System identification of the fermentation system of *Thermoanaerobacter* sp. X514

Jing Yang, Xiaofang Ling, Lishan Yao
The Key Laboratory of Biofuels,
Qingdao Institute of Bioenergy and Bioprocess Technology,
Chinese Academy of Sciences, P.R.China.
Email: yangjing@qibebt.ac.cn

Hualiang Wei, Visakan Kadirkamanathan
Department of Automatic Control & Systems Engineering,
University of Sheffield,
Sheffield, South Yorkshire, United Kingdom.
Email: w.hualiang@sheffield.ac.uk

***Abstract*—Bioethanol production by means of anaerobic thermophilic microorganisms with pentose or hexose as the substrate are of paramount importance in sustainable fuel innovation. Manipulation of microorganisms and the associated experiment conditions by means of various ad-hoc technology is obviously the most straightforward way with the aim of maximizing bioethanol yield. However, methodology by means of mathematical modeling and analysis is often neglected among these routines. In this paper, typical input-output models are applied in the metabolic system analysis of *Thermoanaerobacter* sp. X514 under sole glucose substrate, sole xylose substrate and mixed glucose and xylose substrates conditions. Orthogonal Least Squares (OLS) approach is used for model parameter estimation. Model selection is proposed in order to testify the generality of the suggested model. System identification results illustrate that various forms of Nonlinear AutoRegressive with eXogenous input models (NARX) are applicable in delineating the system where different substrates (glucose or xylose) are utilized during the experiments. The proposed model structure infers that the yields of various products in X514 are mainly driven by the history information of the substrate consumption change. Moreover, the interaction between the main fermentation products of X514 is indirectly connected through the proposed models.**

***Index Terms*—*Thermoanaerobacter* sp., system identification, model estimation, bioethanol**

I. Introduction

System identification and model estimation [1] have long been viewed as common approaches in various applications, and their effects on prediction and control of system behavior are well recognized [2], [3], [4]. Biological systems present complicated whereas regulated behavior thanks to specific genetic codes underlying each individual identity. Understanding biological systems by means of traditional system identification and modeling approaches is of paramount importance in post-genomic era.

The simple yet linear in nature model is stoichiometric model [5], which connects metabolic reaction rates with extracellular metabolites by means of stoichiometric balances. With the basis on biochemical mass balance laws, model parameters are therefore already known, which elegantly avoids the high computational efforts required in system identification steps. Stoichiometric model has been popular for over decades, mainly thanks to the high interest in metabolic pathway optimization as well as metabolic flux quantification[6], [7], [8], [9]. However, the simplicity comes with a price. The model itself is unable to elucidate regulation and manipulation as well as positive or negative feedback between genotype and phenotype. Nonlinear nature underlying the complex mechanism therefore requires models capable of accounting for nonlinear effects among them.

Typical modeling approach in biological system analysis is formulated by a set of coupled ordinary differential equations, which intend to delineate the relationship between networks, metabolites and reaction fluxes involved [10]. The conceptual framework enables these models to be further classified as generalized mass-action-based models [11], [12] and power-law models [13], [14]. Both model types involve a good number of parameters, which obviously incur system identifiability, observability and parameter estimation therein. The most straightforward model is formulated by Michaelis-Menten rate laws and their generalizations, delineating both the inhibitory and activating effects between reactions among a biological network. [15] is one typical example of these model types, where the central metabolic network of *Saccharomyces cerevisiae* was constructed upon the basis of 23 Michaelis-Menten rate equations. Over 80 parameters were involved in these rate equations, whose values mostly originated from published literatures. The proposed model predicted both the stationary and time-dependent metabolic states under *in vivo* condition during glucose pulse experiments. However, as for a 'younger' microorganism whose study history is far much shorter than that of the well-known *Saccharomyces cerevisiae*, in very rare case can all of the parameters in its Michaelis-Menten rate formats be handily available. Parameter estimation is therefore often encountered in models based upon Michaelis-Menten rate laws. However, most of the Michaelis-Menten reaction rate laws are nonlinear and therefore incur high computational efforts during system identification and parameter optimization. On contrary to cumbersome Michaelis-Menten nonlinear model, S-system model [13], [16] has a much regular yet flexible model format. A typical S-system model has the following form:

$$\frac{dX_i}{dt} = a_i \prod_{k=1}^{m} X_{ki}^{p_k} - b_i \prod_{l=1}^{n} X_{il}^{q_l} \tag{1}$$

where X_i is concentration of the i_{th} metabolite in a biological

978-1-4673-4398-5/12/$31.00 ©2012 IEEE

network. X_{ki} represents the k_{th} ancestor of metabolite X_i, whereas X_{il} represents the l_{th} descendant of X_i. m and n are corresponding number of the ancestors and descendants of X_i, respectively, a_i and b_i are coefficients of the model. p_k and q_l are kinetic orders, respectively. S-system model has virtually the capability of capturing any nonlinearity in a system. Although its format can be presented easily for any biological system, parameter estimation is again encountered in S-system identification. Algorithms, for example, genetic algorithm [17] and neural network [18] have been applied in parameter estimation of a S-system. However, huge computational efforts and questioning of the parameter fitting results keep exist. Lin-log and log-lin models [19] provide another flexible representation of a biological system. Both of the two models (whose formulations are not shown here for sake of brevity) examine these states affinity to a specified reference state. Both lin-log or log-lin and S-system models are special cases of BST (Biochemical Systems Theory) [20].

In aforementioned models, linear stoichiometric model is easy to formulate and perform any followed predictions. However, lack of information about system dynamics is its major disadvantage. S-system, lin-log and log-lin approaches all require huge computational efforts and delicately constructed optimization algorithms to carry out model parameter estimation. Their results are often questioned due to the fact that the feasibility of these algorithms is still limited. With the aim of metabolic engineering in mind, an intuitive thought is then to investigate the impact of substrates on the synthesized products, whereas, at the same time, account for the reverse effect of major and side products on substrate consumption. Traditional input-output data-based modeling as well as the feedback system analysis methodology can be adopted here seamlessly. Unfortunately, very rare research has been conducted concerning this topic.

In this paper, *Thermoanaerobacter* sp. X514, which was first discovered as a metal-reducing thermophilic anaerobe [21], was taken as an example. X514 can ferment pentose and hexose to produce ethanol at a simultaneous manner. However, only few research work was done on X514 [22], [23], [24], not to mention the sophisticated link between substrate consumption and product formation. Adopting glucose or xylose as the feeding substrates, the central objective of this work is to propose an empirical and data-based input-output modeling approach which can produce an accurate whereas simple description of the relationship between glucose or xylose, and the main products of X514, including ethanol, acetate and lactate.

The paper is organized as follows: in Section 2, experiment setup and the measuring procedure of the fermentation experiment are described; in Section 3, the model structure, model parameter estimation, model selection and validation steps are presented at a detailed manner. Section 4 presents and compares the fitting results and the model validation results of the proposed methodology. Section 5 concludes the paper and presents future prospectives.

II. Experiments and Models

A. Material and Methods

X514 (ATCC BAA-938) was received from Institute for Environmental Genomics, University of Oklahoma, United States. X514 was incubated under a N_2 (99.99%) headspace in a flask in the mineral medium which was a modified version of the one used for *Thermoanaerobacter pseudoethanolicus* ATCC 33223 by Wiegel [25] at 60°C. The mineral medium contained (per litre of distilled water) 4.2 g $Na_2HPO_4{\cdot}12H_2O$, 1.5 g KH_2PO_4, 1.0 g NH_4Cl, 0.2 g $MgCl_2{\cdot}6H_2O$, 10.0 ml Wolfe's mineral solution [26], 1.0 ml 0.1% (m/v) resazurin. The mineral medium was autoclaved for 20 minutes at 115°C. Reducing agents which were composed of 5 g/l $Na_2S_2O_4$ and 3 g/l $NaHCO_3$ were injected into the medium by a syringe after the basal medium was sterilized and cooled to room temperature. glucose or xylose was injected into the mineral medium at the start of the fermentation. All chemicals were purchased from Amresoco, USB, Sigma or Sangon with high purity.

Three experiments were carried out here with the substrates being sole glucose, sole xylose, mixed glucose and xylose, respectively. The initial concentrations of glucose and xylose were both 2 g/L under the sole substrate conditions, whereas their concentrations in the mixed substrates condition were 1 g/L each. The sampling frequency is 4 hours per sample. The experiments lasted for 60 hours. Concentrations of the substrates, i.e., glucose or xylose, and the two main products, acetate and lactate, were all measured by ionic chromatography ics-3000 (Dionex, USA). Concentration of the major product, ethanol, were measured by Megazyme Enzyme Kit (Ireland). All experiments were performed in triplicate cultures.

B. Data-based Modeling Methodology

It has been proved that under some mild conditions a discrete-time or discretized continuous-time dynamical system can be described by the following difference equation model

$$\begin{aligned} y(n) &= f(y(n-1),\cdots,y(n-q),u(n-1),\cdots,u(n-p), \\ &\quad e(n),\cdots,e(n-t)) \end{aligned} \tag{2}$$

where $y(n)$, $u(n)$ and $e(n)$ are system output, input and noise at the n_{th} time step, respectively. p, q and t are the maximum lags in the input, output and noise, respectively, and $f(\cdot)$ is some unknown linear or nonlinear mapping between the input and output signal. In terms of the linearity or nonlinearity nature of the model, the model is termed AutoRegressive with eXogenous inputs model (ARX), AutoRegressive Moving Average with eXogenous inputs model (ARMAX) [27] or Nonlinear AutoRegressive Moving Average with eXogenous inputs model (NARMAX) [28], [29].

A general form of Eq. (2) with a nonlinear degree order d is illustrated below, with the assumption that system noise is independent identically distributed noise sequence $e(n)$,

978-1-4673-4398-5/12/$31.00 ©2012 IEEE

$$
\begin{aligned}
y(n) &= k_0 + \sum_{i=1}^{r} k_i x_i(n) + \sum_{i=1}^{r}\sum_{j=1}^{r} k_{ij} x_i(n) x_j(n) + \cdots \\
&\quad + \sum_{i_1=1}^{r}\sum_{i_2=1}^{r}\cdots\sum_{i_d=1}^{r} k_{i_1 i_2 \cdots i_l} x_{i_1}(n) x_{i_2}(n)\cdots x_{i_d}(n) \\
&\quad + e(n) \qquad (3)
\end{aligned}
$$

where

$$
\begin{aligned}
r &= p + q \\
x_i(n) &= \begin{cases} y(n-i) & 1 \le i \le q \\ u(n-i+q) & q+1 \le i \le p+q \end{cases} \qquad (4)
\end{aligned}
$$

When none output terms are involved in the model shown above, the time series chain is termed volterra series [30]. Though theoretically many modern terms can be involved in the model above, in practice, only a few model terms are needed. In this paper, a simplified model format is utilized, which is shown in Eq. (5)

$$
y_l(k) = a_{0l} + \sum_{i=1}^{m} a_{il}\phi_i + e(k) \qquad (5)
$$

Here m is the number of model terms, a_i is the corresponding model coefficient for the i_{th} model term, and $y_l(k)$ is the representation of the k_{th} sample of the l_{th} product, ie, ethanol, lactate and acetate, respectively. Model order of 3 is enough to incorporate most of the situation, therefore, the system model order is set as 3 in this paper. Each ϕ_i is a combination of the lagged versions of the input and output variables of the system, whose term comes from the dictionary Φ with the system model order of 3,

$$
\begin{aligned}
\Phi &= \{1\} \cup \{x_i(n) : 1 \le i \le 3\} \\
&\quad \cup \{x_i(n)x_j(n) : 1 \le i < j \le 3\} \\
&\quad \cup \{x_i(n)x_j(n)x_l(n) : 1 \le i < j < l \le 3\} \qquad (6)
\end{aligned}
$$

with $x_i(n)$ defined as in Eq. (4).

C. Model Structure and Parameter Estimation

Though ordinary least squares approach [31] can be used to perform parameter estimation when the model structure is already known *a prior* with the assumption that the system noise is in Gaussian, it is not applicable in the situation where model structure and model parameters both require estimation. A common approach to overcome the dilemma is by means of Orthogonal Least Squares (OLS) method [32], which repeatedly estimates the model structure and parameters associate with the specific structure in a forward regression manner. OLS operates by representing a linear-in-the-parameter model illustrated below

$$
y(t) = \sum_{i=1}^{P} a_i \phi_i(t) + e(t) \qquad (7)
$$

with an auxiliary model given as

$$
y(t) = \sum_{i=1}^{P} b_i w_i(t) + e(t) \qquad (8)
$$

where P is the number of candidate model terms involved in the model and N is the number of data. When $i \neq j$

$$
\sum_{t=1}^{N} w_i(t) w_j(t) = 0 \qquad (9)
$$

Assume that

$$
w_0(t) = \phi_0(t) = 1 \qquad (10)
$$

$w_i(t)$ is then

$$
w_i(t) = \phi_i(t) - \sum_{j=1}^{i-1} b_{ji} w_j(t) \qquad (11)
$$

$i = 1, \cdots, P$ and

$$
a_{ji} = \frac{\sum_{t=1}^{N} \phi_i(t) w_j(t)}{\sum_{t=1}^{N} w_j^2(t)} \qquad (12)
$$

The parameter b_i can then be derived

$$
b_i = \frac{E[y(t) w_i(t)]}{E[w_j^2(t)]} \qquad (13)
$$

Error Reduction Ratio (ERR) provides the criteria for model candidate terms selection,

$$
ERR_i = \frac{g_i^2 E[w_i^2(t)]}{E[y^2(t)]} \qquad (14)
$$

with

$$
g_i = \frac{E[y(t) w_i(t)]}{E[w_j^2(t)]} \qquad (15)
$$

with predefined threshold η and θ, the criteria for selecting the candidate model terms are then,

$$
1 - \sum_{i=1}^{\eta} ERR_i < \theta \qquad (16)
$$

where θ is the desired tolerance for the model terms. The larger the ERR_i is, the more significant the model term is.

D. Model selection

Since multiple models are suggested above, model selection is a critical step in determining which model(s) fits the data in an satisfactory level. Several criteria, for example Akaike Information Criteria (AIC) [33] and the Bayesian Information Criteria (BIC) [34] can be applied in model selection for both linear and nonlinear models. In this paper, AIC with the following format is used, whose formulation is given below,

$$
\mathtt{AIC} = 2k - 2ln(L) \qquad (17)
$$

978-1-4673-4398-5/12/$31.00 ©2012 IEEE

where k is the number of the model parameters, L is the likelihood of the model estimates. In the paper, the measurement noise is assumed Gaussian distribution, L can then be represented below,

$$L = \prod_{i=1}^{m} \sqrt{\frac{1}{2\pi\sigma^2}} \exp\left(-\sum_{i=1}^{m} \frac{(\hat{y_l}(i) - y_l(i))^2}{2\sigma^2} \right) \quad (18)$$

where $\hat{y_l}(i)$ is the estimated result of the i_{th} product, $y_l(i)$, σ is the standard deviation of the corresponding measurement.

III. Results

The measurement data and the Model Predicted Outputs (MPOs) for the three fermentation systems with different substrates are illustrated in Fig. 1, 2 and 3, respectively. Please note that MPO utilized here is different from one step predicted output, where for the former case, the model predictions obtained in each step are adopted to predict the estimates in next step, whereas the latter case uses original model measurements to predict the estimates in each step. Therefore, even a very bad model could sometimes produce nice fitting results in the latter case. However, MPO encompasses predicted error accumulation, therefore, is more robust and reliable.

The corresponding functional forms of the three conditions are illustrated in Eq. (19), (20) and (21), respectively. From these models, it is clear that the studied systems are all nonlinear and the history consumptions of the substrate(s) are the only limiting factor of these studied products. However, it is not sensible to conclude that there is no interaction between different products since the interaction can be indirectly expressed through the input substrate.

$$\begin{aligned} y_1(k) &= 9.82 - 0.95u(k-2) + 0.57u^2(k-2) \\ y_2(k) &= 5.29 - 0.52u(k-2) + 0.03u^2(k-2) \\ y_3(k) &= 2.85 - 0.24u(k-2) + 0.01u^2(k-2) \end{aligned} \quad (19)$$

$$\begin{aligned} y_1(k) &= -0.38u(k-1) + 2.34u(k-2) \\ &\quad -0.15u^2(k-2) \\ y_2(k) &= 0.94u(k-2) - 0.07u^2(k-2) \\ y_3(k) &= 0.34u(k-2) - 0.01u^2(k-1) \end{aligned} \quad (20)$$

$$\begin{aligned} y_1(k) &= 8.89 - 0.39u2(k-3) - 0.04u_1^3(k-1) \\ y_2(k) &= 2.33 + 0.32u_2(k-3) \\ y_3(k) &= 0.67u_2(k-3) + 0.13u_1(k-3)u_2(k-3) \\ &\quad -0.03u_2(k-2)u_2^2(k-3) \end{aligned} \quad (21)$$

IV. Conclusion and Future Prospects

This paper proposes an input-output model, which seamlessly links substrate consumption of X514 to its product formation. Detailed model format and associated parameters indicate underlying biological meaning between substrate and product of X514. It is concluded that the yields of the fermentation products of X514 are mainly driven by the substrate consumptions. Meanwhile, there are strong interactions between these products. Similar model structure can be applied in both the case where the substrate is glucose and where the substrate is xylose, which indicates the metabolic fermentation similarity of X514 when glucose and/or xylose are substrates.

Mathematical modeling is of key importance in biological system analysis, however, how to accommodate more general conditions as well as more complex interactions are key problems in this area. Future work should be focused on finding out models which is applicable and robust in extensive conditions. Furthermore, bioethanol optimization by means of mathematical model should also be considered in future research study.

Acknowledgment

This work is financially supported by National Natural Science Foundation of China (NSFC) and Royal Society, United Kingdom [grant numbers 30800202, 31011130157] and Scientific Research Foundation for the Returned Overseas Chinese Scholars, respectively. Professor Jizhong Zhou from Institute for Environmental Genomics, University of Oklahoma, United States, provided the strain Thermoanaerobacter sp. X514 used in this work. We also thank Prof Jian Xu, Prof. Qiu Cui, Dr. Chenggang Xu and Ms Yuetong Ji for their helps and useful discussions during the work.

References

[1] L. Ljung and E. Ljung, *System identification: theory for the user*. Prentice-Hall Englewood Cliffs, NJ, 1987, vol. 11.

[2] G. Kerschen, K. Worden, A. Vakakis, and J. Golinval, "Past, present and future of nonlinear system identification in structural dynamics," *Mechanical Systems and Signal Processing*, vol. 20, no. 3, pp. 505–592, 2006.

[3] W. Favoreel, B. De Moor, and P. Van Overschee, "Subspace state space system identification for industrial processes," *Journal of Process Control*, vol. 10, no. 2-3, pp. 149–155, 2000.

[4] C. Griesinger, G. Otting, K. W
"uthrich, and R. Ernst, "Clean tocsy for proton spin system identification in macromolecules," *Journal of the American Chemical Society*, vol. 110, no. 23, pp. 7870–7872, 1988.

[5] G. Smolders, J. Van der Meij, M. Van Loosdrecht, and J. Heijnen, "Stoichiometric model of the aerobic metabolism of the biological phosphorus removal process," *Biotechnology and bioengineering*, vol. 44, no. 7, pp. 837–848, 1994.

[6] J. Pramanik and J. Keasling, "Stoichiometric model of escherichia coli metabolism: Incorporation of growth-rate dependent biomass composition and mechanistic energy requirements," *Biotechnology and bioengineering*, vol. 56, no. 4, pp. 398–421, 1997.

[7] C. Schilling, S. Schuster, B. Palsson, and R. Heinrich, "Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era," *Biotechnology progress*, vol. 15, no. 5, pp. 296–303, 1999.

[8] U. Sauer, "Metabolic networks in motion: 13c-based flux analysis," *Molecular Systems Biology*, vol. 2, no. 1, 2006.

[9] J. Yang, S. Wongsa, V. Kadirkamanathan, S. Billings, and P. Wright, "Metabolic flux estimation-a self-adaptive evolutionary algorithm with singular value decomposition," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, pp. 126–138, 2007.

[10] H. Resat, L. Petzold, and M. Pettigrew, "Kinetic modeling of biological systems," *Methods Mol Biol*, vol. 541, pp. 311–335, 2009.

[11] A. Marín-Sanguino and N. Torres, "Optimization of biochemical systems by linear programming and general mass action model representations," *Mathematical biosciences*, vol. 184, no. 2, pp. 187–200, 2003.

978-1-4673-4398-5/12/$31.00 ©2012 IEEE

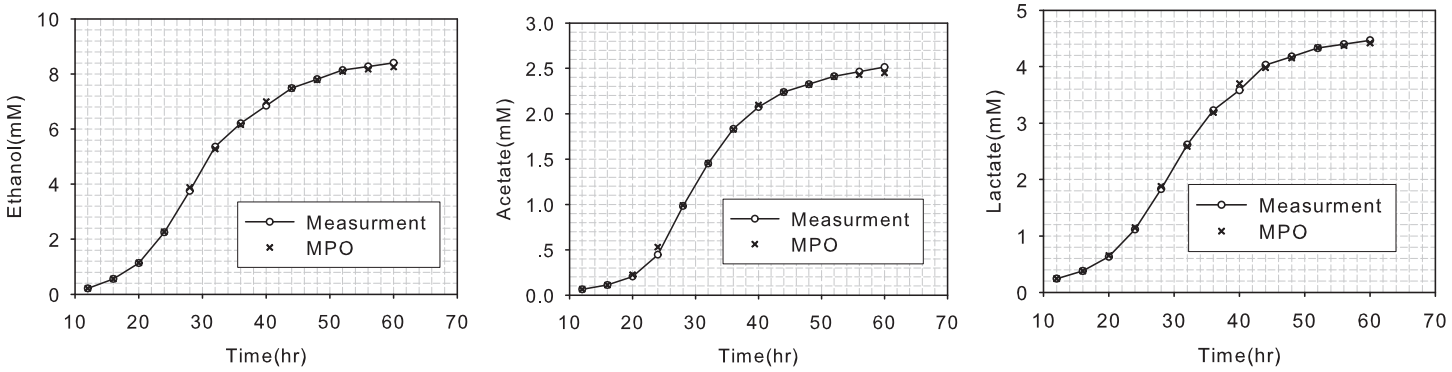

Fig. 1. Comparison of the MPOs and original measurements of X514 fermentation with glucose being the sole substrate

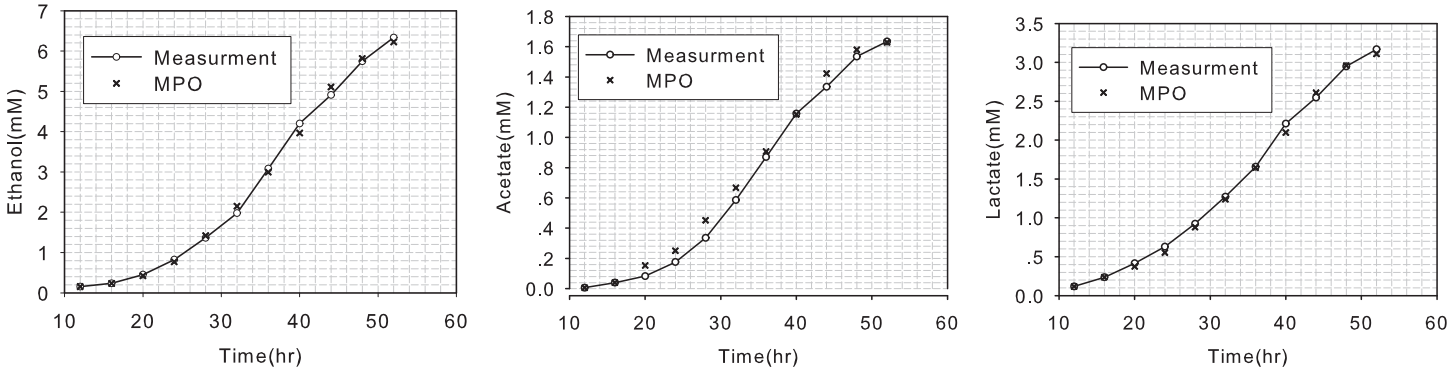

Fig. 2. Comparison of the MPOs and original measurements of X514 fermentation with xylose being the sole substrate

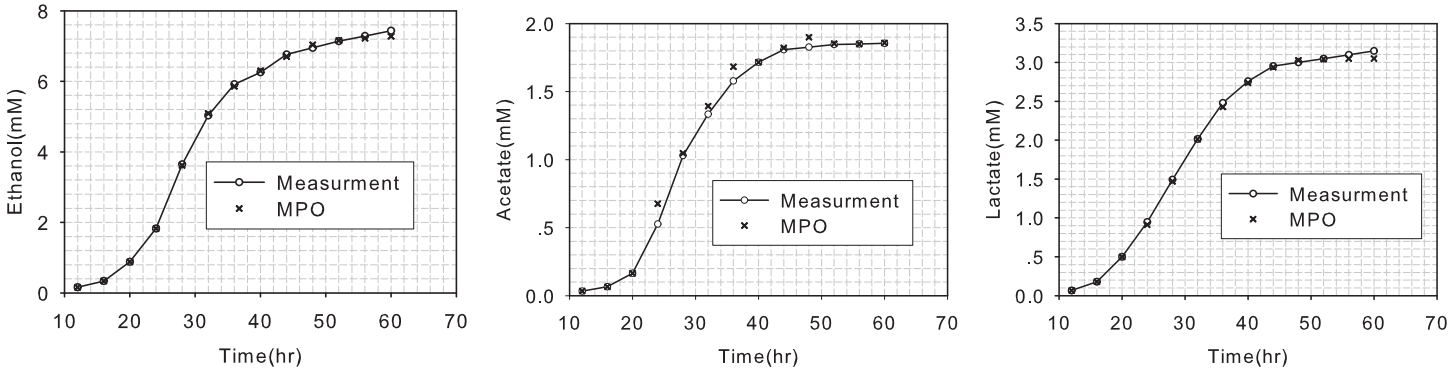

Fig. 3. Comparison of the MPOs and original measurements of X514 fermentation utilizing mixed substrates of glucose and xylose

[12] W. Tucker, Z. Kutalik, and V. Moulton, "Estimating parameters for generalized mass action models using constraint propagation," *Mathematical biosciences*, vol. 208, no. 2, pp. 607–620, 2007.

[13] S. Kimura, K. Ide, A. Kashihara, M. Kano, M. Hatakeyama, R. Masui, N. Nakagawa, S. Yokoyama, S. Kuramitsu, and A. Konagaya, "Inference of s-system models of genetic networks using a cooperative coevolutionary algorithm," *Bioinformatics*, vol. 21, no. 7, p. 1154, 2005.

[14] M. Vilela, I. Chou, S. Vinga, A. Vasconcelos, E. Voit, and J. Almeida, "Parameter optimization in s-system models," *BMC systems biology*, vol. 2, no. 1, p. 35, 2008.

[15] M. Rizzi, M. Baltes, U. Theobald, and M. Reuss, "In vivo analysis of metabolic dynamics in saccharomyces cerevisiae: Ii. mathematical model," *Biotechnology and bioengineering*, vol. 55, no. 4, pp. 592–608, 1997.

[16] S. Kikuchi, D. Tominaga, M. Arita, K. Takahashi, and M. Tomita, "Dynamic modeling of genetic networks using genetic algorithm and s-system," *Bioinformatics*, vol. 19, no. 5, p. 643, 2003.

[17] P. Liu and F. Wang, "Inference of biochemical network models in s-system using multiobjective optimization approach," *Bioinformatics*, vol. 24, no. 8, p. 1085, 2008.

[18] J. Almeida and E. Voit, "Neural-network-based parameter estimation in s-system models of biological networks," *GENOME INFORMATICS*

978-1-4673-4398-5/12/$31.00 ©2012 IEEE

SERIES, pp. 114–123, 2003.
[19] L. Wu, W. Wang, W. Van Winden, W. Van Gulik, and J. Heijnen, "A new framework for the estimation of control parameters in metabolic pathways using lin-log kinetics," *European Journal of Biochemistry*, vol. 271, no. 16, pp. 3348–3359, 2004.
[20] F. Wang, C. Ko, and E. Voit, "Kinetic modeling using s-systems and lin-log approaches," *Biochemical engineering journal*, vol. 33, no. 3, pp. 238–247, 2007.
[21] Y. Roh, S. Liu, G. Li, H. Huang, T. Phelps, and J. Zhou, "Isolation and characterization of metal-reducing thermoanaerobacter strains from deep subsurface environments of the piceance basin, colorado," *Appl Environ Microb*, vol. 68, no. 12, p. 6013, 2002.
[22] L. Lin, H. Song, Y. Ji, Z. He, Y. Pu, J. Zhou, and J. Xu, "Ultrasound-mediated DNA transformation in thermophilic gram-positive anaerobes," *PloS one*, vol. 5, no. 9, p. e12582, 2010.
[23] Q. He, P. Lokken, S. Chen, and J. Zhou, "Characterization of the impact of acetate and lactate on ethanolic fermentation by *Thermoanaerobacter ethanolicus*," *Bioresource Technol*, vol. 100, no. 23, pp. 5955–5965, 2009.
[24] X. Feng, H. Mouttaki, L. Lin, R. Huang, B. Wu, C. Hemme, Z. He, B. Zhang, L. Hicks, J. Xu *et al.*, "Characterization of the central metabolic pathways in *Thermoanaerobacter* sp. strain X514 via isotopomer-assisted metabolite analysis," *Appl. Environ. Microbiol.*, vol. 75, no. 15, p. 5001, 2009.
[25] J. Wiegel, L. Ljungdahl, and J. Rawson, "Isolation from soil and properties of the extreme thermophile *Clostridium thermohydrosulfuricum*." *J Bacteriol*, vol. 139, no. 3, p. 800, 1979.
[26] E. Wolin, M. Wolin, and R. Wolfe, "Formation of methane by bacterial extracts," *J Biol Chem*, vol. 238, no. 8, p. 2882, 1963.
[27] K. Brewer, "Some consequences of temporal aggregation and systematic sampling for arma and armax models," *Journal of Econometrics*, vol. 1, no. 2, pp. 133–154, 1973.
[28] S. Chen and S. Billings, "Representation of non-linear systems: the narmax model," *International Journal of Control*, vol. 49, no. 3, pp. 1012–1032, 1989.
[29] S. Billings and H. Wei, "The wavelet-narmax representation: A hybrid model structure combining polynomial models with multiresolution wavelet decompositions," *International journal of systems science*, vol. 36, no. 3, pp. 137–152, 2005.
[30] R. Brockett, "Volterra series and geometric control theory," *Automatica*, vol. 12, no. 2, pp. 167–176, 1976.
[31] C. Dismuke and R. Lindrooth, "Ordinary least squares," *Methods and Designs for Outcomes Research*, p. 93, 2006.
[32] S. Chen, S. Billings, and W. Luo, "Orthogonal least squares methods and their application to non-linear system identification," *International Journal of Control*, vol. 50, no. 5, pp. 1873–1896, 1989.
[33] H. Akaike, "Likelihood of a model and information criteria," *Journal of Econometrics*, vol. 16, no. 1, pp. 3–14, 1981.
[34] G. Schwarz, "Estimating the dimension of a model," *The annals of statistics*, pp. 461–464, 1978.

978-1-4673-4398-5/12/$31.00 ©2012 IEEE