Pergamon

Wat. Sci. Tech. Vol. 36, No. 5, pp. 81–88, 1997.

Printed in Great Britain.
0273–1223/97 $17·00 + 0·00

PII: S0273-1223(97)00497-6

MODELING ERROR IDENTIFICATION OF ACTIVATED SLUDGE MODELS

Zhigou Yuan, Peter A. Vanrolleghem and Ghislain C. Vansteenkiste

Department of Applied Mathematics, Biometrics and Process Control (BIOMATH), University of Ghent, Coupure 653, B-9000 Ghent, Belgium

ABSTRACT

Information about the location of modeling errors is crucial for the efficient improvement of an invalid model. This article discusses how to pinpoint modeling errors through comparison of experimental data with data obtained through simulation of the invalid model. An observer-based approach is presented. By designing a dedicated observer for the system using the invalid model, a signal vector is generated, on which each modeling error imposes an easily identifiable feature. An algorithm to analyze the featured signal is then presented. With this algorithm, the features of each of the modeling errors are extracted. The approach is illustrated for a denitrification reactor model in which errors in the dimension of the state vector, in the structure of the biokinetic relationship and in the values of the parameters could be identified. © 1997 IAWQ. Published by Elsevier Science Ltd

KEYWORDS

Biological denitrification, modeling error identification, model validation, observer

INTRODUCTION

The use of modeling and simulation has proven to be invaluable in the design, analysis and optimization of activated sludge wastewater treatment plants (see for instance Coen *et al.*, 1996). A typical modeling process starts by identifying an Experimental Frame. The frame represents the experimental conditions under which the modeller wants to investigate the system. As such, it reflects the modeller's goals and questions. Based on the identified frame, a class of matching models can be identified. Through structure characterization, the appropriate model structure is selected based on the *a priori* knowledge and/or the measurement data. Subsequently, parameter estimation and/or calibration yield parameter values. Using the identified model and parameters, simulation allows one to mimic the system behavior. The validity of the model is then substantiated by comparing new experimental data sets (different from those used for model structure and parameter identification) to those produced by simulation (model validation).

A model has to be corrected once proven invalid. Considering the important efforts required during model building, *e.g.* for the collection of the experimental data, it must be quite obvious to the reader that an indication of the source of model error is very valuable for more economic model improvement as it allows to direct the efforts.

As one might intuitively expect, different modeling errors usually cause the behavior of the model to deviate in different ways from that of the real system. Or, in other words, different modeling errors correspond to different "patterns" in the error signal, the difference between experimental data and

simulated data. These "patterns", if extractable, can obviously be used to identify the modeling errors. Starting from this conjecture, an observer-based approach to modeling error identification has been developed in Yuan *et al.* (1996b). By designing a dedicated observer of the system with the invalid model, a signal vector was generated, on which each modeling error imposed an identifiable feature. An algorithm to analyze the featured signal was also developed, with which the features of the modeling errors were extracted. As a result of this specific approach, the type as well as the location of the errors can be determined once a model is found to contain significant errors. Then, an appropriate action to improve the model can be taken.

The aim of the current paper is to show how the errors associated with activated sludge models can be identified with such an approach. This is achieved by showing the identification of the modeling errors of a biological denitrification model. The theory presented in Yuan *et al.* (1996b) is briefly summarized first. The denitrification model and its enumerable modeling errors are then discussed. Finally simulation studies on identifying the modeling errors of the denitrification model are reported.

AN OBSERVER BASED APPROACH TO MODELING ERROR IDENTIFICATION

Model and System Representation

Assume that the parametrized model under evaluation takes the form,

$$\dot{x}_m(t) = f_m(x_m(t), \theta_m, u(t), t) \tag{1}$$

where $x_m(t) \in R^n$ is the state variable vector of the model, $u(t) \in R^p$ is the input vector, and θ_m is the model parameter vector, which is known (*e. g.* from parameter estimation). On the basis of this model, the real behavior of the system can generally be represented as (Yuan *et al.*, 1996a),

$$\dot{x}_r(t) = f_r(x_r(t), \theta_r, u(t), t) = f_m(x_r(t), \theta_m, u(t), t) + e_m(t) \tag{2}$$

where $x_r(t) \in R^n$ is the state vector of the system, $e_m(t) \in R^n$ is the modeling error vector. It is assumed in equation (2) that the real system has the same number of state variables as the model. This representation does not limit the generality of the representation since the errors introduced by erroneous state aggregations in deriving model (1) can also be represented by the error term $e_m(t)$. The error term $e_m(t)$ can further be represented as $e_m(t) = \sum_0^l F_i d_i(t)$, and hence system equation (2) is rewritten as (Yuan *et al.*, 1996b),

$$\dot{x}_r(t) = f_m(x_r(t), \theta_m, u(t), t) + \sum_{i=0}^{l} F_i d_i(t) \tag{3}$$

where $F_i \in R^{n \times s_i}$, a constant matrix, is called the feature matrix of the i-th modeling error; $d_i(t) \in R^{s_i}$, which is generally unknown and time-varying, represents the magnitude of the i-th modeling error. To apply the modeling error identification approach to be presented below, one is required to construct the feature matrices based on *a priori* analysis of the possible modeling errors. However, it is not necessary to specify $d_i(t)$.

Feature Equation Derivation

To present the concept, only the case where all the states $x_r(t)$ of system (3) are measurable is discussed here. The approach for the case of incomplete state measurement can be found in Yuan *et al.* (1996b). Based on model (1), a dedicated observer of system (3) is designed as follows,

$$\dot{x}_O(t) = f_m(x_r(t), \theta_m, u(t), t) + g \cdot (x_r(t) - x_O(t)) \tag{4}$$

which is asymptotically stable as long as g, a scalar, is positive. Note that observer (4) differs from a conventional one in that: (1) $f_m(x_r(t), \theta_m, u(t), t)$, instead of $f_m(x_O(t), \theta_m, u(t), t)$, is used on the

right-hand side of the observer equation; (2) the observer gain matrix is designed as $g \cdot I$. As will be seen in the sequel, these special designs make the modeling error identification possible.

The dynamics of the error signal $e(t) = x_r(t) - x_O(t)$, is characterized by $\dot{e}(t) = -ge(t) + \sum_{i=0}^{l} F_i d_i(t)$. Thus,

$$\sum_{i=0}^{l} F_i \bar{d}_i(t) = e(t) \tag{5}$$

where $\bar{d}_i(t) = \int_{t_0}^{t} e^{-g(t-\tau)} d_i(\tau) d\tau$ is the manifestation of the i-th modeling error in the error signal $e(t)$. In equation (5) $e(t_0) = 0$ is assumed since $x_r(t)$ are completely measurable.

Equation (5) clearly shows that each modeling error imposes a subspace, which is spanned by its feature matrix, on the n-dimensional $e(t)$ space. This indicates that one may identify the modeling errors associated with the model by investigating the properties of the $e(t)$ space. Equation (5) is called a feature equation. Correspondingly, $e(t)$ is called the featured error signal.

Modeling Error Identification

Equation (5) implies that information about modeling errors can be obtained when the featured signal $e(t)$ is localized to a certain subspace. Obviously, the identification process involves two essential steps, localizing $e(t)$ to a subspace and interpreting the localization based on the feature equation.

Given a signal $e(t) \in R^n$ and a subspace $\mathcal{S} \subseteq R^n$, to verify if $e(t)$ lies in $\mathcal{S}$ one may investigate the projection of $e(t)$ on $\mathcal{S}^\perp$, the orthogonal complementary subspace of $\mathcal{S}$, making use of the fact that any signal lying in $\mathcal{S}$ has zero projection on $\mathcal{S}^\perp$. Assume $V = [V_1\ V_2] \in R^{n \times n}$, where the columns of V_1 and V_2 constitute a unitary, orthogonal basis of the subspaces $\mathcal{S}$ and $\mathcal{S}^\perp$, respectively. The projection of signal vector $e(t)$ on $\mathcal{S}^\perp$, represented under the basis constituted by the columns of V_2, is $V_2{}^T e(t)$. If $e(t) \in \mathcal{S}$, we have $V_2{}^T e(t) = 0$. This equation gives a theoretical criterion to judge whether or not $e(t)$ lies in $\mathcal{S}$. Practically, however, it will never strictly hold due to, for instance, noise contained in the measurement of $x_r(t)$, the existence of insignificant modeling errors, *etc.*. Noting that the objective is to identify the most significant modeling errors, this criteria is relaxed to,

$$\|V_2^T e(t)\|_2 / \|V_1^T e(t)\|_2 \leq \epsilon_1 \tag{6}$$

where $V_1^T e(t)$ is the projection of $e(t)$ on $\mathcal{S}$, represented under the basis constituted by the columns of V_1; ϵ_1 is a threshold, which should be chosen according to the magnitude of the system noise and of the measurement noise.

Let $\mathbf{M_{ex}}$ be a subset of the pre-defined modeling error set $\mathbf{M_p} = \{F_1, F_2, \cdots, F_l\}$. Note that symbol F_i, defined as the feature matrix of the i-th modeling error before, is also used to denote the i-th modeling error for simple notation. Assume $\mathbf{M_{ex}}$ satisfies, $\sum_{F_i \in \mathbf{M_{ex}}} \Re(F_i) = \mathcal{S}$ where $\Re(F_i)$ denotes the space spanned by the columns of F_i. By introducing feature equation (5) in inequality (6), one obtains,

$$\| \sum_{F_j \in \mathbf{M_p} - \mathbf{M_{ex}}} V_2^T F_j \bar{d}_j(t) \|_2 / \| \sum_{F_i \in \mathbf{M_{ex}}} V_1^T F_i \bar{d}_i(t) + \sum_{F_j \in \mathbf{M_p} - \mathbf{M_{ex}}} V_1^T F_j \bar{d}_j(t) \|_2 \leq \epsilon_1 \tag{7}$$

When inequality (6) holds, inequality (7) implies that the most significant modeling errors are among those that are contained in $\mathbf{M_{ex}}$ or those whose subspaces spanned by their feature matrices are 'close' to $\mathcal{S}$. The latters constitute a set $\mathbf{M_{cl}}$, which can be constructed with,

$$\mathbf{M_{cl}} = \{F_j \mid F_j \in \mathbf{M_p} - \mathbf{M_{ex}},\ \sigma(V_2^T F_j)/\sigma(V_1^T F_j) \leq \epsilon_2 \text{ where } \epsilon_2 > \epsilon_1\} \tag{8}$$

where $\sigma(V_2^T F_j)$ and $\sigma(V_1^T F_j)$ are the largest singular values of matrices $V_2^T F_j$ and $V_1^T F_j$, respectively. Generally, ϵ_2 should be chosen sufficiently large in order to reduce the risk of excluding significant modeling errors contained in $\mathbf{M_p} - \mathbf{M_{ex}}$. However, a too high ϵ_2 decreases the effectiveness of the algorithm. A user specified compromise is thus needed. When inequality (6) does not hold, which implies that $e(t)$ does not lie in subspace $\mathcal{S}$, no information about which modeling errors are the most significant ones can be drawn. The above discussion leads to the following algorithm for identifying the

modeling errors.

Step 1 : Construct set $\mathcal{S}_\mathbf{c}$ as,

$$\mathcal{S}_\mathbf{c} = \{\mathcal{S}_i \mid \mathcal{S}_i = \sum_{F_{i,j} \in M_i} \Re(F_{i,j}) \quad \text{where} \quad \mathbf{M_i} = \{F_{i,1}, F_{i,2}, \cdots, F_{i,k_i}\} \subseteq \mathbf{M_p}\,,\ dim(\mathcal{S}_i) < n\} \tag{9}$$

Apparently, $\mathcal{S}_\mathbf{c}$ is a set constituted by all the true subspaces of R^n that are spanned by the feature matrices of one or more pre-defined modeling errors.

Step 2 : For each $\mathcal{S}_i \in \mathcal{S}_\mathbf{c}$, construct matrices V_1 and V_2 such that their columns constitute unitary bases of $\mathcal{S}_i$ and $\mathcal{S}_i^\perp$, respectively. Define a modeling error set $\mathbf{M}_{\mathcal{S}_i}$ as,

$$\mathbf{M}_{\mathcal{S}_i} = \{F_j \mid F_j \in \mathbf{M_p}\,,\ V_2 F_j = 0 \ \text{ or } \ \sigma(V_2^T F_j)/\sigma(V_1^T F_j) \leq \epsilon_2\} \tag{10}$$

According to the above analysis, $\mathbf{M}_{\mathcal{S}_i}$ is such a set that all the modeling errors in $\mathbf{M_p} - \mathbf{M}_{\mathcal{S}_i}$ can be excluded when inequality (6) holds.

Step 3 : For each $\mathcal{S}_i \in \mathcal{S}_\mathbf{c}$, define f_i as a logical vector. $f_i = 0$ if inequality (6) holds; $f_i = 1$ otherwise. The most significant modeling errors are thus localized to,

$$\mathbf{M_e} = \bigcap_i \{\mathbf{M}_{\mathcal{S}_i} \cup (\mathbf{M_p} - \mathbf{M}_{\mathcal{S}_i}) f_i\} \tag{11}$$

MODELING A BIOLOGICAL PROCESS

Process Model

Figure 1 shows a biological denitrification process of a post-denitrification plant, which aims to remove the nitrate as well as the carbon compounds contained in the influent water by means of biological reactions. It consists of two functional units, a bioreactor and a settler. In the reactor, which is often completely mixed, biomass is present in the form of sludge flocs. The biomass oxidizes the carbon compounds with nitrate as the oxygen source. The carbon compounds and nitrate are thus both removed and biomass is produced. In order to prevent the sludge concentration in the reactor from becoming too high due to its continuous growth, surplus sludge is removed via the waste flow (see Figure 1). The *a priori* knowledge allows one to model the process by making mass balances for three materials,

$$\begin{aligned}
\dot{X}(t) &= \mu(t)X(t) - bX(t) - \frac{Q_w(t)}{V}X(t) \\
\dot{S}_S(t) &= -\frac{1}{Y_H}\mu(t)X(t) - \frac{Q_{in}(t)}{V}S_S(t) + \frac{Q_{in}(t)}{V}S_{S,in}(t) \\
\dot{S}_{NO}(t) &= -\frac{1-Y_H}{2.86Y_H}\mu(t)X(t) - \frac{1-f_P}{2.86}bX(t) - \frac{Q_{in}(t)}{V}S_{NO}(t) + \frac{Q_{in}(t)}{V}S_{NO,in}(t)
\end{aligned} \tag{12}$$

where X, S_S, S_{NO} denote the biomass, the carbon compounds and the nitrate concentrations in the bioreactor, respectively; $S_{S,in}$ and $S_{NO,in}$ denote the carbon compounds and the nitrate concentrations in the influent, respectively; Q_{in} is the influent flow rate; Q_w is the waste flow rate; V is the volume of the bioreactor; Y_H is the yield coefficient; b is the biomass decay coefficient; f_P is the fraction of the inert materials in biomass; $\mu(t)$ is the specific biomass growth rate, which is still to be modeled. Note that a point settler is assumed in the model.

The following double Monod law is commonly used to model the specific biomass growth rate in the considered system (Henze *et al.* 1987),

$$\mu(t) = \mu_{max}\frac{S_S(t)}{K_S + S_S(t)}\frac{S_{NO}(t)}{K_{NO} + S_{NO}(t)} \tag{13}$$

where μ_{max} is the maximum specific growth rate, K_S and K_{NO} are the so-called half saturation coefficients for the carbon compounds and nitrate, respectively. Equation (12), together with equation (13), gives a parametric model of the denitrification process. All the parameters involved are plant dependent and hence have to be specifically calibrated or estimated for each individual case.

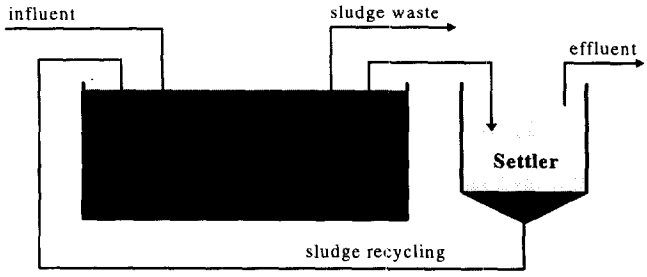

Figure 1: Biological denitrification process

Modeling the Enumerable Modeling Errors

Basically, modeling errors may be introduced in each stage of the modeling process. In defining the experimental frame, some important components may be missed, some significant disturbances to the system may be improperly neglected, state variables may be deliberately aggregated to reduce the model order and so on. All of these introduce errors into the model. In identifying a model structure, a wrong one may be assumed due to for instance lack of knowledge of the mechanism of the process or due to an oversimplification. Similarly, incorrect model parameter values may be used due to improper or inadequate data used for model estimation and/or ill defined estimation algorithms. Analysis of the denitrification process allows for instance to enumerate the following modeling errors,

modeling error due to oxygen presence in influent
An assumption underlying model (12) is that no other reactions occur in the process which affect the mass balance of the concerned materials. One knows, however, that this assumption is not valid when dissolved oxygen is present in the influent. In fact, when dissolved oxygen is fed to the bioreactor aerobic oxidation will also occur. With r_o denotes the oxidation reaction rate, this reaction introduces an r_o term into the first equation of (12) and an r_o/Y_H term into the second equation. The modeling error term in equation (2) thus takes the form of $e_{m,o}(t) = [1 \ -1/Y_H \ 0]^T r_o(t)$, where $[1 \ -1/Y_H \ 0]^T$ is the feature matrix/vector of the concerned modeling error, $r_o(t)$ is the unknown, time-variant magnitude of the modeling error.

modeling error due to an improperly characterized reaction rate
There does not exist a fundamental law that precisely characterizes the dependence of the specific denitrification reaction rate on the concentrations of the substrates. The "laws" which have hitherto been proposed are all quite empirical. A problem of this type of laws is that they have a limited applicability range. An inappropriate choice of the "laws" will introduce the following error term (assuming that the real reaction rate: $\mu_r(t) = \mu(t) + \delta\mu_s(t)$, where $\delta\mu_s(t)$ is the modeling error) into equation (2) : $e_{m,\mu}(t) = [1 \ -1/Y_H \ -(1-Y_H)/(2.86Y_H)]^T \delta\mu_s(t)X(t)$.

modeling errors due to inaccurate parameter values
modeling error of b. Assuming $b_r = b + \delta b$, where b_r is the real decay coefficient and δb is the modeling error, one obtains $e_{m,b}(t) = [-1 \ 0 \ -(1-f_P)/2.86]^T \delta b X(t)$.
modeling error of f_P. Assuming $f_{P,r} = f_P + \delta f_P$, where $f_{P,r}$ is the real inert fraction in a biomass cell and δf_P is the modeling error, one obtains $e_{m,f_P}(t) = [0 \ 0 \ 1]^T \delta f_P b X(t)/2.86$.
modeling error of Y_H. Assuming $1/Y_{H,r} = 1/Y_H + \delta(1/Y_H)$, where $Y_{H,r}$ is the real yield coefficient and $\delta(1/Y_H)$ is the modeling error, one obtains $e_{m,Y_H}(t) = [0 \ -1 \ -1/2.86]^T \delta(1/Y_H)\mu(t)X(t)$
modeling errors of μ_{max}, K_{NO} **and** K_S. Assuming $\mu_r(t) = \mu(t) + \delta\mu_p(t)$, where $\mu_r(t)$ is the real specific reaction rate and $\delta\mu_p(t)$ is the error caused by the modeling error of μ_{max}, K_{NO} or K_S, one obtains $e_{m,\mu_{max},K_{NO},K_S}(t) = [1 \ -1/Y_H \ -(1-Y_H)/(2.86Y_H)]^T \delta\mu_p(t)X(t)$.

One finds that all the modeling errors shown above, caused either by missing state variables, by choosing

inappropriate model structures or using inaccurate parameter values, has been modeled with a term $F_i d_i(t)$.

SIMULATION STUDY

Design

Assume the real behavior ($\dot{x}_r(t) = f_r(x_r(t), \theta_r, u(t), t)$) of the denitrification process shown in Figure 1 is characterized by,

$$\begin{aligned}
\dot{X}_r(t) &= \mu_r(t)X_r(t) + \mu_o(t)X_r(t) - b_r X_r(t) - \frac{Q_w(t)}{V}X_r(t) \\
\dot{S}_{S,r}(t) &= -\frac{1}{Y_{H,r}}\mu_r(t)X_r(t) - \frac{1}{Y_{H,r}}\mu_o(t)X_r(t) - \frac{Q_{in}(t)}{V}S_{S,r}(t) + \frac{Q_{in}(t)}{V}S_{S,in}(t) \\
\dot{S}_{NO,r}(t) &= -\frac{1-Y_{H,r}}{2.86Y_{H,r}}\mu_r(t)X_r(t) - \frac{1-f_{P,r}}{2.86}b_r X_r(t) - \frac{Q_{in}(t)}{V}S_{NO,r}(t) + \frac{Q_{in}(t)}{V}S_{NO,in}(t) \\
\dot{S}_O(t) &= -\frac{1-Y_{H,r}}{Y_{H,r}}\mu_o(t)X_r(t) - \frac{Q_{in}(t)}{V}S_O(t) + \frac{Q_{in}(t)}{V}S_{O,in}(t)
\end{aligned} \tag{14}$$

where S_O and $S_{O,in}$ are the oxygen concentrations in the reactor and in the influent, respectively; $\mu_r(t)$ and $\mu_o(t)$ are the specific denitrification and oxidation rates, respectively,

$$\mu_r(t) = \mu_{max,r}\frac{S_{S,r}(t)}{K_{S,r}+S_{S,r}(t)}\frac{S_{NO,r}(t)}{K_{NO,r}+S_{NO,r}(t)}\frac{K_{OI}}{K_{OI}+S_O(t)} \tag{15}$$

$$\mu_o(t) = \mu_{max,r}\frac{S_{S,r}(t)}{K_{S,r}+S_{S,r}(t)}\frac{S_O(t)}{K_O+S_O(t)} \tag{16}$$

Note how $\mu_r(t)$ has a different structure from its model shown in equation (13). The extra term $K_{OI}/(K_{OI}+S_O(t))$ in $\mu_r(t)$ reflects the effect of oxygen inhibition of the denitrification rate.

While states $S_{S,r}(t)$ and $S_{NO,r}(t)$ and inputs $S_{S,in}(t)$ and $S_{NO,in}(t)$ in system (14) can be measured through chemical analysis, state $X_r(t)$, the biomass concentration, is not directly measurable. However, with a newly developed biological sensor, DECADOS (patent pending), the combination of parameters and the state $\mu_{max,r}X_r(t)/Y_{H,r}$, denoted as $\bar{X}_r(t)$, can be measured (Bogaert *et al.* 1996). In order to make this measurement usable, model (12) is rewritten as,

$$\begin{aligned}
\dot{\bar{X}}(t) &= \mu_{max}(t)\bar{\mu}(t)\bar{X}(t) - b\bar{X}(t) - \frac{Q_w(t)}{V}\bar{X}(t) \\
\dot{S}_S(t) &= -\bar{\mu}(t)\bar{X}(t) - \frac{Q_{in}(t)}{V}S_S(t) + \frac{Q_{in}(t)}{V}S_{S,in}(t) \\
\dot{S}_{NO}(t) &= -\frac{1-Y_H}{2.86}\bar{\mu}(t)\bar{X}(t) - \frac{Y_H}{\mu_{max}}\frac{1-f_P}{2.86}b\bar{X}(t) - \frac{Q_{in}(t)}{V}S_{NO}(t) + \frac{Q_{in}(t)}{V}S_{NO,in}(t) \\
\bar{X}(t) &= \frac{\mu_{max}}{Y_H}X(t) \quad \bar{\mu}(t) = \frac{S_S(t)}{K_S+S_S(t)}\frac{S_{NO}(t)}{K_{NO}+S_{NO}(t)}
\end{aligned} \tag{17}$$

Based on model (17), an observer is designed (equation (4)) with $g = 3$. The error signal,

$$e(t) = [\bar{X}_r(t)\ \ S_{S,r}(t)\ \ S_{NO,r}]^T - [\bar{X}_O(t)\ \ S_{S,O}(t)\ \ S_{NO,O}]^T$$

is a featured signal. Corresponding to the reformulation of the model, the modeling errors modeled in the previous sub-section are reformulated as,

$$\begin{aligned}
e_{m,o}(t) &= F_1 d_1(t) = [\mu_{max}\ \ -1\ \ 0]^T d_1(t) \\
e_{m,\mu(t)}(t) &= F_2 d_2(t) = [\mu_{max}\ \ -1\ \ -(1-Y_H)/2.86]^T d_2(t) \\
e_{m,b}(t) &= F_3 d_3(t) = [-1\ \ 0\ \ -Y_H(1-f_P)/(2.86\mu_{max})]^T d_3(t) \\
e_{m,f_P}(t) &= F_4 d_4(t) = [0\ \ 0\ \ 1]^T d_4(t) \\
e_{m,Y_H}(t) &= F_5 d_5(t) = [0\ 0\ 1]^T d_5(t) \\
e_{m,K_{NO},K_S}(t) &= F_6 d_6(t) = [\mu_{max}\ \ -1\ \ -(1-Y_H)/2.86]^T d_6(t) \\
e_{\mu_{max}}(t) &= F_7 d_7(t) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -0.125 \end{bmatrix}^T \begin{bmatrix} d_{7,1}(t) \\ d_{7,2}(t) \end{bmatrix}
\end{aligned}$$

With the "estimated" model parameters ($\mu_{max} = 1day^{-1}$, $b = 0.24day^{-1}$ $f_P = 0.2$, $Y_H = 0.67$, $K_{NO} = 0.5mg/l$, $K_S = 2mg/l$), the feature matrices F_1~F_7 are calculated. In order to improve the identifiability of the modeling errors, a row wise transformation is applied to all the feature matrices, and the improved feature matrices are shown in Table 1 where W, the transformation matrix, is chosen as $diag(1\ 1\ 8)$. Correspondingly, the feature equation (5) is changed to $\sum_{i=1}^{7} F_{iw} \int_{t_0}^{t} e^{-g(t-\tau)} d_i(\tau) d\tau = \bar{e}(t) = We(t)$.

Table 1: Feature matrices

$F_{1w} = WF_1$	$F_{2w} = WF_2$	$F_{3w} = WF_3$	$F_{4w} = WF_4$	$F_{5w} = WF_5$	$F_{6w} = WF_6$	$F_{7w} = WF_7$
1	1	−1	0	0	1	1 0
−1	−1	0	0	0	−1	0 0
0	−0.923	−1.499	8	8	−0.923	0 −1

With feature matrices F_{1w}~F_{7w}, set $\mathcal{S}_c$ is constructed as (with (9)),

$$\begin{aligned}\mathcal{S}_c = \{&\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3, \mathcal{S}_4, \mathcal{S}_5, \mathcal{S}_6, \mathcal{S}_7, \mathcal{S}_8\} = \{\Re(F_{1w}), \Re(F_{2w}), \Re(F_{3w}),\\ &\Re(F_{4w}), \Re(F_{7w}), \Re(F_{1w}) + \Re(F_{2w}), \Re(F_{1w}) + \Re(F_{3w}), \Re(F_{2w}) + \Re(F_{3w})\}\end{aligned}$$

With $\epsilon_2 = 0.2$, sets $\mathbf{M}_{\mathcal{S}_i}$, $i = 1, 2, \cdots, 8$, can be calculated as follows with (10),

$$\begin{array}{ll}\mathbf{M}_{\mathcal{S}_1} = \{F_{1w}\} & \mathbf{M}_{\mathcal{S}_2} = \{F_{2w}, F_{6w}\}\\ \mathbf{M}_{\mathcal{S}_3} = \{F_{3w}\} & \mathbf{M}_{\mathcal{S}_4} = \{F_{4w}, F_{5w}\}\\ \mathbf{M}_{\mathcal{S}_5} = \{F_{7w}, F_{3w}, F_{4w}, F_{5w}\} & \mathbf{M}_{\mathcal{S}_6} = \{F_{1w}, F_{2w}, F_{4w}, F_{5w}, F_{6w}\}\\ \mathbf{M}_{\mathcal{S}_7} = \{F_{1w}, F_{3w}\} & \mathbf{M}_{\mathcal{S}_8} = \{F_{2w}, F_{3w}, F_{6w}\}\end{array}$$

Simulation

System (14) and its observer were simulated for different modeling errors, with random noise added to all three measurements: $\bar{X}_r$, $S_{S,r}$ and $S_{NO,r}$. The conditions and parameters that were used in the simulations, when not specifically mentioned, are: $S_{O,in} = 0$, $K_O = 0.5\ mg.l^{-1}$, $K_{OI} = 1.5\ mg.l^{-1}$, b_r, $f_{P,r}$, $Y_{S,r}$, $K_{NO,r}$ and $K_{S,r}$ took the same values as those of the model (see above). The cases studied are,

$$\begin{array}{ll}\textbf{Case1}: S_{O,in} = 2mg.l^{-1}; & \textbf{Case2}: S_{O,in} = 10mg.l^{-1};\\ \textbf{Case3}: b_r = 0.3day^{-1}; & \textbf{Case4}: f_{P,r} = 0.3;\\ \textbf{Case5}: Y_{H,r} = 0.5; & \textbf{Case6}: K_{NO,r} = 1mg.l^{-1};\\ \textbf{Case7}: \mu_{max,r} = 1.2day^{-1}; & \textbf{Case8}: b_r = 0.3day^{-1}, K_{NO,r} = 1mg.l^{-1};\\ \textbf{Case9}: b_r = 0.28day^{-1}, K_{NO,r} = 0.2mg.l^{-1}; & \textbf{Case10}: \mu_{max,r} = 1.2day^{-1}, K_{NO,r} = 0.3mg.l^{-1};\end{array}$$

Note that for Case 1 and Case 2, the presence of oxygen in the influent not only introduces modeling error F_{1w} but also F_{2w} due to the oxygen inhibition of denitrification.

The identification results are shown in Table 2, where r_i denotes $\|V_{i2}^T We(t)\|_2 / \|V_{i1}^T We(t)\|_2$, the ratio of the projection of $We(t)$ on $\mathcal{S}_i^{\perp}$ to that on $\mathcal{S}_i$, for $i = 1, 2, \cdots, 8$. M_e is obtained using equation (11) where f_i is obtained by comparing r_i with $\epsilon_1 = 0.1$. The following observations are made on Table 2.

- For Case 2 to Case 8, the algorithm successfully localizes the modeling errors to a subset of $\mathbf{M_p}$ in which the real modeling errors are contained ($M_e \subseteq M_{real}$);
- For Case 1 and Case 9, the modeling errors identified are only a subset of those that are really associated with the model. In Case 1, both F_{1w} and F_{2w} are associated with the model due to the existence of oxygen in the influent. F_{2w} is not identified because the influence of the oxygen inhibition of denitrification, when the oxygen concentration in the reactor is low, is much smaller than the influence of oxygen on carbon compounds consumption and biomass growth. A similar reasoning explains why F_{3w} is not identified in Case 9. This is considered to be a nice property since it means that the algorithm will only identify the most significant modeling errors that have caused the invalidity of the model.

Table 2: Identification results

Case	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	M_e	M_{real}
1	0.14	0.87	1.73	7.38	0.98	**0.00**	**0.06**	0.27	$\{F_{1w}\}$	$\{F_{1w}, F_{2w}\}$
2	0.30	1.17	1.29	3.39	0.92	**0.00**	0.12	0.31	$\{F_{1w}, F_{2w}, F_{4w}, F_{5w}, F_{6w}\}$	$\{F_{1w}, F_{2w}\}$
3	2.53	6.42	**0.03**	0.63	**0.02**	0.42	**0.01**	**0.00**	$\{F_{3w}\}$	$\{F_{3w}\}$
4	37.6	1.62	0.64	**0.03**	**0.02**	**0.00**	0.47	0.34	$\{F_{4w}, F_{5w}\}$	$\{F_{4w}\}$
5	27.8	1.42	0.71	**0.04**	**0.02**	**0.00**	0.47	0.34	$\{F_{4w}, F_{5w}\}$	$\{F_{5w}\}$
6	0.65	**0.00**	8.03	1.54	0.74	**0.00**	0.24	**0.00**	$\{F_{2w}, F_{6w}\}$	$\{F_{6w}\}$
7	1.40	5.81	0.38	1.40	**0.01**	0.69	0.28	0.21	$\{F_{7w}, F_{3w}, F_{4w}, F_{5w}\}$	$\{F_{7w}\}$
8	0.23	0.53	2.65	8.11	0.67	0.19	0.23	**0.00**	$\{F_{2w}, F_{3w}, F_{6w}\}$	$\{F_{3w}, F_{6w}\}$
9	1.02	0.24	2.65	0.99	0.67	**0.09**	0.23	**0.00**	$\{F_{2w}, F_{6w}\}$	$\{F_{3w}, F_{6w}\}$
10	0.47	1.02	1.32	9.52	0.37	0.45	0.35	0.15	M_p	$\{F_{6w}, F_{7w}\}$

- For Case 10, the algorithm fails to give any conclusion on the modeling errors.

CONCLUDING REMARKS

An approach to modeling error identification has been presented in this paper. It consists of three major steps: (i) analyzing the process and its model to predict the possible modeling errors and extract the feature matrices of these errors, (ii) designing observers for the process based on the invalid model to generate the featured signal, and (iii) analyzing the featured signal to identify the modeling errors. Since the approach requires enumeration of the modeling errors in advance, it is more suitable for mechanistic models than for black box ones. For grey box models, one may consider each small black box as a unit and localize the modeling errors to this level.

The method has been illustrated for a denitrification reactor model in which different modeling error types could be identified, *i. e.* state dimension deficiency, model structure error and parameter value inaccuracy. The potential for more efficient model improvement is evident as efforts for model adjustment e. g. via additional experiments can be directed much better.

REFERENCE

Bogaert, H., Vanderhasselt, A., Gernaey, K., Yuan, Z., Thoeye, C. and Verstraete, W. (1997) A new sensor based on pH-effect of the denitrification process, *J. of Env. Eng.* . (in press)

Coen, F., Vanderhaegen, B., Boonen, I., Vanrolleghem, P. and Van Meenen, P. (1996) Nitrogen removal upgrade of a WWTP within existing reactor volumes: A simulation supported scenario analysis. *Wat. Sci. Tech.* , 34(3-4), 339-346

Henze, M. *et al.* (1987) Activated Sludge Model No.1, *IAWPRC Scientific and Technical Report No.1*

Yuan, Z. and Vansteenkiste G. (1996a) An Approach to Validation of Stochastic Dynamic Models with Initial State Uncertainty, *SCS Transactions on Simulation*, Vol. 13, No. 1, 3-18.

Yuan, Z., Vangheluwe, H. and Vansteenkiste, G. (1996b) An Observer-based Approach to Modeling Error Identification, submitted to *SCS Transactions on Simulation.*