

## 1. HRT Optimization Based on ASM2d-Guided Reward

Based on the Equation (7) in the main article paper,  $x_i^j$  is a function of time  $k$ , and it is also the functions on  $x_i^1, x_i^2, \dots, x_i^m$  at time  $k$  based on ASM2d. As we consider whatever function  $x_i^j$  might be on any variable, the variation of  $x_i^j$  is the same from time 0 to time  $k_0$  at this variable. Hence, the following integral function can be evaluated according to the numerical iteration of the ASM2d:

$$x_{k_0}^j \Big|_i - x_0^j \Big|_i = \int_0^{k_0} g_i^j \left( x_k^1 \Big|_i, \dots, x_k^m \Big|_i \right) \cdot dk = \sum_{t=0}^{k_0} g_i^j \left( x_k^1 \Big|_i, \dots, x_k^m \Big|_i \right) \quad (1)$$

Thus, for all the concentrations of the components in the anaerobic reaction tank ( $i=1$ ), the value of Equation (1) can also be calculated through the numerical iteration method of ASM2d, as shown in Equation (2):

$$\int_0^{k_0} g_i \left( x_k^1 \Big|_i, \dots, x_k^m \Big|_i \right) \cdot dk = \sum_{k=0}^{k_0} g_i \left( x_k^1 \Big|_i, \dots, x_k^m \Big|_i \right) \quad (2)$$

Thus, it can be seen that,  $\int_0^{k_0} g_i \left( x_k^1 \Big|_i, \dots, x_k^m \Big|_i \right) \cdot dk$  is the integral upper limit function on time  $k$ , so, we set  $\int_0^k g_i \left( x_k^1 \Big|_i, \dots, x_k^m \Big|_i \right) \cdot dk = F^i(k)$ , Equation (3) is obtained through formula transformation of Equation (1):

$$s_1 - s_0 = F^1(k_0 \Big|_1) - F^1(0) \quad (3)$$

In Equation (3),  $s_0$  is the influent value,  $F^1(0)$  is the initial value of  $g_1 \left( x_k^1 \Big|_1, \dots, x_k^m \Big|_1 \right)$  when  $k=0$ ; that is,  $F^1(0) = g_1(s_0)$ , and the period from 0 to  $k_0 \Big|_1$  means the  $i$ th reactor tank HRT. Thus, Equation (3) can be replaced, as follows:

$$s_1 = F^1(k_0 \Big|_1) - g_1(s_0) + s_0 \quad (4)$$

We set  $\sigma_1(s_0, k_0 \Big|_1) = F^1(k_0 \Big|_1) - g_1(s_0) + s_0$ , then:

$$s_1 = \sigma_1(s_0, k_0 \Big|_1) \quad (5)$$

Thus, it can be obtained that  $s_1$  is the function of  $s_0$  and  $t_0$ . In exactly the same way, it can be obtained that  $s_2 = \sigma_2(s_1, k_0 \Big|_2)$  and  $s_3 = \sigma_3(s_2, k_0 \Big|_3)$ ; thus, we set:

$$s_i = \sigma_i(s_{i-1}, k_0 \Big|_i) \quad (6)$$

When conducting the optimal analysis of the HRTs in anaerobic, anoxic, and oxic tanks, the IRR in the QL-AAO system is regarded as the fixed value, and vice versa; when conducting the optimal analysis of the IRR, the HRT is regarded as the fixed value. Therefore, the real effluent concentration value from the anoxic reaction tank is shown in Equation (7):

$$s_1 \Rightarrow \frac{s_1 + s_3 \times q}{1 + q} \quad (7)$$

where  $q$  represents the IRR.

According to Equation (8) in the main article, the expression of  $r_i$  could be obtained in terms of  $s_i$  and  $k_0|_i$ :

$$r_i = r(s_i, k_0|_i) \quad (8)$$

Thus, the optimization expression of the wastewater treatment process can be transferred into Equation (9):

$$\pi^* = \arg \max_{\pi} V^{\pi}(s_i) \quad (9)$$

where  $\pi^*$  represents the  $\pi$  function when the maximum value of  $V^{\pi}(s_i)$  is obtained. Equation (10) can be further expressed as follows:

$$\pi^* = \arg \max_{k_0|_i} V^{\pi}(s_i) \quad (10)$$

Consequently, the value of  $k_0|_i$  under each process  $i$  can be achieved when the maximum value of  $V^{\pi}(s_i)$  is obtained.

Here we suppose in this study, the boundary condition of the whole HRT for the AAO system is less than or equal to 8 h, as shown in Equation (11):

$$\sum_{i=1}^3 k_0|_i \leq 8 \quad (11)$$

Thus, based on Equation (5) in main article paper and Equation (11), we can obtain the value of  $k_0|_i$  when the boundary condition of HRT is less than or equal to 8. Based on the analyses above, Equation (10) can be replaced, as shown in Equation (12):

$$\pi^* = \arg \max_{k_0|_i} V^{\pi}(s_i) = \arg \max_{k_0|_i} \left\{ r_i(s_i, k_0|_i) + V^{\pi^*}[\sigma_i(s_i, k_0|_i)] \right\} \quad (12)$$

Hence, when  $V^{\pi}(s)$  is maximum in this QL-AAO system, the Q function under the optimal solution of HRT can be obtained by Equation (13):

$$Q(s_i, k_0|_i) = r_i(s_i, k_0|_i) + V^{\pi^*}[\sigma_i(s_i, k_0|_i)] \quad (13)$$

Thus, the optimization expression of the wastewater treatment process can be further transferred into Equation (14):

$$\pi^* = \arg \max_{k_0|_i} Q_i(s_i, k_0|_i) \quad (14)$$

In terms of the above expression, the improved Q function is iterated and updated, as shown in Equation (15):

$$Q(s_i, k_0|_i) = r_i(s_i, k_0|_i) + \max_{k_0|_{i+1}} Q(s_{i+1}, k_0|_{i+1}) \quad (15)$$

## 2. IRR Optimization Based on ASM2d-Guided Reward

In terms of Equation (6) and Equation (7), for instance, the transition function of the effluent concentrations from anoxic tank is shown in Equation (16):

$$s_2 = \sigma_1 \left( \frac{s_1 + \sigma_2 (s_2, k_0|_2) \times q}{1 + q}, k_0|_1 \right) \quad (16)$$

Thus, the effluent value from anoxic reaction tank can be obtained on the functions of anaerobic reaction tank,  $k_0|_1$ ,  $k_0|_2$  and  $q$ . Then, its transition function is set as  $\sigma_q$ . Equation (17) can be obtained as follows:

$$s_2 = \sigma_q (s_1, k_0|_1, k_0|_2, q) \quad (17)$$

When conducting the optimal analysis of the IRR, the HRTs are regarded as the fixed values. In this step,  $k_0|_1$  and  $k_0|_2$  are regarded as the fixed values. Hence, the effluent concentration from the anoxic reaction tank can be obtained based on the transition function of  $q$  (Equation (18)):

$$s_2 = \sigma_q (s_1, q) \quad (18)$$

By combining Equation (7), Equation (11), and Equation (18), the expression of the reward function of anoxic reaction tank ( $i=1$ ) is shown in Equation (19):

$$r_1 = r_q (s_1, q) \quad (19)$$

The Q function expression of the IRR optimization is shown in Equation (20):

$$Q(s_1, q) = r_q(s_1, q) + V^{\pi^*} [\sigma_q(s_1, q)] \quad (20)$$

Then, the optimal control parameter  $q$  of the IRR for the QL-AAO is obtained based on a function  $\pi$  that depends on  $q$ , as shown in Equation (21):

$$\pi^* = \arg \max_q Q(s_1, q) \quad (21)$$

**Table S1.** Kinetic parameters and values in ASM2d model.

Symbol	Definition	Unit	Value
Hydrolysis process			
$K_h$	Hydrolysis rate constant	$d^{-1}$	3.00
$\eta_{NO3}$	Correction factor for reduction of anoxic hydrolysis rate	-	0.60
$\eta_{fe}$	Correction factor for reduction of anaerobic hydrolysis rate	-	0.10
$K_{O2}$	Oxygen saturation/inhibition coefficient	$g(O_2)/m^3$	0.20
$K_{NO3}$	Nitrate saturation/inhibition coefficient	$g(N)/m^3$	0.50
$K_X$	Saturation coefficient of particulate COD	$g(X_S)/g(X_H)$	0.10
Heterotrophic bacteria			
$\mu_H$	Maximum matrix growth rate	$g(X_S)/(g(X_H) - d)$	6.00
$q_{fe}$	Maximum fermentation rate	$g(X_F)/(g(X_H) - d)$	3.00
$\eta_{NO3}$	Denitrification reduction correction factor	-	0.8
$b_H$	Specific attenuation (death) rate	$d^{-1}$	0.40
$K_{O2}$	Oxygen saturation coefficient	$g(O_2)/m^3$	0.20
$K_F$	Growth saturation coefficient based on $S_F$	$g(COD)/m^3$	4.00
$K_{fe}$	$S_F$ fermentation saturation coefficient	$g(COD)/m^3$	4.00
$K_A$	Growth saturation coefficient based on $S_A$	$g(COD)/m^3$	4.00
$K_{NO3}$	Nitrate saturation coefficient	$g(N)/m^3$	0.50
$K_{HN4}$	$S_{NH4}$ saturation coefficient	$g(N)/m^3$	0.05
$K_P$	$S_{PO4}$ saturation coefficient	$g(P)/m^3$	0.01
$K_{ALK}$	$S_{ALK}$ saturation coefficient	$mol(HCO_3)/m^3$	0.10
Phosphorus accumulating organisms			
$q_{PHA}$	PHA Storage rate constant	$g(X_{PHA})/(g(X_{PAO}) - d)$	3.00
$q_{PP}$	Poly-P Storage rate constant	$g(X_{PP})/(g(X_{PAO}) - d)$	1.50
$\mu_{PAO}$	PAOs Maximum growth rate	$d^{-1}$	1.00
$\eta_{NO3}$	Correction factors for hypoxic activity reduction	-	0.60
$b_{PAO}$	$X_{PAO}$ Bacteriolysis rate constant	$d^{-1}$	0.20
$b_{PP}$	$X_{PP}$ Decomposition rate constant	$d^{-1}$	0.20
$b_{PHA}$	$X_{PHA}$ Decomposition rate constant	$d^{-1}$	0.20
$K_{O2}$	Oxygen saturation/inhibition coefficient	$g(O_2)/m^3$	0.20
$K_{NO3}$	Nitrate saturation/inhibition coefficient	$g(N)/m^3$	0.50
$K_A$	$S_A$ saturation coefficient	$g(COD)/m^3$	4.00
$K_{HN4}$	$S_{NH4}$ saturation coefficient	$g(N)/m^3$	0.05
$K_{PS}$	Poly-P Saturation coefficient of stored phosphorus	$g(P)/m^3$	0.20
$K_P$	$S_{PO4}$ saturation coefficient	$g(P)/m^3$	0.01
$K_{ALK}$	$S_{ALK}$ saturation coefficient	$mol(HCO_3)/m^3$	0.10
$K_{PP}$	Polyphosphate saturation coefficient	$g(X_{PP})/g(X_{PAO})$	0.01
$K_{MAX}$	$X_{PP}/X_{PAO}$ maximum ratio	$g(X_{PP})/g(X_{PAO})$	0.34
$K_{IPP}$	$X_{PP}$ inhibition coefficient of storage	$g(X_{PP})/g(X_{PAO})$	0.02
$K_{PHA}$	PHA saturation coefficient	$g(X_{PHA})/g(X_{PAO})$	0.01
Autotrophic bacteria			
$\mu_{AUT}$	$X_{AUT}$ maximum growth rate	$d^{-1}$	1.00
$b_{AUT}$	$X_{AUT}$ decay rate	$d^{-1}$	0.15
$K_{O2}$	Oxygen saturation coefficient	$g(O_2)/m^3$	0.50
$K_{HN4}$	$S_{NH4}$ saturation coefficient	$g(N)/m^3$	1.00
$K_{ALK}$	$S_{ALK}$ saturation coefficient	$mol(HCO_3)/m^3$	0.50
$K_P$	$S_{PO4}$ saturation coefficient	$g(P)/m^3$	0.01

**Table S2.** Influent components of the AAO system in the ASM2d simulation.

<i>j</i>	Components	Description	Values	Units
1	$S_{O_2}$	Dissolved oxygen	0, 3	$gO_2 / m^3$
2	$S_F$	Readily biodegradable organic compound	0	$gCOD / m^3$
3	$S_A$	Fermentation products	244.62–385.26	$gCOD / m^3$
4	$S_{NH_4}$	Dissolved ammonium ion	13.12–25.15	$gN / m^3$
5	$S_{PO_4}$	Inorganic soluble Phosphorus	1.06–3.96	$gP / m^3$
6	$S_{ALK}$	Bicarbonate alkalinity	7.00	$molHCO_3^- / m^3$
7	$S_{NO_3}$	Dissolved nitrate ion	0	$gN / m^3$
8	$S_{N_2}$	Nitrogen (N <sub>2</sub> ), 20 °C, 79 kPa	0	$gN / m^3$
9	$X_{PAO}$	Phosphate accumulating bacteria	247.5	$gP / m^3$
10	$X_S$	Slowly biodegradable organic compounds	32.4	$gCOD / m^3$
11	$X_{TSS}$	Particulate matters	3207.7	$gTSS / m^3$
12	$X_I$	Inert organic matters	435.6	$gCOD / m^3$
13	$X_H$	Heterotrophic bacteria	2080.3	$gCOD / m^3$
14	$X_{PP}$	Polyphosphate stored in PAO	27.6	$gP / m^3$
15	$X_{PHA}$	Intracellular storage product PHA in PAO	30.8	$gCOD / m^3$
16	$X_{AUT}$	Autotrophic bacteria	353.5	$gCOD / m^3$