

# Estimation of Delay in Time Series Using ARX Model Fitting

Linjian Xiang and Qing Zhao

Department of Electrical and Computer Engineering

Univ. of Alberta, Edmonton, Alberta, Canada

**Date: June 19, 2018**

The objective of this report is to introduce a method for calculating the time-delays between time series that have dynamic relations. In a large-scale complex process, the dynamics among different time series are usually too complicated for the commonly used correlation-based time-delay estimation (TDE) to work well. For this reason, we investigate the Auto-Regression eXogenous (ARX) model identification based TDE method. In this approach, the delay between two time series is specifically modelled, the dynamic relation among the time series and their past histories are represented in a form of linear regression, and the best time-delay is found based on the estimated model. Additionally the model order is also determined. The order and delay together can be used to choose an appropriate value for the total delay shifts (i.e. the  $h$  value), an important tuning parameter in the DPCA algorithm.

## 1. Notations and Problem Formulations

For convenience, we introduce the *forward shift operator*  $q$  by

$$qu(k) = u(k+1) \quad (1)$$

and the *backward shift operator*  $q^{-1}$

$$q^{-1}u(k) = u(k-1) \quad (2)$$

Assuming that there exists a time delay of  $d$  steps between two time-series  $u(k)$  and  $y(k)$  which is introduced by the dynamic process, hence  $u(k)$  and  $y(k)$  are not only correlated but also have auto-correlations. Then the system equation relating the sequence  $u(k)$  to the sequence  $y(k)$  is

$$y(k) = g(k-d) * u(k) = \sum_{i=0}^{n-1} g(i-d)u(k-i)$$

The above equation is the discrete-time convolution sum between the delayed  $g(k)$ , the (unknown) characteristic response of the system, and the input sequence  $u(k)$ .  $d$  is the total number of delays. It can be written as

$$y(k) = \sum_{i=0}^{n-1} g(i)q^{-d}q^{-i}u(k) = \left[ \sum_{i=0}^{n-1} g(i)q^{-i} \right] q^{-d}u(k) = G(q)u(k-d) \quad (3)$$

by migrating the delay into the sequence  $u(k)$ ,  $G(q) = \sum_{i=0}^{n-1} g(i)q^{-i}$  represents the delay-free linear time-invariant model with unknown order and parameters. In this report, using the model (3), an explicit model identification method is introduced for time-delay estimation (TDE). This method is tested on various data sets, the results are included and discussed in the report.

## 2. ARX model based method

### 2.1. Delay estimation

Given model (3), note that  $G(q)$  can also be represented in a fractional form

$$G(q) = \frac{B(q)}{A(q)},$$

where  $A(q)$  and  $B(q)$  are  $n$ th order and  $m$ th order polynomial, respectively, and  $m \leq n$ ,

$$A(q) = 1 + \sum_{i=1}^n a_i q^{-i}, \quad B(q) = b_0 + \sum_{j=1}^m b_j q^{-j}$$

$a_i$  ( $i = 1, 2, \dots, n$ ) and  $b_j$  ( $j = 1, 2, \dots, m$ ) are the unknown parameters, then we have

$$A(q)y(k) = B(q)u(k-d)$$

By considering noises in the equation, the general ARX model (with delay) can be written as follows

$$A(q)y(k) = B(q)u(k-d) + e(k) \quad (4)$$

where  $e(k)$  is the Gaussian white noise.

By substituting  $A(q)$  and  $B(q)$  in, the ARX model can be expressed by

$$y(k) + a_1 y(k-1) + \dots + a_n y(k-n) = b_0 u(k-d) + \dots + b_m u(k-m-d) + e(k) \quad (5)$$

The prediction form of the above model with the unknown delay  $d$  can be given as

$$\begin{aligned} \hat{y}(k, d) = & b_0 u(k-d) + b_1 u(k-1-d) \dots + b_m u(k-m-d) \\ & - a_1 y(k-1) + \dots + a_n y(k-n) \end{aligned} \quad (6)$$

now we define the regressor matrix and the parameter vector

$$\phi(k, d) = [-y(k-1) \dots -y(k-n) \ u(k-d) \dots u(k-m-d)]^T$$

$$\theta = [a_1 \dots a_n \ b_0 \dots b_m]^T$$

Then the equation (5) can be rewritten as

$$\hat{y}(k, d) = \theta^T \phi(k, d) = \phi^T(k, d) \theta \quad (7)$$

The unknown parameter  $\theta$  can be estimated by Least-squares Estimation (LSE) directly. Then by iterating over  $d$ , the output error cost for each  $d$  can be calculated. At last the best delay estimate  $T_d$  is chosen with lowest output error cost, i.e.

$$T_d = \arg_d \left\{ \min \{ [y(k) - \hat{y}(k, d)]^T [y(k) - \hat{y}(k, d)] \} \right\} \quad (8)$$

## 2.2. Order Selection

In the above, usually  $m$  and  $n$  ( $m \leq n$ ) are unknown, and also need to be determined based on data. For simplicity, we assume  $m = n$  so that only the model order  $n$  needs to be decided. The *Akaike Information Criterion* (AIC) is used for model order selection. The AIC is considered to be a measure for the relative quality of the model for a given dataset. The AIC value of the model can be written as:

$$AIC = 2\mathbf{k} - 2\ln \mathbf{L} \quad (9)$$

Where  $\mathbf{k}$  is the number of estimated parameters in the model and the  $\mathbf{L}$  is the maximum value of the likelihood function for the model. The log-likelihood term  $-2\log \mathbf{L}$  is called the *goodness-of-fit* term, the larger *goodness-of-fit* the better AIC. On the other hand, the AIC also includes the penalty term  $2\mathbf{k}$  for the number of parameters, called bias correction, which is used to avoid model over-fitting.

The log-likelihood function for  $N$  number of values in data set is [1],

$$\ln \mathbf{L}(\mu, \sigma^2) = -\frac{N}{2} \ln(2\pi) - \frac{N}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^N (x_i - \mu)^2 \quad (10)$$

where  $\mu$  and  $\sigma$  are the mean and standard deviation, respectively. We substitute the equation (10) and the numerically calculated standard deviation  $\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$  into (9), then AIC becomes

$$AIC = 2\mathbf{k} + N \ln\left(\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}\right) + N(\ln(2\pi) + 1) \quad (11)$$

For sets of candidate model orders, the preferred order  $n$  is chosen as the one with the minimum AIC value.

### 3. Delay estimation for DPCA parameter tuning

In our project with Honeywell Process Solutions (HPS), it is proposed to use dynamic PCA (DPCA) and dynamic IPLS (DIPLS) for monitoring an industrial process. Details can be found from the previous Engage Grant report [2]. In a dynamic process, variables are time correlated. This means the value of each process variables at current time is dependent on the past histories of itself or other variables. To incorporate this feature, in DPCA an augmented step is added to form the augmented matrix  $\bar{X}_i$  ( $i$  being the  $i$ th data point) from the original  $X_i$  as  $\bar{X}_i = [X_i^T \ X_{i-\tau}^T \ \dots \ X_{i-(h-1)\tau}^T]$ .  $\tau$  is the unit interval allowing for scaling in time for the augmented matrix. It is usually chosen as  $\tau = 1$ , as we can see from commonly used difference equation, the unit step for differencing is 1, e.g. in equation (5). The parameter  $h$  is total number of sample lags (shifts). In another word, the  $h$  parameter determines the number of shifting action for generating the augmented  $\bar{X}$  matrix. By using the ARX method described above, the total number of shifts  $h$  can be determined by the order and the delay estimate in Section II. If we compare the form of  $\bar{X}$  and the regressor matrix  $\phi(k, d)$ , it is noted that  $h$  can be chosen as the maximum number of delay shifts in the regressor, which is  $\max\{n, m + d\}$ . Because we assume  $m = n$  for simplicity, then we choose  $h = n + T_d$ , the sum of the order and estimate of the delay.

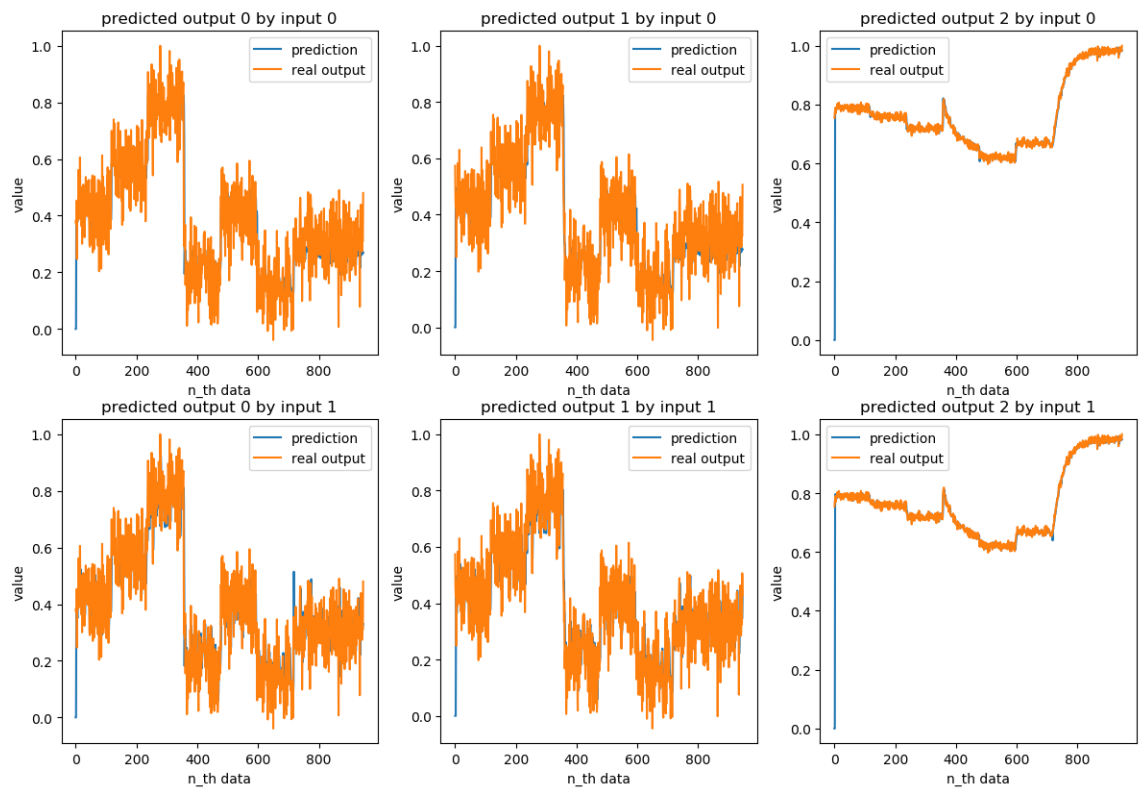
### 4. Testing Result

The TDE method is tested on the simulated CSTR and Tennessee Eastman model(TE) data. For example, the CSTR model has 3 output variables and 2 input variables. The TDE method is applied on 1000 samples of CSTR data and the ARX fitted curve is shown as figure (??). For this data set delay we got is 1, the order is 2 and the suggested  $h$  for DPCA is 3.

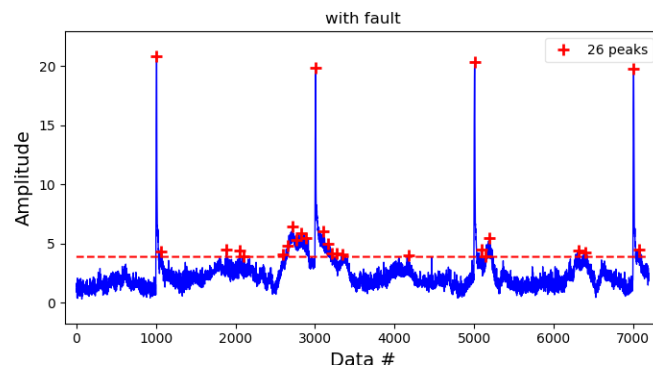
To certify the  $h$  calculated by TDE works for DPCA, we firstly applied the DPCA on non-faulted data. Then insert numbers of fault into the system model and apply DPCA on model data. At last, the  $T^2$ Hotteling method is used to identify faults.

The data from TE stripper is used for testing, the TE process manipulated variables  $XMV(4)$  and  $XMV(8)$  are used for TDE inputs and the TE process measurements  $XMEAS(15)$  to  $XMEAS(19)$  are used for TDE outputs. The fault detection between the  $h$  calculated by ARX TDE method and the  $RR - 13$ [?] method are compared and shown as figure (??) and figure (??). The  $h$  computed by ARX method is 19 and the  $h$  calculated by  $RR - 13$  is 2.

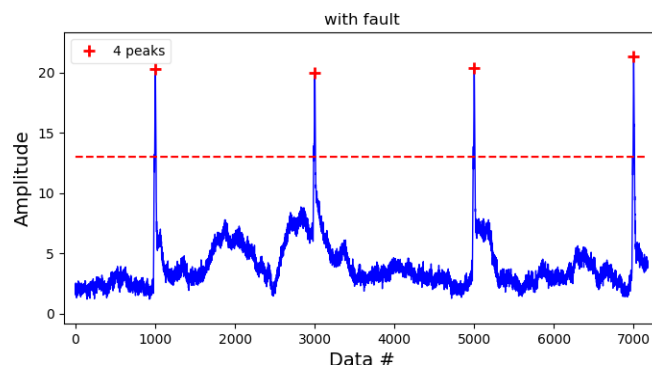
There are four faults added into the simulation, and the  $T^2$ Hotteling method were applied on these two simulation data sets. From the figures, it is clear to say that the  $h$  calculated by ARX TDE method has better performance.



**Figure 1. ARX curve fitting**



**Figure 2.  $h$  calculated by original method,  $h=2$**



**Figure 3.  $h$  calculated by ARX method,  $h=19$**

## References

- [1] M. Lovric (ed.), International Encyclopedia of Statistical Science, DOI 10.1007/978-3-642-04898-2, Springer-Verlag, Berlin, Heidelberg, 2011
- [2] B. Rashidi and Q. Zhao, "Progress Report for Phase 1", the Engage Grant Project with HPS, Sept. 2, 2016, Univ. of Alberta.
- [3] T.J. Rato, M.S. Reis, Defining the structure of DPCA models and its impact on process monitoring and prediction activities, Chemom. Intell. Lab. Syst. 125 (2013) 7486, <http://dx.doi.org/10.1016/j.chemolab.2013.03.009>.