABLE 2
FORWARD SELECTION FOR MODEL 1

Step	1	2	3	4	5
Selected	$y(t\text{-}1)$	$u(t\text{-}1)$	$y(t\text{-}2)$	$y(t\text{-}5)$	$u(t\text{-}4)$
FPE	0.1090	0.0850	0.0809	0.0800	0.0818

The selection result in Table 2 generates the absences of y(t-3) and y(t-4). An exhaustive order search with *max_ny* = 5, *max_nu* = 1 and *max_d* = 0 is then conducted. The result is summarized in Table 3.

TABLE 3
EXHAUSTIVE ORDER SEARCH FOR MODEL 1

	ny = 1	ny = 2	ny = 3	ny = 4	ny = 5
nu = 1, d= 0	0.0850	0.0809	0.0814	0.0846	0.0878

The minimum FPE is 0.0809 at *ny* = 2. It then reports three regressors, [y(t-1), y(t-2), u(t-1)]. The results for the first four models are summarized in Table 4. The results using time-ordered data are also provided in Table 4.

TABLE 4
RESULTS OF ORDER DETERMINATION FOR MODELS 1-4

Model	Time Order	SNNR	Truth
1	y(t-1)y(t-2)u(t-1)	y(t-1)y(t-2)u(t-1)	y(t-1)y(t-2)u(t-1)
2	y(t-1)u(t-1)u(t-2)	y(t-1) y(t-2) y(t-3) u(t-1)	y(t-1)y(t-2)u(t-1)
3	y(t-1) u(t-1)	y(t-1) y(t-2) y(t-3) y(t-4)u(t-1)	y(t-1)y(t-2)y(t-3) u(t-1)u(t-2)
4	y(t-1) y(t-2)u(t-1)	y(t-1) u(t-1)	y(t-1)u(t-1)

Modeling could make two types of mistakes with the choice of regressors: Either missing regressors that should be included or adding ones that do not exist in the true models. As observed in Table 4, both approaches are tie in Model 1 test. In the Model 2 test, the Time Order arrangement misses $y(t\text{-}2)$ but adds $u(t\text{-}2)$, while the SNNR arrangement adds $y(t\text{-}3)$. 'Time Order' makes three and one mistakes in the Model 3 and 4 tests. 'SNNR' makes two mistakes in the Model 4 test. For the first 4 tests, 'Spatial' has three mistakes while 'Time Order' makes 6 mistakes. By this criterion, data arranged by SNNR outperforms that by Time Order.

The proposed order determination technique is also compared to the geometric method [4]. The testing is conducted on Models 6-9, and results are summarized in Table 5. The best method is the one in which the identified regressors provide a closest match to the truth. As observed, the geometric method performs well for deterministic dynamic models while poorly for stochastic models. The geometric method is sensitive to noise. The proposed technique makes one mistake in the Model 8 test. Its

performance is much less influenced by noise than the geometric method.

TABLE 5
RESULTS OF ORDER DETERMINATION FOR MODELS 6-9

Model	SNNR	Geometric [4]	Truth
6	y(t-1)	y(t-1)	y(t-1)
7	y(t-1)	y(t-1) y(t-2) y(t-3)	y(t-1)
8	y(t-1) y(t-2) y(t-3)	y(t-1) y(t-2)	y(t-1)y(t-2)
9	y(t-1) y(t-2)	y(t-1) y(t-1)y(t-3)y(t-4)	y(t-1)y(t-2)

D. Nonlinear component detection

The last testing task considers nonlinear component detection. The implementation detail is given for the Model 1 test with selected regressors, [$y(t\text{-}1)$, $y(t\text{-}2)$, $u(t\text{-}1)$]. The result is recorded in Table 6.

TABLE 6
EXHAUSTIVE COMBINATORIAL TRIALS FOR MODEL 1

Regressors	1	2	3	12	13	23	123
MSE	0.0832	0.0856	0.0598	0.0849	0.0711	0.0736	0.0751

In the first trial, the SNNR is conducted based on y(t-1). The resultant data is then used in a recursive estimation that results in a MSE of 0.0832. The trial continues until all combinations are exhausted. The minimum MSE is 0.0598 that corresponds to the third regressor, u(t-1). If the model structure (23) is adopted, one might have the following structure for the Model 1, where g represents a nonlinear function, $y(t) = a_1 y(t-1) + a_2 y(t-2) + g\left(u(t-1)\right)$

The results of nonlinear component detection for the first four models are summarized in Table 7.

TABLE 7
RESULSTS FOR NONLINEAR COMPONENT DETECTION

Model	Detected nonlinear components
1	u(t-1)
2	y(t-2)
3	y(t-1) y(t-2) y(t-3) y(t-4) u(t-1)
4	u(t-1)

The discrepancy between detected and desired nonlinear components is clear in the Model 2 test. Only y(t-2) is found to be a nonlinear component while y(t-1) is missed. One possible reason is that the nonlinear affect of y(t-1) to y(t) is not strongly expressed. Another reason is probably due to fact that only the minimum MSE is reported. In Table 1, the MSE without SNNR is 0.074. The MSE due to y(t-2) is 0.0263, which is the minimum. The MSE due to y(t-1) is 0.0301. The MSE due to [y(t-1) y(t-2)] is however 0.0311. The affect of y(t-1) on y(t) cannot be denied. However, it seems that y(t-2) totally dominates y(t-1) in this case. Therefore, a fair comment on the proposed technique is that it is able to find the most influential nonlinear component.

VI. CONCLUSIONS

The SNNR based technique is shown a useful tool to determine the dynamic order directly from data. It is shown that the technique has low sensitivity to noise, making it possible to deal with data from a stochastic system. In addition, the technique can also be used to determine if a system is linear or nonlinear, or extended to detect the most influential nonlinear regressors.

REFERENCES

[1] P.J. Brockwell and R.A. Davis, "Time series: theory and methods", 2nd, New York: Springer, 1998.

[2] A.J. Miller, "Subset selection in regression", New York: Chapman And Hall, 1990, pp 43-83.

[3] I. Lind and L. Liung, " Regressor and structure selection in NARX models using a structured ANOVA approach", Automatica, vol. 44, 2008, pp 383 -395.

[4] C. Molina, N. Sampson, W. J. Fitzgerald, and M. Niranjan, "Geometric techniques for finding the embedding dimension of time series", Neural Networks for signal processing [1996] VI. Proceedings of the 1996 IEEE signal processing society workshop, vol. 1, no. 4-6, Sep 1996, pp 161 - 169.

[5] C. Rhodes and M. Morari, "Determining the model order of nonlinear input/output systems directly from data", Proceedings of the American Control Conference 1995, vol. 3, no. 21-23, June 1995, pp 2190 - 2194

[6] X. He and H. Asada, "A new method for identifying orders of input-output models for nonlinear dynamic systems", Proceedings of the American Control Conference, June 1993, pp 2520 - 2523

[7] P. Young, "Recursive estimation and time-series analysis: An introduction", New York: Springer-Verlag, 1984.

[8] L. Ljung and T. Soderstrm, "Theory and practice of recursive identification", New York: MIT Press, 1987.

[9] P. Young, "Time variable and state dependent modeling of non-stationary and nonlinear time series", in Developments in time series analysis, T.S Rao, Ed. New York: Chapman and Hall, 1993, pp 374-413.

[10] L. Ljung, "System identification: Theory for the user", 2nd, New York: Prentice Hall PTR, 1999.

[11] K.S. Narendra and K. Parthasarathy, "Identification and control of dynamical systems using neural networks", IEEE transactions on neural network, vol. 1, no.1, March 1990, pp 4-27.