Multivariable State-Space Identification of Dissolve Oxygen Control

Sharatul Izah Samsudin[1], Sani Irwan Md Salim[1], Mohd Fua'ad Rahmat[2], Norhaliza Abdul Wahab[2] and Mohd Hafiz Sulaiman[1]

[1]*Faculty of Electronic and Computer Engineering, Universiti Teknikal Malaysia Melaka (UTeM), Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia.*

[2]*Faculty of Electrical Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Malaysia.*

sharatul@utem.edu.my

***Abstract*—This work is proposed to identify a linear time-invariant dynamic model of dissolve oxygen (DO) of a wastewater treatment plants with multilevel pseudo random signals as an excitation input. DO is always known as a main variable in wastewater control. For this purpose, a state-space model that emphasize on numerical subspace state-space system identification (N4SID) is applied. The works include the development of perturbation input signals, Identifying the estimation model continued by validating the model performances by Variance Accounted For and mean relative squared error. It was observed that the estimated model with multilevel input offers good predicted behavior's compared to two-level pseudo random binary input signal. Benchmark Simulation Model BSM1 was applied as data generator for identification procedures.**

***Index Terms*—BSM1; Dissolve Oxygen; Multilevel Input Signal.**

I. INTRODUCTION

The modeling can be defined as a process to describe the dynamic behaviour of a system [1]. Two basic ways of modelling includes the mathematical modelling which is analytical approach that commonly use the physics law to represents the process' behavious. Another is system identification that referred to experimental approach. The experiments are performed on the system while the model is then fitted based on the data recorded [2]. The biological process of the ASP was first developed on IAWQ's Activated Sludge Model No. 1 (ASM1) [3]. It then continued by a series of mathematical models known as Activated Sludge Model No. 1 (ASM2) and Activated Sludge Model No. 3 (ASM3). Among them, the ASM1 is the most successful one used to represent the processes dynamics of the ASP [4, 5]. Undoubtly, derivation on physical behaviour of the system offering more exciting appearances, but it is clearly difficult and time consuming specifically when dealing with a large system. The direct usage of the ASM1 is difficult for control purposes since more computer intensive, hardest calibration and longer time consumption will be required [4, 6, 7]. Therefore, system identification technique becomes a good alternative in predicting the behaviour of the activated sludge.

A perturbation input signal is frequently used at the input of the identified system in obtaining more informative data [1, 2]. The perturbation signal must be persistently exciting hence the bandwidth of the perturbation signal may span with respect to the system. For a linear system, a pseudo random binary sequences (PRBS) is usually used in identification due to its similarity of white noise autocorrelation function besides easy implementation [8]. However, the PRBS input is not always well suited for the nonlinear problems [9]. With only two levels input signal, the resulting data may not provide sufficient information to identify the nonlinear behaviour. The problem is underlined where the alternative perturbation input signal is demanded in identifying a good model for greater control action. As a result, a multi-level pseudo-random sequences (MPRS) is proposed to be applied. The design methodology for MPRS proposed by Lara and Milani [9] and Lara, et al. [10] are referred in the simulation. The multilevel maximum-length pseudo-random sequences and a priori knowledge of desired system including the relevant process bandwidth are used. The advantages of multi-level input signal over PRBS for nonlinear identification have been successfully proved in pH neutralization and rapid thermal processing wafer reactor and in modulation of metal oxide micro-hotplate gas sensors [8, 11, 12].

The multilevel random signal has been used for wastewater model identification identification such presented in [9, 10, 13]. But the input signal is applied to Activated Sludge Wastewater Treatment Plant-University of Sao Paulo (ASWWTP-USP) benchmark model and not to BSM1 plant. In relation to BSM1, the multile-level input signal was used in identifying nitrogen removal process in [14]. Meanwhile, the application N4SID algorithm to BSM1 has been covered in the work by Wahab, et al. [15] and Abdul Wahab, et al. [16] but with the PRBS input signal. The importance to control the DO in activated sludge is obviously undeniable. The DO concentration should be sufficiently supplied so that adequate oxygen can be delivered to the microorganisms in the sludge. Our main objective here is to obtain a simple model which describes a dynamic behaviour of the last three aerated tank concentrations. To highlight, the work in [14] and [15] are extended to updated version of BSM1 [17] while the DO is newly identified with multi-level input signal that seems worthy of further investigation. The performance of the identified model is then investigated by the percentage of mean variance–accounted-for (MVAF) and the percentage of mean relative squared error (MRSE).

II. BENCHMARK SIMULATION MODEL NO. 1

A Benchmark Simulation Model No. 1 (BSM1) is a standard tool for modelling and simulation control for wastewater control application [18, 19]. Figure 1 shows the plant layout of the BSM1. The plant consists of five tanks where the first two compartments are anoxic tanks followed by three aerobic tanks. Each of anoxic and aerated tanks is assumed to have constant volume of 1000 m3 and 1333 m3, respectively therefore the total biological volume of the plant is 5999 m3. Besides, the effluent from the last tank is connected in series to a settler of constant volume of 6000 m3. The detail of BSM1 can be referred in [17].

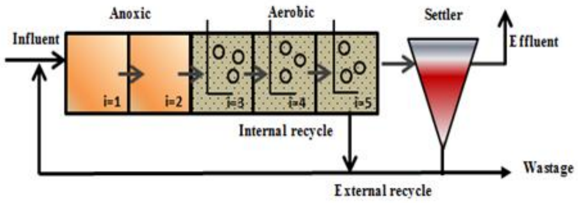

Figure 1: The plant layout of the BSM1

A. Identifying of the State-space Model

The BSM1 is used as a data generator for identification model in the ASP. Identification process is suggested to model of DO in the last three aerated tank (DO3, DO4 and DO5) concentrations. The input signals are oxygen transfer coefficient, KLa at each of aerated tanks. The model is identified under constant influent and dynamic dry influent. Specifically for dynamic dry influent, in order to obtain a good identified model, the variations of the influent flow rate, Qin, influent ammonium concentration, SNH and influent substrate, SS were included as disturbances. In addition, for a better identification result, the data were first subtracted from the sample mean as to remove the offsets. No data filtering is required since the data set was generated from the simulation model. The input-output pair and the disturbances involved in identification procedures are simplified in block diagram as shown in Figure 2.

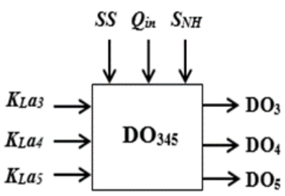

Figure 2: The block diagram of DO concentrations

The resulted identified models are then expressed as

$$\begin{aligned} x(k+1) &= Ax(k) + B_p u(k) + B_d d(k) \\ y(k) &= Cx(k) + Du(k) \end{aligned} \tag{1}$$

where $x(k)$ is the state vector, $u(k)$ is the input vector, $y(k)$ is the output vector and $d(k)$ is the measurable disturbance vector. A, B_p, B_d, C and D are matrices of appropriate dimensions. In this case, the measurable disturbances, $d(k)$ are considered as an input vector. Nevertheless, the direct D term was not considered in this study thus the Equation (2) is then rewritten as:

$$\begin{aligned} x(k+1) &= Ax(k) + \begin{bmatrix} B_p & B_d \end{bmatrix} \begin{bmatrix} u(k) \\ d(k) \end{bmatrix} \\ y(k) &= Cx(k) \end{aligned} \tag{2}$$

B. Validation of the State-space Model

In order to assess the performance of the identified model with MPRS input signal, the two validation tools referred to percentage of mean variance–accounted-for (MVAF) and percentage of mean relative squared error (MRSE) are applied. The description of MVAF and MRSE described in Equation (3) and (4).

$$\%MVAF = \frac{1}{l}\sum_{i=1}^{l}\left(1 - \frac{\mathrm{variance}(y_i - \hat{y}_i)}{\mathrm{variance}(y_i)}\right) x\,100 \tag{3}$$

y_i and $\hat{y}_i$ are the actual measured output and predicted output of the i^{th} data, respectively. Meanwhile, the number of measured and predicted data are indicated by l.

$$\%MRSE = \frac{1}{l}\sum_{i=1}^{l}\sqrt{\frac{\sum_{j=1}^{l}(y_i(j) - \hat{y}_i(j))^2}{\sum_{j=1}^{l}(y_i(j))^2}}\, x\,100 \tag{4}$$

III. RESULTS AND DISCUSSION

A. Introduction to MPRS

The MPRS input signal is developed in a similar way to the pseudo random binary sequences (PRBS) where a shift registers and modulo addition were used. As in PRBS, the MPRS input is periodic, deterministic signals and have an autocorrelation function similar to white noise. As discussed by Lara and Milani [9], the MPRS input signals are generated from sequences of maximum period in Galois Fields, $GF(q)$. q is equal to prime or a power of a prime $p(> 1)$ such as q =2,3,5,7,9,11,… The generator of the MPRS and the guidelines can be referred in [14].

B. MPRS Design

The procedures to develop the MPRS:

i. The excitation signal bandwidth, ωs that places the power of the input signal in the frequency range is calculated. The α_s and β_s that are related to the high and low frequency content were defined. Based on [12], the αs is set to 2 by assuming the closed loop speed to be twice faster than the open loop speed. To satisfy the low frequency content requirement., the $\beta s = 5$ that referred to 99% settling time is applied. Therefore, according to (2.8)-(2.10), the ωs is calculated at 1.25 rad/day $\leq \omega s \leq$ 10 rad/day.

ii. Determining the switching time, Tsw. that refers to the minimum commutation time for the signal level changes. The Tsw of the MPRS input signal of DO3, DO4 and DO5 were calculated to Tsw ≤ 0.0278 day. The Tsw is set to 0.01 day with respect to the time constant of DO (10-15 minutes).

iii. Selecting the best level of the MPRS input signal. The best of q-level MPRS was done to result a good identified model in simulation. For this task, the performances of the identified model are assessed by the percentage of mean variance–accounted-for (MVAF) as described in Equation (3). The best-identified models are presented by smaller deviations between the actual measured and the predicted outputs. The comparative performance of q-level MPRS under constant and dry influents are shown in Table 1 and 2, respectively.

The result reveals that the best signal level is identified at $q = 17$ thus the GF (17) with primitive polynomial of degree three was selected. The length of MPRS signal was calculated to 4912. The generation of the MPRS input signal is based on the generator of q-level pseudo random binary sequence as shown in Figure 3.

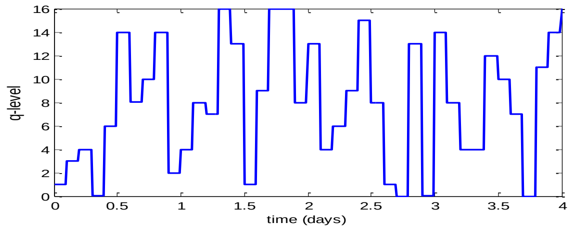

Figure 3: The MPRS input signal

C. Data Collection

The MPRS input signal is specifically designed for a multivariable identification in the simulation. Several tests have been carried out so as to adequately excite the system. The simulation ran from steady state values for all given influent files. The model was tested for two different operating conditions; steady-state input with respect to constant influent flow and dynamic input that refers to dry influent flow within 7 days' simulation time. The constant influent flow was first simulated where the three input signals in the last three aerated tanks are illustrated in Figure 4. Notice that 1/d is the unit per day where d is the day of simulation. The KLa is initially set to $84/d$ [17], but has slight enhancement due to excitation of MPRS input signal.

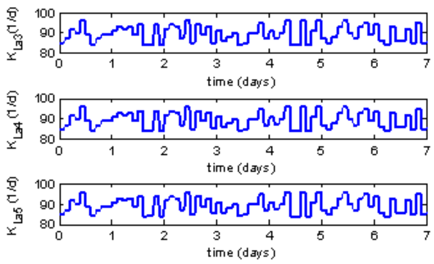

Figure 4: Input signal to activated sludge process for constant influent

Figure 5 shows the input signals applied under dry influent. In order to enhance the performance of the identified model under dry influent, three measurable disturbances include the Qin, SNH and SS presented in Figure 6 are applied. The measurable disturbances are not used in constant influent identification but only applied in dry influent.

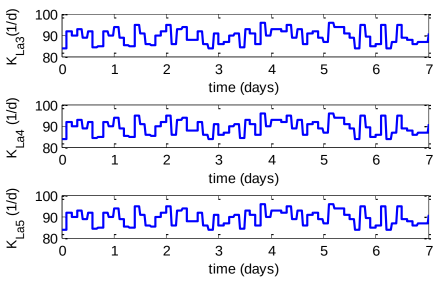

Figure 5: Input signal to activated sludge process for dry influent

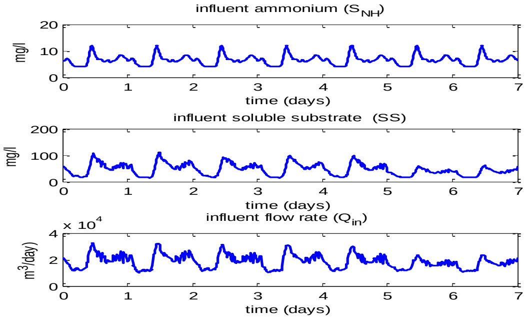

Figure 6: Measurable disturbances for dry influent flow

The identification can be performed by perturbing the plant inputs using multi-level input signal and the response on the plant output was recorded. For constant influent, the best order of the identified model was achieved at four. The estimated fourth order of the DO345 model under constant influent flow is described in Equation (5). As observed, the poles' are located at 0.5900, 0.6675, 0.7704 and 0.863 which are inside the unit circle.

$$A = \begin{bmatrix} 0.6665 & 0.0043 & -0.0045 & -0.0008 \\ -0.0511 & 0.7695 & -0.0193 & -0.0006 \\ 0.0431 & -0.0590 & 0.6281 & 0.0255 \\ -0.0969 & -0.0050 & 0.3121 & 0.8274 \end{bmatrix}$$
$$B = \begin{bmatrix} 0.0077 & -0.0002 & 0.0002 \\ 0.0021 & -0.0035 & 0.0043 \\ -0.0004 & -0.0031 & 0.0211 \\ 0.0032 & 0.0010 & -0.0215 \end{bmatrix} \tag{5}$$
$$C = \begin{bmatrix} 0.7211 & -0.1334 & 0.0545 & 0.1192 \\ 0.4521 & -1.0544 & 0.4000 & 0.4410 \\ 0.1502 & -0.3167 & 0.4010 & 0.3500 \end{bmatrix}$$
$$D = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Similarly, the resulting model for dry influent flow has the forth order as described in Equation (6). Again, the poles of the identified model were lie inside the unit circle of 0.5885, 0.7650, 0.8668 and 0.8362.

$$A = \begin{bmatrix} 0.7665 & 0.0023 & -0.0055 & -0.0018 \\ -0.0411 & 0.860 & -0.0189 & -0.0004 \\ 0.0422 & -0.0587 & 0.6301 & 0.0260 \\ -0.0988 & -0.0049 & 0.31211 & 0.7999 \end{bmatrix}$$
$$B = \begin{bmatrix} 0.0067 & -0.0003 & 0.0004 & 0.1616 & -0.0834 & 0.2431 \\ 0.0019 & -0.0042 & 0.0039 & 0.4007 & -0.4018 & -0.0357 \\ -0.0002 & -0.0051 & 0.0216 & -0.0942 & -0.1345 & -0.0028 \\ 0.0029 & 0.0010 & -0.0215 & -0.1252 & -0.0573 & 0.3869 \end{bmatrix} \tag{6}$$
$$C = \begin{bmatrix} 0.6211 & -0.1534 & 0.0550 & 0.1189 \\ 0.4519 & -1.0499 & 0.3900 & 0.4410 \\ 0.1502 & -0.3170 & 0.4010 & 0.4000 \end{bmatrix}$$
$$D = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

D. Data Validation

Figure 7 and 8 illustrate the DO3, DO4 and DO5 concentrations under constant and dry influents, respectively. The solid line denotes the real data while the dotted lines represent the predicted data. The BSM1 was simulated in 7 days while the input and the output of the plant were recorded. The identification procedure was carried out off-line with the first 4 days while the remaining 3 days were applied in validation purposes for both constant and dry influents.

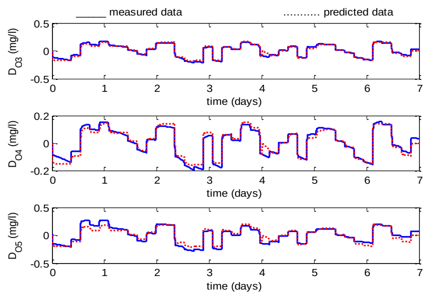

Figure 7: DO3, DO4 and DO5 concentrations for constant influent flow

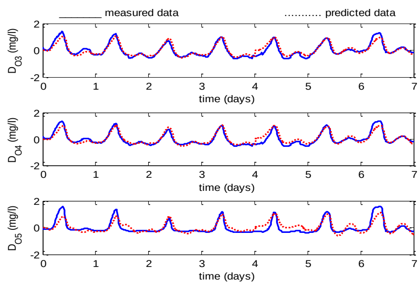

Figure 8: DO3, DO4 and DO5 concentrations for dry influent flow

The performance of the identified model in estimating the physical behaviour of the system was investigated where the model was cross-validated on the validation data. To identify the quality of the models, the MVAF and MRSE as described in Equation (3) and (4) are applied. As previously discussed, the best-identified models are indicated by smaller deviations between the actual measured and the predicted outputs with smaller relative error. The MVAF and MRSE analysis of identification and validation data sets of DO345 model are depicted in Table 1 and 2. It is observed that obvious improvement of VAF analysis was recorded with MPRS rather than PRBS under constant and dry influent.

Table 1
Validation of (a) MVAF (b) MRSE under constant influent

(a)

MVAF	PRBS (%)		MPRS (%)	
	Identification	Verification	Identification	Verification
DO_3	85.5185	85.2670	97.9148	98.1811
DO_4	95.7523	96.1288	96.6363	97.2863
DO_5	89.3583	89.0138	97.3314	96.9992

(b)

MRSE	PRBS (%)		MPRS (%)	
	Identification	Verification	Identification	Verification
DO_3	38.0935	61.1635	14.4434	13.9325
DO_4	34.4538	23.8037	8.7771	17.0816
DO_5	15.9202	47.2417	9.8525	18.1176

Table 2
Validation of (a) MVAF (b) MRSE under dry influent

(a)

MVAF	PRBS (%)		MPRS (%)	
	Identification	Verification	Identification	Verification
DO_3	84.6709	79.3092	82.7295	82.6734
DO_4	83.4075	80.7068	85.1931	89.1770
DO_5	70.6188	82.9731	86.9497	86.2059

(b)

MRSE	PRBS (%)		MPRS (%)	
	Identification	Verification	Identification	Verification
DO_3	39.1324	37.3336	35.1022	37.7335
DO_4	27.1159	37.2053	25.2445	20.2109
DO_5	38.9141	36.8033	25.4833	27.1157

IV. Conclusion

A linear state-space model that represents the dynamic natures of the ASP is identified. The activated sludge is a multivariable process with highly nonlinear variables and control parameters thus difficult to be modelled. Therefore, a system identification approach which deals with input and output data in estimating the linear model of a dynamic multivariable system has been explored. This leads to the application of numerical subspace state-space identification (N4SID). In order to obtain more informative data in signal excitation, the multi-level pseudo random signal (MPRS) was generated. It has been proven that the quality of the identified model with multi-level excitation input is better compared to a two-level input signal.

Acknowledgment

The authors would like to thank the Ministry of Higher Education (MOHE), Universiti Teknikal Malaysia Melaka (UTeM) and Universiti Teknologi Malaysia (UTM). Their support is gratefully acknowledged.

References

[1] L. Ljung, *System Identification: Theory for the User*, 2nd ed.: Prentice Hall, New Jersey, 1999.
[2] T. Söderström and P. Stoica, *System identification*: Prentice-Hall, Inc., 1988.
[3] M. Henze, C. L. Grady, W. Gujer, G. Marais, and T. Matsuo, "A General Model for Single-Sludge Wastewater Treatment Systems," *Water Research*, vol. 21, pp. 505-515, 1987.
[4] T. Yang, W. Qiu, Y. Ma, M. Chadli, and L. Zhang, "Fuzzy model-based predictive control of dissolved oxygen in activated sludge processes," *Neurocomputing*, vol. 136, pp. 88-95, 2014.
[5] J. Wu and J. Luo, "Aeration control of activated sludge wastewater treatment process using optimal control," in *Intelligent Control and Automation (WCICA), 2012 10th World Congress on*, 2012, pp. 4969-4973.
[6] P. Samuelsson, B. Halvarsson, and B. Carlsson, "Interaction analysis and control structure selection in a wastewater treatment plant model," *IEEE Transactions on Control Systems Technology*, vol. 13, pp. 955-964, 2005.
[7] A. Stare, D. Vrečko, N. Hvala, and S. Strmčnik, "Comparison of control strategies for nitrogen removal in an activated sludge process in terms of operating costs: a simulation study," *Water research*, vol. 41, pp. 2004-2014, 2007.
[8] M. Braun, D. Rivera, A. Stenman, W. Foslien, and C. Hrenya, "Multi-level pseudo-random signal design and" model-on-demand" estimation applied to nonlinear identification of a RTP wafer reactor," in *American Control Conference, 1999. Proceedings of the 1999*, 1999, pp. 1573-1577.
[9] J. M. Lara and B. Milani, "Identification of neutralization process using multi-level pseudo-random signals," in *Proceedings of the American Control Conference*, 2003, pp. 3822-3827.

[10] J. M. Lara, B. E. Milani, and O. A. Sotomayor, "Adaptive predictive control of ammonium level based on dissolved oxygen variable set-point control in an activated sludge process," in *Mercosur Congress on Process Systems Engineering*, 2005.

[11] A. Vergara, E. Llobet, J. Brezmes, P. Ivanov, X. Vilanova, I. Gracia*, et al.*, "Optimised temperature modulation of metal oxide micro-hotplate gas sensors through multilevel pseudo random sequences," *Sensors and Actuators B: Chemical,* vol. 111, pp. 271-280, 2005.

[12] A. Vergara, E. Llobet, J. Brezmes, P. Ivanov, C. Cané, I. Gracia*, et al.*, "Quantitative gas mixture analysis using temperature-modulated micro-hotplate gas sensors: Selection and validation of the optimal modulating frequencies," *Sensors and Actuators B: Chemical,* vol. 123, pp. 1002-1016, 2007.

[13] O. Sotomayor, S. Park, and C. Garcia, "Subspace-based optimal control of N-removal activated sludge plants," in *Proceedings of the 2001 IEEE International Conference on Control Applications (CCA'01)*, 2001, pp. 588-593.

[14] S. I. Samsudin, M. Fuarsquo, N. A. Wahab, M. S. Gaya, and I. W. Jamaludin, "Multilevel input signal for multivariable state-space identification of wastewater treatment plant," *International Journal of Physical Sciences,* vol. 7, pp. 1847-1856, 2012.

[15] N. A. Wahab, R. Katebi, and J. Balderud, "Multivariable PID control design for activated sludge process with nitrification and denitrification," *Biochemical Engineering Journal,* vol. 45, pp. 239-248, 2009.

[16] N. Abdul Wahab, M. Katebi, and J. Balderud, "Multi-variable PID control design for wastewater system," in *15th IEEE Mediterranean Conference on Control and Automation,(MED07), 27-29 June 2007*, 2007, pp. 1-6.

[17] K. Gernaey, L. DTU, D. U. Jeppsson, and S. Winkler, "Benchmark simulation model no. 1 (BSM1)," 2008.

[18] J. Copp, "COST Action 624: The COST Simulation Benchmark," *Description and Simulation Manual Office for Official Publications of the European Communities*, 2002.

[19] K. V. Gernaey, J. B. Copp, P. A. Vanrolleghem, and U. Jeppsson, "Lessons learned from the WWTP benchmarking exercise," in *Proceedings 11th IWA Conference on Instrumentation, Control and Automation (ICA2013)*, 2013, pp. 18-20.