Desalination and Water Treatment
www.deswater.com

35 (2011) 110–117
November

doi: 10/5004/dwt.2011.3137

Comparison of phenomenological membrane bioreactor activated sludge biological models with alternative versions based on time series input-output approaches

Parneet Paul

Room 252, Howell Building, School of Engineering and Design, Brunel University, Uxbridge, Middlesex UB8 3PH, United Kingdom
Tel. +44 1895 265435, Ext. 65435; email: Parneet.Paul@brunel.ac.uk

Received 19 November 2010; Accepted 15 July 2011

ABSTRACT

Data was taken from a full scale sidestream membrane bioreactor (MBR) plant that has been treating industrial wash water from a salad processing factory based in Worcestershire. Various online and offline measurements were taken during an intensive sampling period for the complete flow train. Offline tests included measuring extra-cellular polymeric substance (EPS) and soluble microbial product (SMP) levels in the sludge water which are the main foulants on the associated membrane [1]. Two modified phenomenological Activated Sludge models based upon the ASM1 and ASM3 that included EPS/SMP concentrations were calibrated and validated using these offline and online measurements [2–6]. In order to see whether a simpler model structure could be formulated for advanced control purposes that was based wholly upon measured historical data sets, further conceptual models were developed based on system identification procedures and input-output times series analysis methods [7]. These model forms utilised autoregressive, state-space and subspace formulations, and were calibrated and validated in Matlab© using the collated plant data. A poor model fit was shown for the modified Oliveira and Lu Activated Sludge models, although their parameters were changed from the original models' values to reflect an MBR system [5,6]. In comparison the subspace and ARX formulated biological models were reasonably accurate when compared to the Activated Sludge version, although a much longer historical data set is needed to confirm these initial findings.

Keywords: Wastewater; MBR; Activated sludge; EPS; SMP; Modelling; System identification; Time series

1. Introduction

This research work uses phenomenological models based on membrane bioreactor (MBR) biochemical processes to measure the effectiveness of alternative time series input-output (IO) models based upon system identification methods. Both model types are calibrated an d validated using a similar plant layout and data set derived for this purpose. The focus of this research is to create practical MBR computer models which can then be applied in MBR design, control and optimisation [8].

The eventual purpose of this study is to create an integrated biological and membrane filtration model based on IO modelling methods to accurately predict fouling propensity of the mixed liquor on the membranes. This fouling propensity is determined by various factors such

as biological conditions, operational/cleaning regimes, environmental factors (e.g., temperature), and influent conditions. Primary among these factors is the concentration of unbound extra-cellular polymeric substance (EPS) within the liquor which determine irreversible fouling, and suspended solids levels which determine the reversible fouling [9].

1.1. Problems with using phenomenological activated sludge biological models for design, operation and control of MBR plant

For a MBR system treating wastewater, capturing biological membrane fouling phenomena in the form of mathematical models has been a task of many different research teams around the globe for the past decade. Most researchers model the biological processes occurring in the reactor using a phenomenological Activated Sludge approach that obeys the basic laws of physics and can be deduced from first principles and scientific theories [2]. However, all Activated Sludge models do suffer from the following disadvantages:

- A purely mathematical and automated optimisation of Activated Sludge variety model can prove problematic due to the complexity and resulting unidentifiable nature of the highly non-linear processes involved, especially for the International Water Association (IWA) ASM1 which has its circular death regeneration concept [3]. Hence mathematical optimisation should always be supported by sufficient expert process knowledge, since an optimisation algorithm cannot differentiate between more defined (i.e., stoichiometric parameters) or less defined parameter values, and will often end up producing rather small modifications to a considerable number of parameters.
- Another major problem encountered in calibration of these models is the lack of identifiability of the model parameters since often more than one combination of influent characteristics and model parameters can give a reasonable fit based on the available data. Hence expert knowledge is required for these model types, so that all obtained information is carefully assessed, and so that the model parameters are constrained within realistic boundaries for the specific wastewater treatment processes under investigation.
- A proper Activated Sludge model that introduces new soluble microbial product (SMP) processes to include mechanisms for membrane fouling requires considerable expertise and process knowledge to fully develop. This expertise in model development ranges from checks to the model processes to ensure balancing of components, through to parameter estimation and performance of the model calibration and validation procedures. It also includes extensive knowledge to carry out a proper influent and sludge characterisation procedure to determine the state components for the model [10].
- The complexity of this variety of phenomenological models means they require specialist knowledge to set up and often prove difficult to use in practice especially for existing plant operation and control. In the main, they have tended to be used for research purposes or for the concept design of new plant.
- For many applications insufficient data is available to allow a full model calibration and validation, and thus the verified model is not omnipotent for every situation.
- The general application of such complex models, which in themselves require considerable calibration experience to give sufficient predictive accuracy, means their take up for process control and the development of future operational strategies will always prove limited [11].

In order to overcome the inherent deficiencies in the traditional activated sludge modelling approach, a growing group researchers are utilising non-traditional approaches to describe the biological processes for a MBR that impact on membrane fouling.

1.2. Input-output (IO) models as a possible alternative – time series system identification methods

In an ideal world, a quick and easy approach to wastewater treatment modelling is required that can be easily applied to a real life situation. This would ideally be coupled with very simple calibration procedures so that any model can be constantly "retrained" on newer plant data sets as and when they become available. Since this "retraining" would prove straight forward, it could be performed as many times as necessary. To make this proposed new approach easy to apply, it should not require an intimate knowledge of the exact processes occurring in the MBR, so it could be applied by any non-specialist who was new to wastewater treatment modelling [12].

Very few alternative approaches have been used to date when compared to the traditional mechanistic models developed for wastewater treatment plant [11]. A lesser known approach is time series modelling using autoregressive models. It is more commonly used in econometric system forecasting for international financial markets [7]. It has only been used in a limited manner for wastewater treatment modelling, and even then, only for the simple modelling of effluent leaving a plant [13]. It has been hypothesized under this study that a formulation based on simplified IO times series

models, should be developed as an alternate, simpler and faster way of calibrating and verifying MBR Activated Sludge models. This would mean that the exact nature of the biology in the bioreactor and its effects on the membrane fouling process need not be fully understood, as the time series models would be based solely on historical IO data sets that would be used to predict future plant output. This procedure if it proves effective is largely linearised around an operating point or range so that any solutions are easily obtained. It would then be very useful for plant control and operation, and be much quicker to develop than a phenomenological model since an intimate knowledge of the physics and chemistry behind the process is not required. Additionally, complex theory and mathematics to describe this theory would not be needed thus again saving time in model development [12].

Under this study two different model types, namely phenomenological model structures and IO times series model structures, were tested to ascertain which gave best results. The main research questions posed were:

i) How easy is it in practice to calibrate and validate a relatively simple phenomenological biological Activated Sludge model for a real life MBR plant which is still rich enough in complexity to include the major biological/biochemical agents involved in the fouling of MBRs?
ii) Is a system identification procedure using time series analysis a simpler, quicker modelling approach to use to accurately determine the biological interactions within the bioreactor of a real life MBR plant? Can it give the same degree of accuracy as a phenomenological model? Is it as useful? Is it as robust?

An IO models based on standard mathematical formulations such as ordinary differential equations or difference equations of various orders can be used as a quick method for model prediction as no prior process knowledge is required for model calibration and validation [7]. The procedure automatically selects the best order model based on the number of lags in output data that give the optimal prediction. Little skill is needed by the simulator to obtain best fit, and a significant amount of time is saved when compared to the complex needs of verifying a typical mechanistic model.

2. Description of models utilised

2.1. Phenomenological models used

The most widely known and used Activated Sludge model is the IWA's ASM1, which has become a major reference for many scientific and practical projects. It was first introduced in 1987 by Henze, et al. and is still considered as a "state-of-the-art" global model with its kinetic and stoichiometric parameters having being extensively studied and calibrated on an international basis [3]. This model contains seven soluble, *S*, and, six particulate, *X*, components in the wastewater. Each of the thirteen components represents an independent state and has different growth and decay processes. The IWA's ASM3 is another powerful and commonly used biological model [4]. It was formulated by the same international working group that initially created the ASM1. The basic idea of the ASM3 is to give a much clearer and easier distinction between the soluble and particulate components that practically reflect the reality faced by plant operators.

Many researchers have modified these basic Activated Sludge models to reflect the processes occurring in a MBR especially those thought to contribute to membrane fouling. Hence the start point of this research was to take two versions of existing modified Activated Sludge models that take into account the production of EPS as SMP were tested using real life data produced from a full scale sidestream configuration MBR plant.

2.1.1. ASM1 Lu (2002) model

The first version of the modified biological model used the ASM1 combined with a SMP model and is based on the work carried out by Lu, et al. [3,5]. The mechanisms involved in this combined ASM1 and SMP model are described in Fig. 1. This figure depicts the SMP formation–degradation processes based on the ASM1 where the SMP consists of two types, namely utilisation associated products (UAP) and biomass associated products (BAP).

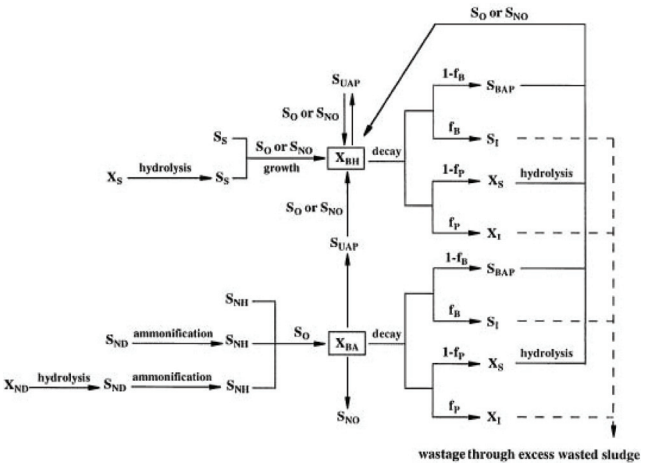

Fig. 1. Schematic of modified ASM1 Lu model with SMP formation and decay.

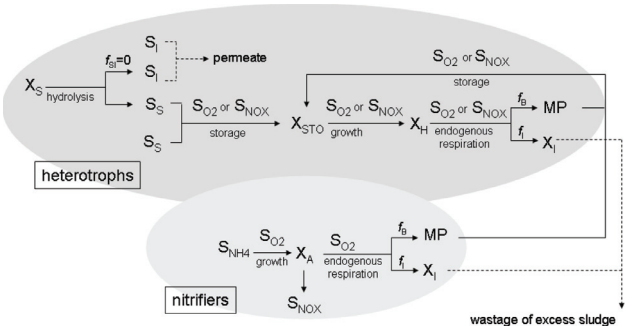

Fig. 2. Schematic of modified ASM3 Oliveira model with MP products.

2.1.2. ASM3 Oliveira (2005) model

The second version of the modified biological model used the ASM3 combined with a microbial product (MP) model based on the work carried out by Oliveira-Esquerre, et al. [4,6]. Fig. 2 outlines the combined ASM3 Oliveira-Esquerre model which takes into account the microbial product (MP) as part of the biotransformation process [6]. This MP is analogous to total SMP in other models.

2.2. IO models used – Autoregressive model structures

System identification is an iterative process in which models with different structures are identified from data, and the individual model performance compared. The normal start point is by estimating the parameters of very simple model structures. If the performance still proves poor, then the model structure is gradually increased in complexity. Ultimately the simplest of all model structures tested is eventually selected that best describes the dynamics of the system under scrutiny. In this iterative process, which can be automated, the system identification procedure commences by initially using linear continuous IO polynomial model structures, such as autoregressive exogenous (ARX) and autoregressive exogenous moving average (ARMAX) ones. Later on linear continuous IO state-space model structures are also tested using the supplied times series data [7]. The best fit structure is then chosen as the optimal model formulation.

Incidentally the ARX model is the simplest one of a group of linear prediction formulas based upon a general linear case. This model type attempts to predict an output $y[n]$ of a system based on the previous outputs ($y[n-1]$, $y[n-2]$...) and inputs ($x[n]$, $x[n-1]$, $x[n-2]$...). Deriving the linear prediction model for the estimated output, $y_e[n]$, involves determining the coefficients a_1, a_2, .. and b_0, b_1, b_2, ... in Eq. (1).

$$y_e[n] = a_1 y[n-1] + a_2 y[n-2] \ldots + b_0 x[n] + b_1 x[n-1] + b_2 x[n-2] + \ldots \quad (1)$$

An ARX model formulation is simple and has good noise-to-signal ratios, while the ARMAX is designed when the dominate disturbances enter via the input states which is the case for wastewater treatment plant. The state-space models are first order versions of the autoregressive form that utilise intermediate state vectors in the calculation procedure. The state space model structure is a good choice for quick estimation because it requires only two parameters, namely the model order and one or more input delays.

All these model formulations are solved using iterative optimisation techniques and algorithms like the least squares method. However, this requires a lot of computing power and they are prone to inherent inaccuracies. A much more attractive model formulation is the subspace one which does not need to be solved using iterative optimisation techniques and algorithms, but by only using algebraic calculations [14]. This means the subspace model formulation is a very powerful version of the state-space one that uses only a single-shot solving procedure with improved accuracy.

3. Model calibration and validation

3.1. Experimental procedure – full scale MBR plant

Both biological model types have been tested on data obtained from an full scale MBR plant designed and installed by Aquabio Ltd., and located in Worcestershire that treats salad wash water as industrial effluent. This plant has been operational since 2001 and is owned by Kanes Foods Ltd., a UK vegetable processor. This plant enables the reuse of up to 75% of the wastewater and is based on pre-treatment and aerated flow balancing followed by a crossflow sidestream MBR, reverse osmosis and ultraviolet disinfection (see Figs. 3 and Fig. 4).

In order to effectively separate the biomass, 4 banks of specialized ultrafiltration modules are used, fed by a recirculation system from the two bioreactor tanks. To allow a fully calibrated model to be produced measurements were taken at the following points in the flow train: i) at the inflow; ii) in the bioreactors; iii) in the permeate flow; and, iv) at the wastage point. Table 1 summarizes some of the typical biological and nutrient loading data that was collected from the plant during an intensive three weekly sampling programme. The figures are given as either a range or as an average with plus or minus the largest variance. The measured data was used to carry out a full wastewater characterisation and then simulations were run on Matlab© [10].

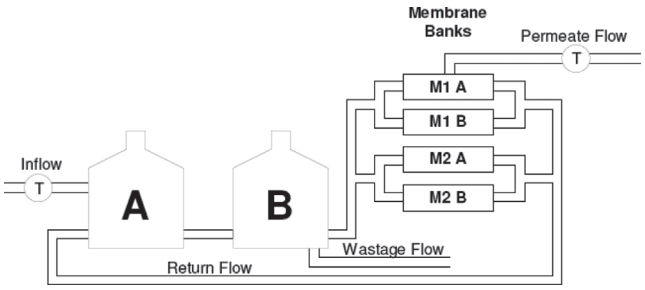

Fig. 3. Flow train of Kanes Foods Ltd. full scale MBR plant.

Fig. 4. Picture of Kanes Foods Ltd. full scale MBR plant.

3.2. Model simulation – results for phenomenological model formulations

One aspect of both these models that was discovered was that neither model was fully tested as should have been done by their developers, so neither model as they original stand are accurate in predicting sludge yields, aeration demands, etc. Hence before being used, both these models were extensively tested, and then modified to make them accurate as possible within the constraints of each model structure. This meant changing the stoichiometry and kinetic parameter values as necessary to reflect a typical MBR scenario.

Several simulations were run using the fractionated influent and bioreactor component state data on both the original ASM1 and ASM3 models followed by runs using the modified ASM1 Lu model and the modified ASM3 Oliveira model. All outputs from each simulation run were plotted on the same graphs so results could be directly compared. Figs. 5 and 6 are plots of measured values against simulated values of COD and Ammonia concentration in Reactor B for the four model runs. It is clearly evident from the results of all four model type simulation runs, that the original untouched models overall performance is nearly always better than the modified versions that include SMP components. This is as would be expected due to the increased parameter sets that in themselves have not been properly verified by extensive global research

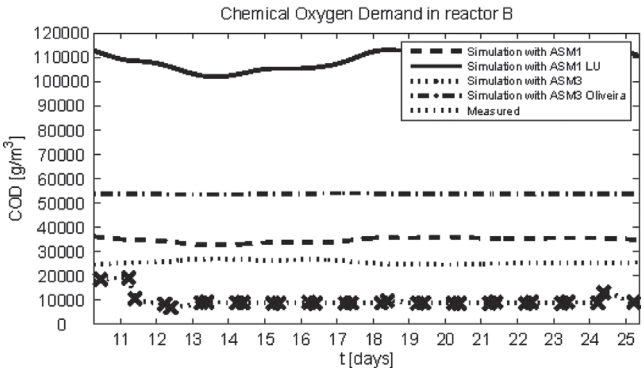

Fig. 5. Measured COD in Reactor B versus simulated values from ASM1, ASM1 Lu, ASM3, and ASM3 Oliveira models.

Table 1
Some of the typical averaged biological, nutrient and other measurements made during 3 week sampling period

Measured data	Inflow	Bioreactor	Permeate	Wastage
COD unfiltered (mg/l)	455 ± 179	7032 – 19240	8 – 30	1012 – 19600
TSS / MLSS (mg/l)	72 – 692	15250 ± 2000	4 – 68	820 – 22890
Total Nitrogen (mg/l)	22 ± 17	53 ± 19	0.7 – 16.9	12 – 204
Ammonia (mg/l)	0.2 – 4.8	5.2 ± 3.1	0 – 0.6	3.2 – 13.0
Total Phosphorous (mg/l)	2.4 – 8.4	3.9 – 22.4	0.2 – 4.4	4.0 – 25.3
Return Flow rate (m^3/hr)	254 ± 6	–	–	–
Viscosity (cP)	–	170 ± 63	–	177 ± 302

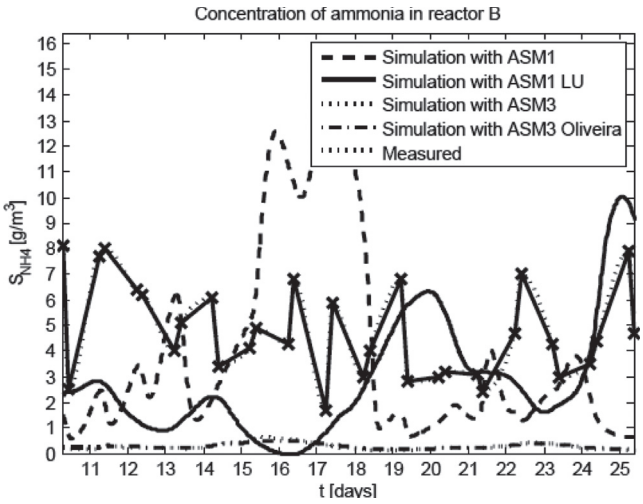

Fig. 6. Measured Ammonia in Reactor B versus simulated values from ASM1, ASM1 Lu, ASM3, and ASM3 Oliveira models.

as is the case with those for the original untouched Activated Sludge model varieties.

3.3. Model simulation – results for IO model formulations

After various assumptions and simplifications of the plant data, a multi-input multi-output (MIMO) model structure was tested. As shown in Eq. (2), the flow into the membrane module and the influent component states were used as variables in the input model vector, x, with a limited number of reactor component states being the variables in the output model vector, y. In this case only a maximum of ten input states was utilised while the output states were limited to two in order to improve the fit, since there was only a limited data set available. The MLSS and viscosity were specifically chosen as output states as they directly and indirectly contribute to membrane fouling processes [1].

$$x = \begin{bmatrix} q_{in} & COD_{in} & COD_{fil,in} & TSS_{in} & TN_{in} \\ NH3_{in} & NO3_{in} & TP_{in} & PO3_{in} & pH_{in} \end{bmatrix}^T$$
$$y = \begin{bmatrix} MLSS & \mu \end{bmatrix}^T \quad (2)$$

Fig. 7 shows the best fit results obtained for simulation of this MIMO model structure when it is run as a subspace formulation. As can be seen the fit for MLSS is reasonable although the viscosity fit is poor.

Several other simulations were run using different autoregressive model formulations, and also a reducing set of input states. The final runs only use a single input multi-output (SIMO) model structure. Table 2 details the results of these simulation runs. As can be seen a reduced input data set gives better model fits as there is less correlation needed between component states, so the final validated model structure is greatly simplified. The subspace formulation performs best with the ARX formulation proving next best in accuracy. The other two standard autoregressive formulations either perform very badly or fail to run.

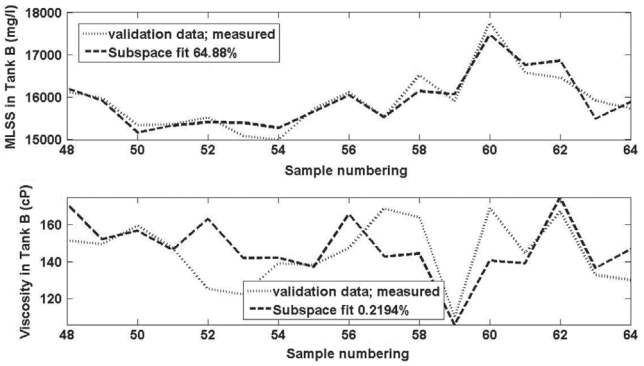

Fig. 7. MIMO Subspace formulation – Best fit using all biological variables as inputs with MLSS and viscosity levels in bioreactor as outputs.

Table 2
Comparison of best fits for various autoregressive model formulations

Input data set type	Subspace	ARX	ARMAX	State-space
All ten input data components	30.29% MLSS –16.50% μ	59.26% MLSS 55.48% μ	did not run	too few data
COD and DO as inputs only	48.16% MLSS –49.74% μ	9.055% MLSS 13.13% μ	did not run	5.127% MLSS 5.672% μ
Only COD as input only	64.88% MLSS 0.2194% μ	10.60% MLSS 18.01% μ	did not run	no fit

Table 3
Summary of results of both model types

	Phenomenological models				MIMO models			
Results from data	Original ASM3	Modified ASM3 Oliveira	Original ASM1	Modified ASM1 Lu	Subspace	ARX	ARMAX	State-space
Full scale MBR plant	Good fit	Poor fit	Good fit	Very poor fit	Reasonable fit	Reasonable fit	–	–

4. Conclusion

Table 3 summarises qualitatively the simulation results for both the phenomenological models and the MIMO models. The following summarised points can be made regarding behavioural IO Activated Sludge variety model structures:

- These model structures are intuitively easily determined structures since they mainly represent the various state variable components in the model.
- They can be easily set up and run without expert knowledge and the input and output component states can be very easily altered to allow numerous simulations scenarios to be tested.
- A major limitation of these models are that they require a considerable amount of data taken over a long time period to run adequately which increases greatly when using large MIMO structures. This can be overcome to some extent by using SIMO structures instead.
- This study proves that a subspace procedure and a standard ARX method can give comparable accuracy when directly compared to the two main unaltered IWA Activated Sludge models [2].

It initially looks like this novel approach has many advantages over traditional mechanistic models while giving comparable results for some IO structures. Early simulation results described in this study prove this especially for subspace methods. However these methods can prove very fragile particularly the ARMAX formulation which is prone to crashing. They also require very large data sets to produce accurate formulations, and these linear models are only useful around a very narrow operating range or operating point.

When answering the research questions posed, then generally speaking the overall performance of both the phenomenological models and the IO models proved similar. All the phenomenological models proved difficult to set up and run for reasons already discussed earlier, while all the IO model forms proved the opposite. The optimal way ahead in this area of research may be the prudent use of a combination of both model types. Hence this means using a conventional phenomenological Activated Sludge model to predict most process states while using a reduced IO model structure to predict the process states that impinge directly on membrane fouling (i.e., SMP levels). This has the advantage of using well defined existing standard IWA models with all their benefits, with the SMP state components/s only being used in the IO model version which would be easy to calibrate. In conclusion, further research is required using longer historical data sets to definitively ascertain whether this autoregressive modelling approach can be further developed and improved upon.

Acknowledgements

The author would like to thank and acknowledge the assistance received especially in terms of data collation from the following organisations: Process Control Water Software Systems research group at De Montfort University; Aquabio Limited; Kanes Foods Limited; ITT Sanitaire (UK); and Northern Ireland Water Services.

Symbols

μ	—	Viscosity (cP)
COD	—	Chemical oxygen demand (mg/l)
COD_{fil}	—	Filtered chemical oxygen demand (mg/l)
TSS / $MLSS$	—	Total suspended solids/mixed liquor suspended solids (mg/l)
TN	—	Total nitrogen (mg/l)
TP	—	Total phosphorous
PO_3^-	—	Phosphate concentration (mg/l)
S_s / X_s	—	Readily/slowly biodegradable substrate (mg/l)
S_I / X_I	—	Soluble/particulate inert concentration (mg/l)
X_{BH} (X_H)	—	Heterotrophic biomass concentration (mg/l)
X_{BA} (X_A)	—	Autotrophic biomass concentration (mg/l)

S_{UAP}	—	Utilisation associated product concentration (mg/l)
S_{BAP}	—	Biomass associated product concentration (mg/l)
MP	—	Concentration of microbial products (mg/l)
S_O (S_{O2} or DO)	—	Dissolved oxygen concentration (mg/l)
S_{NO} (S_{NOX} or NO_3^-)	—	Nitrate and nitrite nitrogen concentration (mg/l)
S_{NH} (S_{NH4} or NH_3^+)	—	Ammonium concentration (mg/l)
S_{ND}/X_{ND}	—	Soluble/particulate organic nitrogen concentration (mg/l)
X_{STO}	—	Cell internal storage product of heterotrophic organisms (mg/l)
F_B (F_l)	—	Inert fraction of biomass leading to soluble products
F_P	—	Fraction of biomass yielding particulate products
F_{SI}	—	Fraction of X_S that hydrolyses to soluble inert products

References

[1] A. Drews, J. Mante, V. Iversen, M. Vocks, B. Lesjean, and M. Kraume, Does fouling in MBRs depend on SMP? Desalination, 231(2008) 141–149.

[2] IWA, Task Group on mathematical modelling for design and operation of biological wastewater treatment, Activated Sludge Models ASM1, ASM2, ASM2d & ASM3, IWA Publishing, London, 2000.

[3] M. Henze, C.P.L. Jr. Grady, W. Gujer, G.V.R. Marais and T. Matsuo, Activated Sludge Model No. 1, IAWPRC Scientific Technical Report No. 1, London, 1987.

[4] W. Gujer, M. Henze, T. Mino and M. van Loosdrecht, Activated Sludge Model No. 3, Water Sci. Technol., 39(1) (1999) 183–193.

[5] S.G. Lu, T. Imai, M. Ukita, M. Sekine, T. Higuchi and M. Fukagawa, A model for membrane bioreactor process based on the concept of formation and degradation of soluble microbial products, Water Res., 35(8) (2002) 2038–2048.

[6] K.P. Oliveira-Esquerre, H. Narita, N. Yamato, N. Funamizu and Y. Watanabe, Modelling of conventional MBR for wastewater treatment, 21st Center of Excellence Program Publication, Department of Urban and Environmental Engineering, Hokkaido University, 2005.

[7] L. Ljung, System Identification: Theory for the User, Prentice Hall, 1999.

[8] G. Olsson and B. Newell, Wastewater treatment systems: modelling, diagnosis and control, IWA Publishing. 1999.

[9] S. Judd, The MBR Book: Principles and Applications of Membrane Bioreactors for Water and Wastewater Treatment, Elsevier Science, 2006.

[10] J.J.W. Hulsbeek, J. Kruit, P.J. Roeleveld and M.C.M. van Loosdrecht, A practical protocol for dynamic modelling of activated sludge systems, Water Sci. Technol., 45(6) (2002) 127–136.

[11] K.V. Gernaey, M.C.M. van Loosdrecht, M. Henze, M. Lind and S.B. Jorgensen, Activated sludge wastewater treatment plant modelling and simulation: state of the art, Environ. Model. Software, 19 (2004) 763–783.

[12] P. Paul, Using traditional modelling approaches for a MBR system to investigate alternate approaches based on system identification procedures for improved design and control of a wastewater treatment process, PhD Thesis, Process Control – Water Software Systems research group, Dept. of Engineering, Faculty of Technology, De Montfort University, 2011.

[13] S.A. Dellana and D. West, Predictive modelling for wastewater applications: Linear and nonlinear approaches, Environ. Model. Software, 24 (2009) 96–106.

[14] H. Chen and J.M. Maciejowski, A new subspace identification method for bilinear systems, Technical Report CUED/F-INFENG/TR.357, University of Cambridge, UK, 1999.