



## OPEN

## A Novel Protocol for Model Calibration in Biological Wastewater Treatment

## SUBJECT AREAS:

ENVIRONMENTAL  
BIOTECHNOLOGY

ENVIRONMENTAL SCIENCES

APPLIED MICROBIOLOGY

WATER MICROBIOLOGY

Ao Zhu<sup>1,3</sup>, Jianhua Guo<sup>1,2</sup>, Bing-Jie Ni<sup>2</sup>, Shuying Wang<sup>1</sup>, Qing Yang<sup>1</sup> & Yongzhen Peng<sup>1</sup>

<sup>1</sup>Key Laboratory of Beijing for Water Quality Science and Water Environmental Recovery Engineering, Engineering Research Center of Beijing, Beijing University of Technology, Beijing 100124, PR China, <sup>2</sup>Advanced Water Management Centre (AWMC), The University of Queensland, St Lucia, Brisbane, QLD 4072, Australia, <sup>3</sup>Tsinghua Holding Human Settlements Environment Institute, Beijing 100083, PR China.

Received  
19 November 2014Accepted  
20 January 2015Published  
16 February 2015

Correspondence and requests for materials should be addressed to J.G. (gjh@bjut.edu.cn) or Y.P. (pyz@bjut.edu.cn)

Activated sludge models (ASMs) have been widely used for process design, operation and optimization in wastewater treatment plants. However, it is still a challenge to achieve an efficient calibration for reliable application by using the conventional approaches. Hereby, we propose a novel calibration protocol, i.e. Numerical Optimal Approaching Procedure (NOAP), for the systematic calibration of ASMs. The NOAP consists of three key steps in an iterative scheme flow: i) global factors sensitivity analysis for factors fixing; ii) pseudo-global parameter correlation analysis for non-identifiable factors detection; and iii) formation of a parameter subset through an estimation by using genetic algorithm. The validity and applicability are confirmed using experimental data obtained from two independent wastewater treatment systems, including a sequencing batch reactor and a continuous stirred-tank reactor. The results indicate that the NOAP can effectively determine the optimal parameter subset and successfully perform model calibration and validation for these two different systems. The proposed NOAP is expected to use for automatic calibration of ASMs and be applied potentially to other ordinary differential equations models.

Activated sludge is the most widely used biological technology for treating domestic and industrial wastewater. After its development with 100 years of history, many novel and modified processes have been developed to meet the more and more stringent discharge and emission limits. However, most of operating systems are suffering some drawbacks, such as substantial energy consumption, excessive greenhouse gas emission, and labour-intensive industry. As a powerful tool, Activated Sludge Models (ASMs) have proven to be very useful in process design, operation and optimization<sup>1,2</sup>. To date, ASMs for the simulation of biological nutrients removal processes have been updated from the first version of ASM1 to more complicated extensions, including ASM2, ASM2d, and ASM3<sup>3</sup>, and further to the extended ASM3s<sup>4–10</sup> in order to satisfy various requirements.

However, ASMs are large and overparameterized models in terms of having many stoichiometric and kinetic parameters. Some of the model parameters as well as the model structure have to be adjusted, since microbial community structure and dominant species can vary in different wastewater treatment systems with different influent characteristics or operation schemes<sup>6,7,11–14</sup>. In addition, the collected data from full-scale plants as well as pilot- or lab- scale reactors can hardly provide reliable estimations of all the parameters simultaneously due to the well-known problem of poorly identifiable parameters<sup>15–20</sup>. Thus, the approach to properly select the subsets of parameters for model calibration plays a crucial role on simulation results and model applications<sup>21–25</sup>.

Until now, substantial studies have been conducted to develop effective model calibration approaches, which can be distinguished into two major categories: the conventional experience-based approaches and the systems analysis approaches<sup>22</sup>. The experience-based approaches that were proposed by WERF, BIOMATH, STOWA, CALAGUA, HSG etc., import programmatic flow based on experts' knowledge and experience<sup>22,26,27</sup>. However, these calibration protocols require specific experimental designs and data processing methods to decouple the ASMs to be small and simple sub ones. The data for model calibration and validation may be from different feeding and operation conditions, and some of the parameters are fixed in order to obtain accurate estimation of the concerning ones. This approach might also ignore shifts in microbial community, eventually resulting in error estimations. Recently, the Good Modeling Practice (GMP) task group from IWA have emphasized that a standardized modeling procedure is needed to distinguish parameter subset that is identifiable with the available data for experience-based approaches<sup>28</sup>.

Systematical analysis approaches have attracted much more attention<sup>15,18,29–32</sup>, which mainly consist of parameter identification, sensitivity analysis, and error propagation<sup>32</sup>. Conventional parameter subset identifiability



analyzing methods are mainly based on the local sensitivity analysis<sup>21,22</sup>. Recently, global sensitivity analysis (GSA) becomes a promising method for parameter identifiability analysis<sup>33–35</sup>, such as the Morris screening method<sup>36</sup>, Fourier Amplitude Sensitivity Testing (FAST), Extended-FAST, the Sobol's method<sup>37–41</sup>, and standardized regression coefficient (SRC) method<sup>33</sup>. Structural identification through Taylor series expansion<sup>16</sup>, generating series<sup>42</sup>, similarity transformation approach<sup>43</sup> and differential algebra approach<sup>17,44</sup>, have been proposed in order to screen out the identifiable combinations. Based on the symbolic algebra system, structural identification has been applied to ASM-type simple sub-models successfully<sup>16,17</sup>. However, structural identification based on symbolic algebra system could be time consuming and unreliable due to the large number of parameters in ASMs. Therefore, it is highly desired to develop a numerical integrated approach by including both proper global parameter sensitivity analysis method and correlation analysis method for the identifiability of the parameters of the ASMs.

The main objective of this study is to develop a novel systematical approach, namely Numerical Optimal Approaching Procedure (NOAP), for efficient calibration and validation of ASMs, taking an extended ASM3 as example. The NOAP integrates a new screening global parameter sensitivity analysis method (Derivative based Global Sensitivity Measures, DGSM) and a numerical correlation analysis method (based on the pseudo-global covariance matrix), with Genetic Algorithm (GA) as the global optimization (GO) method for parameter estimation. The validity and applicability of the approach for efficient model calibration are tested by independent experimental data from a Sequential Batch Reactor (SBR) activated sludge system<sup>45</sup> and a Continuous Stirred-Tank Reactor (CSTR) activated sludge system<sup>46</sup>.

## Results

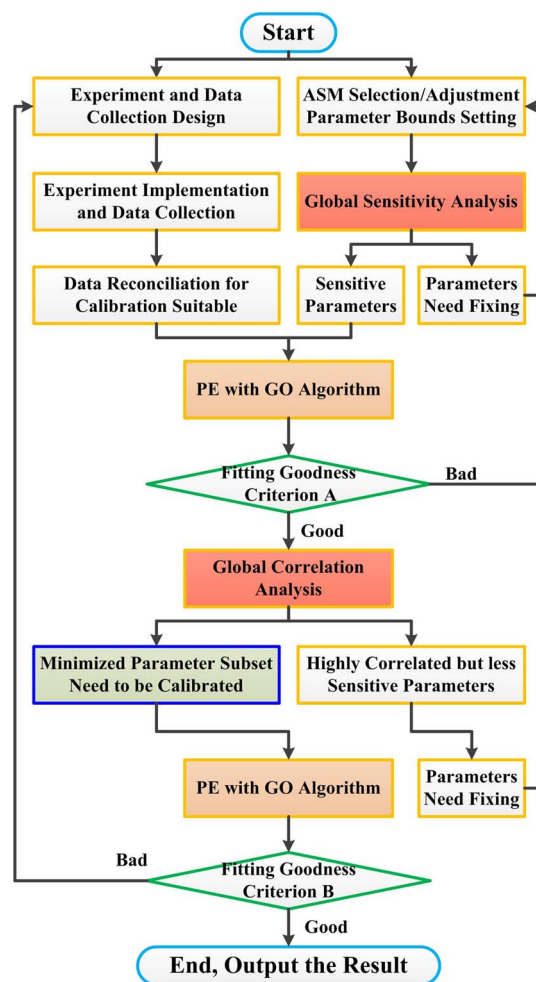
**Development of the NOAP approach.** The iterative algorithm of NOAP approach, as shown in the main scheme flow in Fig. 1, mainly consists of four steps in succession.

**Step I: Model selection and data preparation.** The user should firstly choose an appropriate model and extend it to satisfy the specific requirement from the current available models in previous studies<sup>1,47</sup>. Data collected from lab-, pilot-, and full-scale reactors are often used, but the amount and quality need reconciliation for the basic request of model calibration. The data collection and reconciliation can refer to the procedures suggested by Rieger et al.<sup>28</sup>.

**Step II: GSA for factors fixing.** Global parameter sensitivity analysis would be performed only if (i) the initial values of the state variables are set in a reasonable range based on data analysis and ii) parameter boundaries (lowest and highest) for global sensitivity analysis are proper. A fitting goodness criterion (such as  $A = 85\%$ ) between data and simulation results is used to judge parameter subset remained after factor fixing from DGSM analysis result. If the result is satisfied ( $R^2 \geq A$ ), the parameters with lower sensitivity will be removed with trial, then repeating estimation trial until a proper small enough parameter subset is selected out. Otherwise, more non-sensitive ones should be added into the target subset one by one with the iterative parameter estimation trial. If parameters are all estimated simultaneously without a satisfied result, boundaries of the parameters or model structure might be improper and need further modifications.

**Step III: For the scale reduced parameter subset, a pseudo-global correlation matrix is calculated.** If the correlation coefficient of any parameter pair is high enough ( $>0.95$  as example), then such parameter combinations are located to be highly correlated parameters due to the statistical average effect of the numerical algorithm. Afterwards, less interested members and correlation crossing ones (high correlated with not only just one parameter) can be fixed.

**Step IV: The final parameter estimation is performed to check the efficiency of the procedure.** Another fitting goodness criterion (such



**Figure 1** | Scheme flow of the proposed NOAP (PE, parameter estimation; GO, global optimization).

as  $B = 95\%$ ) is set to roughly test the quality of the observables' collected data.

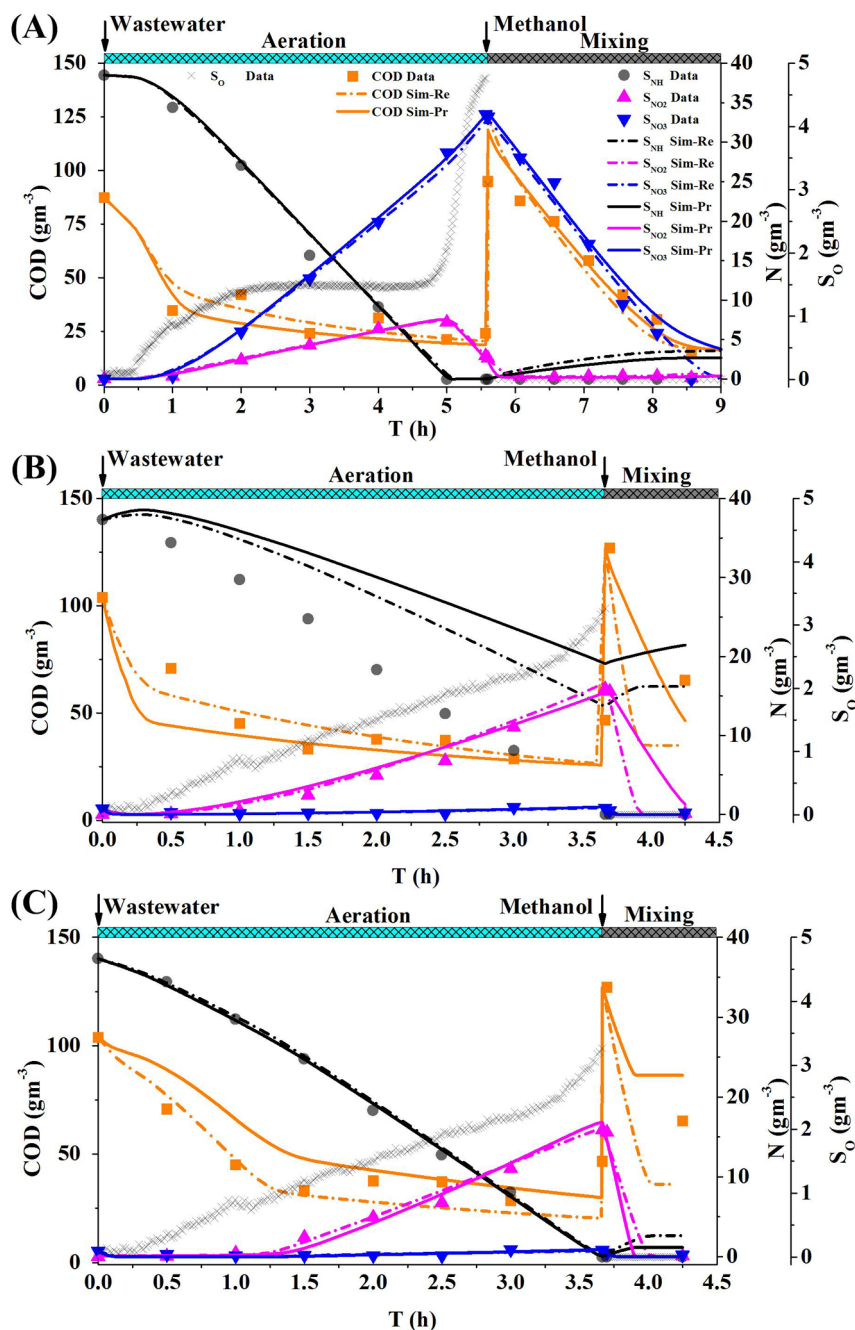
### Preliminary model calibration based on conventional approach.

According to Kaelin et al.<sup>6</sup>, the parameter subset  $[\mu_{AOB}, \mu_{NOB}, \eta_{H,NO_3}, \eta_{H,NO_2}, \eta_{H,end,NO_3}, \eta_{H,end,NO_2}]$  is firstly adopted and used to perform preliminary model calibration. Through this step, it is expected to: (1) determine whether the experimental data from the target system could be simulated by the proposed model; (2) suggest proper initial values of the state variables; and (3) verify directly whether the reported parameter subset from literature is efficient or not.

**Scenario 1: SBR system.** As shown in Fig. 2(a), the simulated results (Sim-Re series) have a good fitness with the experimental data in terms of the profiles of COD, nitrite and nitrate concentrations. However, an unacceptable accumulation of ammonium during anoxic period is observed in the model simulations, which might be due to improper parameter values. Moreover, the validation (Sim-Re series) illustrated in Fig. 2(b) further verifies that the simulated ammonium results are much higher than the experimental data ( $R^2_{(NH)} \text{ Sim-Re} = 0.5011$ ).

Regarding the SBR system for nitrogen removal, the parameter subset selected from literature can partially calibrate the model to describe most of the experimental data. However, the extra efforts are needed to select other parameter subsets due to the failure of model validation.

**Scenario 2: CSTR system.** As shown in Fig. 3(a), the simulated (Sim-Re series) COD and ammonium concentrations in effluent fit well



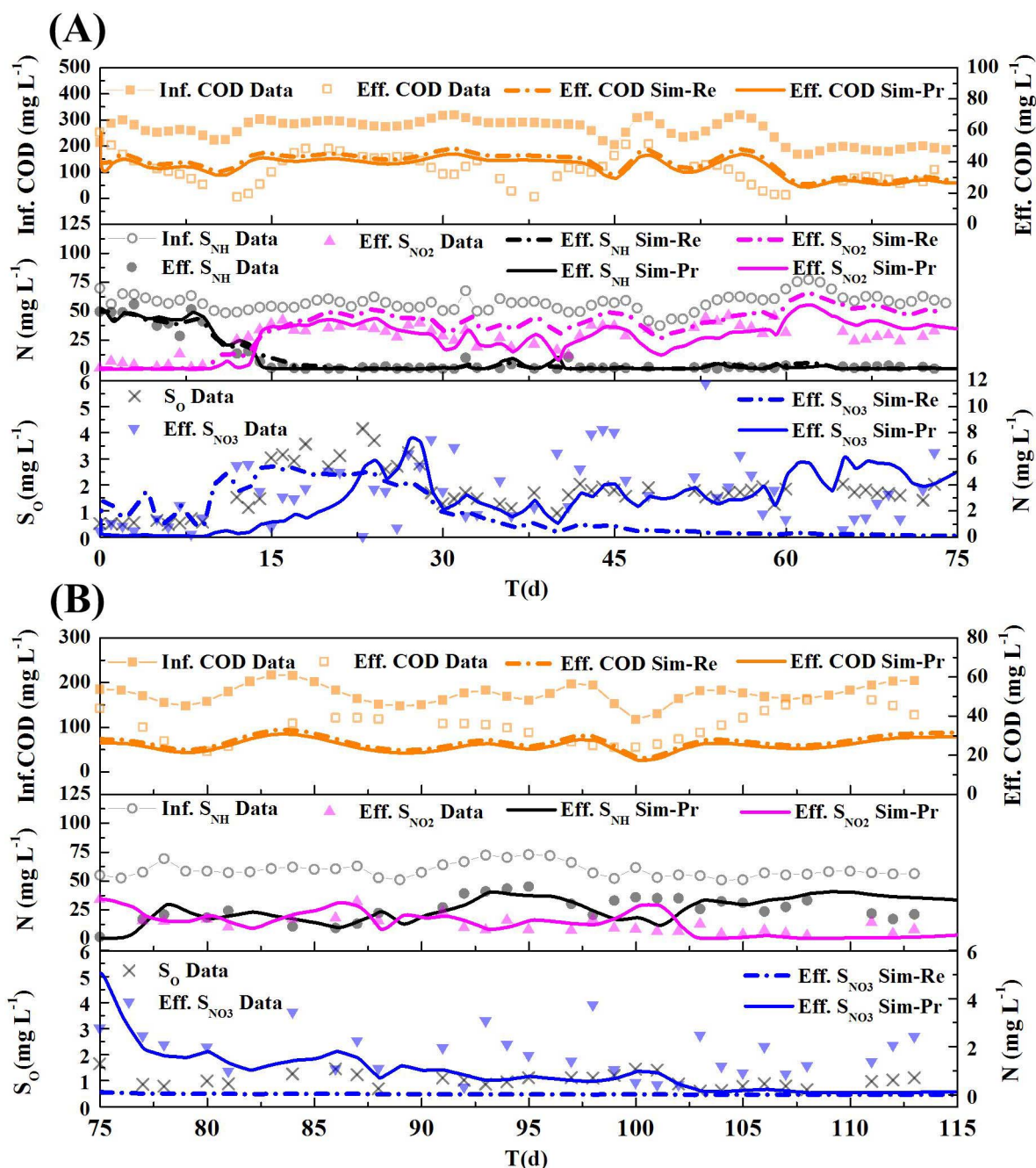
**Figure 2** | Calibration and validation results with the parameter subset suggested by the reference 6 and the proposed procedure NOAP for target SBR system. (A, Calibration results with the parameter subset recommended by the reference (Sim-Re)<sup>6</sup> and the NOAP procedure (Sim-Pr); B, Validation for the calibration results as Fig. 2(a) presented, legends are the same as Fig. 2(a); C, Validation adjusted for the calibration with the parameter subset recommended by the reference<sup>6</sup> and the proposed procedure NOAP with the suggestions of the global parameter sensitivity and correlation analysis results (Here  $\eta_{H,NO_2}$  is added to the calibration subset), legends are the same as Fig. 2(a).)

with the experimental data. However, the simulated nitrite concentrations are slightly higher than the experimental data ( $R^2(\text{Eff. } S_{NO_2} \text{ Sim-Re}) = -0.3617$ ), and the simulated nitrate concentrations are distinctly lower than the experimental data after the 30<sup>th</sup> day ( $R^2(\text{Eff. } S_{NO_3} \text{ Sim-Re}) = -1.0290$ ). The validation results (Sim-Re series) in Fig. 3(b) suggest that all the measured nitrogen species concentrations (including ammonium, nitrite and nitrate) deviate distinctly from the experiment data ( $R^2(\text{Eff. } S_{NH} \text{ Sim-Re}) = -3.2151$ ;  $R^2(\text{Eff. } S_{NO_2} \text{ Sim-Re}) = -16.4224$ ;  $R^2(\text{Eff. } S_{NO_3} \text{ Sim-Re}) = -3.3828$ ). This might be attributed to the improper parameter values of the dissolved oxygen (DO)-related kinetic coefficients, such as  $K_H$ ,  $O_2$ ,  $inh$ ,  $K_{AOB}$ ,  $O_2$ ,

and  $K_{NOB, O_2}$ . Hence, the model calibration is failed when applying the literature reported parameter subset<sup>6</sup> into the CSTR system.

The preliminary calibration results from SBR and CSTR systems indicate that the parameter sensitivity vary across different activated sludge systems in the application of same model. Moreover, the parameter subset reported in literature could lead to poor model calibration results if fixing the remained parameters (e.g., the DO related parameters in Scenario 2). Thus, reliable parameter sensitivity analysis should be imported for model application in different activated sludge systems. It should be noted that the parameter subset reported in literature could be still useful for qualitatively model





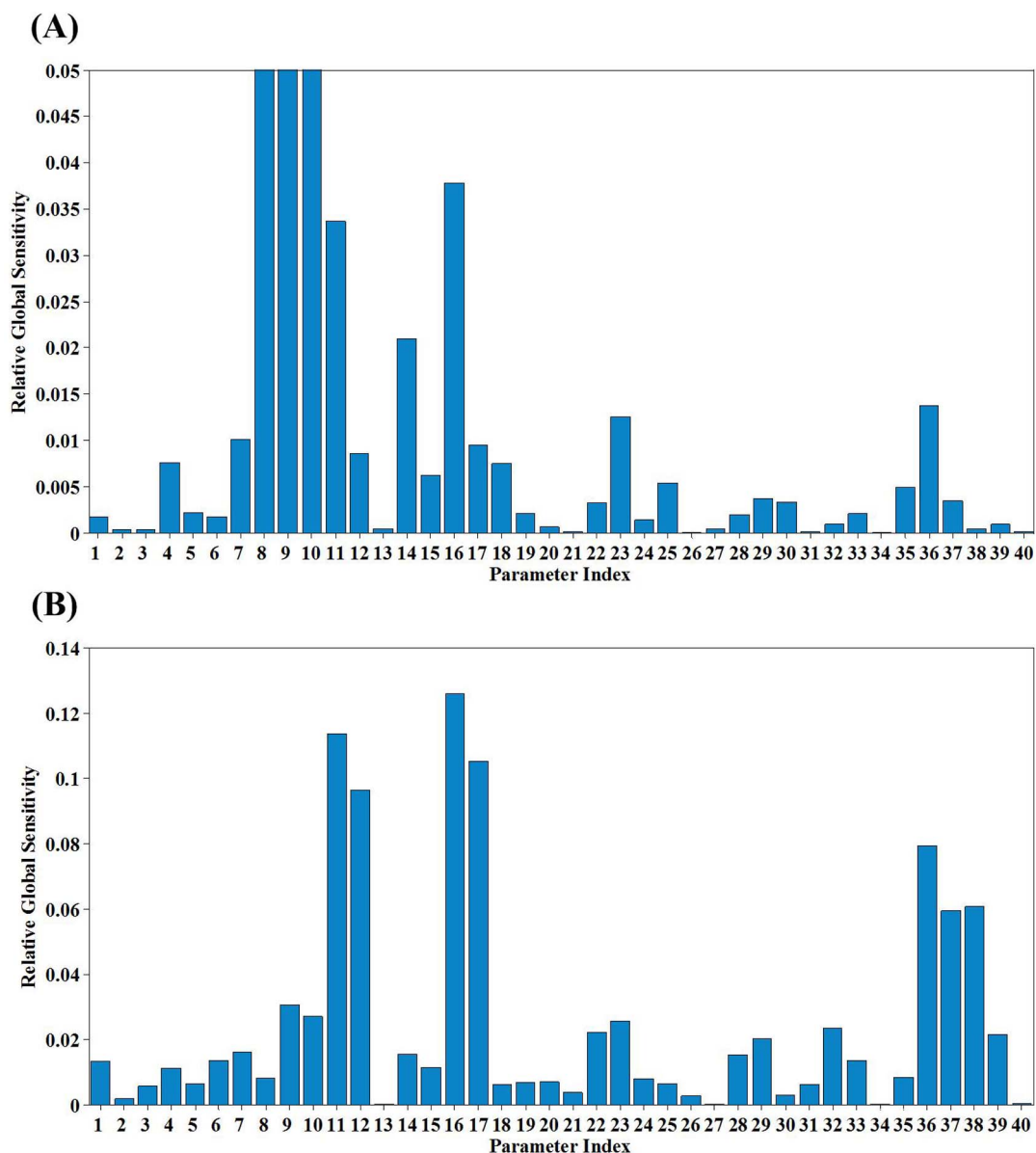
**Figure 3** | Calibration and validation with the parameter subset recommended by the reference<sup>6</sup> and the proposed procedure NOAP for target CSTR system (A, Calibration results with the parameter subset suggested by the reference<sup>6</sup> and the NOAP procedure (Sim-Pr); B, Validation for the calibration results as Fig. 3(a) presented.)

predictions, although it may not serve as a precise and quantitative predictor to match all experiment data.

**Parameter sensitivity and correlation analysis.** According to the preliminary calibration and validation, it can be concluded that the extended ASM3 model could predict the data only if a proper parameter subset is selected. However, the parameter subset reported in literature<sup>6</sup> cannot achieve satisfied calibration or validation results for both SBR and CSTR systems. Thus, numeric global parameter sensitivity and correlation analysis are performed to determine a more proper parameter subset.

In order to assess the identification of all the related parameters listed in Table S3 (SI), parameter sensitivity and correlation analysis

are conducted for both SBR and CSTR systems. The proposed iterative trials (Fig. 1) are implemented in order to capture the best-fit parameter values. The results are shown in Fig. 4 and Fig. 5, respectively. It can be found that both sensitivity analysis and correlation analysis results are quite different between the SBR system and CSTR system. Furthermore, parameters indexed from 1 to 6 in Table S4 (SI) exhibit a similar low sensitivity for both SBR and CSTR systems, suggesting the composition coefficients could be fixed as default values due to their low sensitivity. However, most of the composition coefficients strongly depend on the influent wastewater characteristics. Therefore, it is recommended that the N content of biomass can be fixed as default values from references, while others (e.g. COD transformation fractions) need to be obtained experimentally.



**Figure 4 | Global parameter sensitivity analysis results.** (A, Parameter sensitivity analysis for the SBR system (parameter 8, 9 and 10 reached 0.20, 0.13 and 0.46, respectively; B, Parameter sensitivity analysis results for the CSTR system; The indexes and their corresponding parameters are listed as the follows: 1-  $i_{N,SS}$ , 2-  $i_{N,XB}$ , 3-  $i_{N,BM}$ , 4-  $f_{XB}$ , 5-  $Y_{H,O_2}$ , 6-  $Y_{H,NO_3}$ , 7-  $Y_{H,NO_2}$ , 8-  $Y_{STO,O_2}$ , 9-  $Y_{STO,NO_3}$ , 10-  $Y_{STO,NO_2}$ , 11-  $Y_{AOB}$ , 12-  $Y_{NOB}$ , 13-  $k_{HB}$ , 14-  $k_{STO}$ , 15-  $\mu_{HB}$ , 16-  $\mu_{AOB}$ , 17-  $\mu_{NOB}$ , 18-  $b_{H,O_2}$ , 19-  $b_{STO,O_2}$ , 20-  $b_{AOB}$ , 21-  $b_{NOB}$ , 22-  $\eta_{H,NO_3}$ , 23-  $\eta_{H,NO_2}$ , 24-  $\eta_{H,end,NO_3}$ , 25-  $\eta_{H,end,NO_2}$ , 26-  $\eta_{N,end}$ , 27-  $K_X$ , 28-  $K_{H,O_2}$ , 29-  $K_{H,O_2,inh}$ , 30-  $K_{H,SS}$ , 31-  $K_{H,NH_4}$ , 32-  $K_{H,NO_3}$ , 33-  $K_{H,NO_2}$ , 34-  $K_{H,ALK}$ , 35-  $K_{H,STO}$ , 36-  $K_{AOB,O_2}$ , 37-  $K_{NOB,O_2}$ , 38-  $K_{AOB,NH_4}$ , 39-  $K_{NOB,NO_2}$ , 40-  $K_{N,ALK}$ . The meanings of each parameter can be found in Table S4, SI).

For the SBR system, parameters including  $Y_{H,NO_2}$ ,  $Y_{STO,O_2}$ ,  $Y_{STO,NO_3}$ ,  $Y_{STO,NO_2}$ ,  $Y_{AOB}$ ,  $k_{STO}$ ,  $\mu_{AOB}$ ,  $\mu_{NOB}$ ,  $\eta_{H,NO_2}$  and  $K_{AOB,O_2}$  show high sensitivity, with their normalized relative influences reach 0.01 or even higher (Fig. 4(a)). Meanwhile, according to Fig. 5(a), the highly correlated parameter pairs are  $Y_{H,NO_2}$  vs.  $Y_{STO,NO_2}$  ( $R^2 = 0.98$ ),  $Y_{AOB}$  vs.  $\mu_{NOB}$  ( $R^2 = 0.99$ ),  $Y_{AOB}$  vs.  $K_{AOB,O_2}$  ( $R^2 = 0.99$ ), and  $\mu_{AOB}$  vs.  $K_{AOB,O_2}$  ( $R^2 = 0.99$ ). In addition, the sensitivity of  $Y_{STO,NO_2}$  is much higher than  $Y_{H,NO_2}$  and  $Y_{AOB}$ , and  $K_{AOB,O_2}$  is highly correlated with other parameters, thus,  $Y_{H,NO_2}$ ,  $Y_{AOB}$  and  $K_{AOB,O_2}$  should be eliminated from the parameter subset. Finally, a parameter subset [ $Y_{STO,O_2}$ ,  $Y_{STO,NO_3}$ ,  $Y_{STO,NO_2}$ ,  $k_{STO}$ ,  $\mu_{AOB}$ ,  $\mu_{NOB}$ ,  $\eta_{H,NO_2}$ ] has been selected for the SBR system. Similarly, the parameter subset selected for CSTR system has been determined as [ $Y_{STO,NO_3}$ ,  $Y_{STO,NO_2}$ ,  $\eta_{H,NO_3}$ ,  $\eta_{H,NO_2}$ ,  $K_{H,O_2,inh}$ ,  $K_{H,NO_3}$ ,  $K_{AOB,O_2}$ ,  $K_{NOB,O_2}$ ]. The corresponding calibration and validation achieve satisfied results as shown in Fig. 2(b) and Fig. 3(b).

Optimal parameter subsets of SBR and CSTR systems for the calibration of the extended ASM3 are highly different. In the SBR system, three processes dominate the biological activities, which are storage of  $S_5$  by heterotrophic organisms, growth of autotrophic organisms and denitrification via nitrite as the electron acceptor. These results are consistent with the experimental observations in the target SBR system, in which simultaneous nitrification and denitrification (SND) was observed and the nitrogen loss due to SND under aerobic condition was about 10 ~ 20 mg/L<sup>45</sup>. In contrast, for the CSTR system, the dominated biological activities include storage of  $S_5$  by heterotrophic organisms, denitrification and inhibition of DO on heterotrophic denitrification and autotrophic growth. Hence, different optimal parameter subsets of SBR and CSTR systems may be mainly attributed to the different operation conditions, which lead to different dominant biological activities. From the modeling

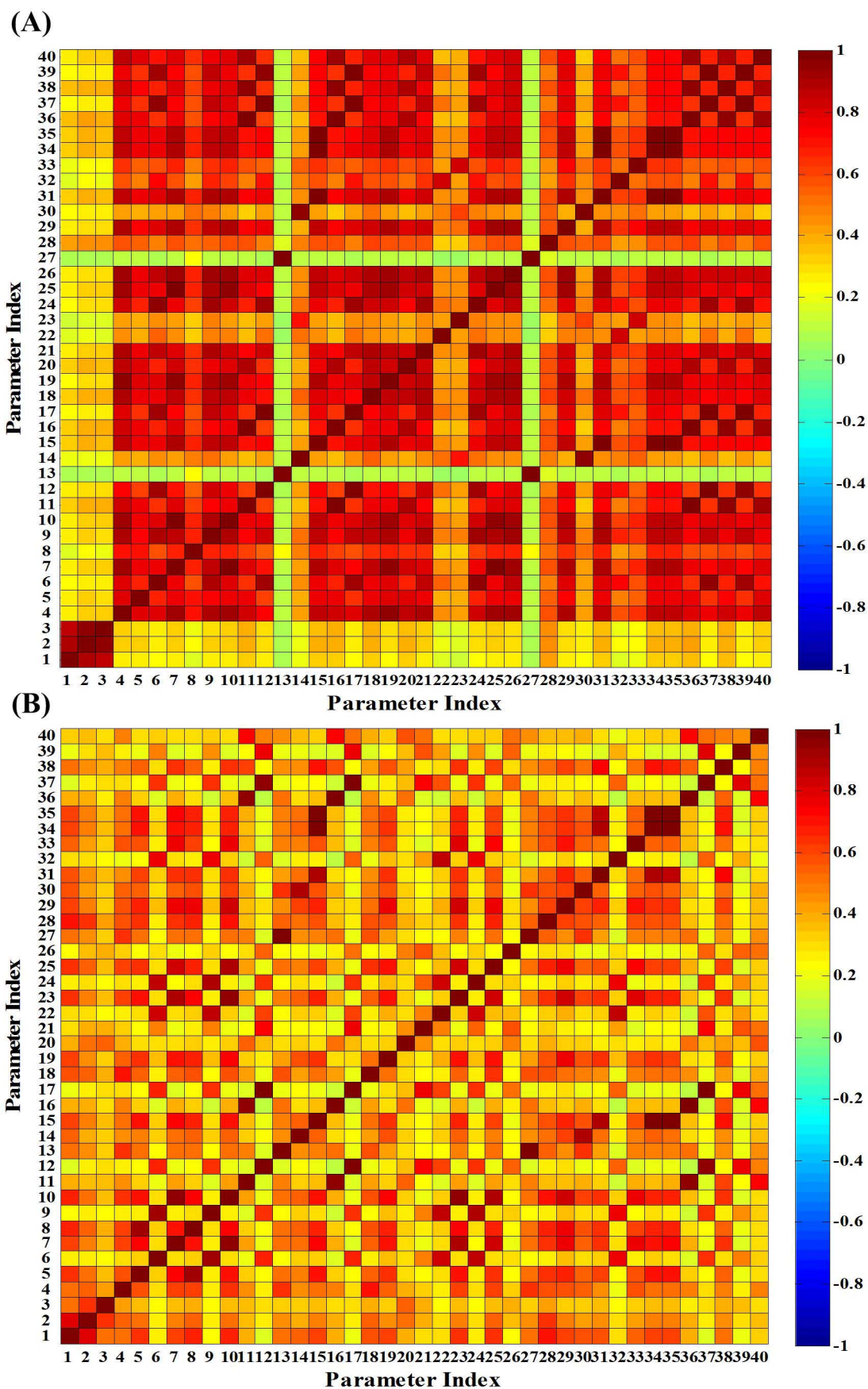


Figure 5 | Global parameter correlation analysis results (color of the off-diagonal elements represents related parameters' correlation, between -1 and 1) (A, Parameter correlation analysis result matrix for the SBR system; B, Parameter correlation analysis result matrix for the CSTR system. The indexes and their corresponding parameters are the same with Fig. 4.)





perspective, different initial values of the state variables, inputs like DO, influent COD and ammonium and the parameter boundaries would lead to different optimal parameter subsets.

**Parameter estimation and model validation based on the NOAP approach.** In order to compare with the results obtained through the experience approach (i.e., preliminary calibration), the parameter subsets for SBR and CSTR systems determined by NOAP are further utilized for calibration and validation in this section. All the parameter estimation results can be found in Table 1.

**Scenario 1: SBR system.** As illustrated in Fig. 2(a), the subset of parameters determined by the NOAP can efficiently calibrate the extended ASM3 through the GO algorithm (see Sim-Pr series). The results indicate that a satisfied calibration could be obtained by different parameter subsets under the condition of limited data availability. Compared to the results of the traditional experience-based approach's calibration results (Fig. 2(a) Sim-Re series), the accumulation of ammonium during anoxic period in the NOAP calibration (Fig. 2(a) Sim-Pr series) is less, demonstrating a better fitting ability of the proposed NOAP approach. Furthermore, the nitrate reduction rate becomes slower when the biodegradable COD concentration is insufficient for denitrification. The simulated nitrate profile deviate slightly from the trend of the experimental nitrate concentrations, which indicates the parameter subset determined by the NOAP is more sensitive than that suggested by the traditional experience-based approach.

Although the calibration is successfully achieved, the initial validation (Fig. 2(b) Sim-Pr series) is still failed ( $R^2(S_{NH} \text{ Sim-Pr}) = 0.0997$ ), similar with the validation by the experience-based parameter subset (Fig. 2(b) Sim-Re series). However, the proposed NOAP provides possible solutions to improve the simulation performance. After the iterative parameter subset selection procedure by NOAP, the parameter  $K_{H,O_2,inh}$  is further added into the parameter subset for re-calibration of the model. As shown in Fig. 2 (c), the performance of re-calibration for both parameter subsets (suggested by the traditional experience-based approach and the NOAP, respectively) are improved after adding the parameter  $K_{H,O_2,inh}$  in the parameter subset. In addition, the parameter subset [ $Y_{STO,O_2}$ ,  $Y_{STO,NO_3}$ ,  $Y_{STO,NO_2}$ ,  $k_{STO}$ ,  $\mu_{AOB}$ ,  $\mu_{NOB}$ ,  $K_{H,O_2,inh}$ ,  $\eta_{H,NO_3}$ ] determined by the NOAP can reach a better fitting goodness, as the simulated ammonium concentrations match better with the original data during the anoxic period (Fig. 2 (c)). Thus, the proposed NOAP demonstrates its ability to provide additional information for improving model calibration.

**Scenario 2: CSTR system.** As present as Fig. 3, the calibration results (see Sim-Pr series) with parameter subset determined by the NOAP procedure achieve a much better fitting goodness with the experimental data (as shown in Fig. 3(a)). The validation of the calibrated parameters

(Fig. 3(b)) also shows a good fitting except the period from day 99 to day 101, in which a lower ammonium and higher nitrite concentrations are predicted. The reason for such phenomena might result from the use of DO as an input for simulation and the DO concentrations during the period are higher than the real concentrations. It can be concluded that parameter subset [ $Y_{STO,NO_3}$ ,  $Y_{STO,NO_2}$ ,  $\eta_{H,NO_3}$ ,  $\eta_{H,NO_2}$ ,  $K_{H,O_2,inh}$ ,  $K_{H,NO_3}$ ,  $K_{AOB,O_2}$ ,  $K_{NOB,O_2}$ ] selected by the proposed NOAP could achieve a better fitting, with the same GO algorithm.

## Discussion

In this work, a novel NOAP approach for the efficient calibration of activated sludge models with limited available data has been proposed. The proposed NOAP integrates a new numerical global parameter sensitivity analysis method (DGSM) for factor fixing and a numerical pseudo-global parameter correlation analysis method for non-identifiable parameter detection to determine the optimal parameter subset for model calibration. The validity and applicability of the approach for efficient model calibration is confirmed by two different activated sludge systems (SBR and CSTR systems). The model calibration results suggested that the optimal sensitive parameter subsets of the SBR and CSTR system are different despite with the same extended ASM3 model to calibrate. Even with the same biomass collected from a municipal WWTP, two SBR reactors finally result in different optimal parameter subsets due to different operational conditions. The results indicate that the parameter subsets determined by NOAP can tail with the state variation of the system. This outstands from the experience-based procedures in calibrating dynamic systems as activated sludge systems whose parameters and structure can vary gradually, which would facilitate modeling automation a lot to support more optimization applications of WWTP.

The optimal parameter subset is different and specific for various systems, because of differences in environmental conditions, influent characteristics, operation modes and biomass population. As a black-box method, conventional experience-based model calibration procedures construct a mapping relation between data and model parameter subset, in which mapping routines are based on experts' empirical knowledge. However, uncertainty would be inevitable due to the arbitrary subset selection. For example, these risks may be distinctly enlarged when modeling the dynamic SBR scenario. An efficient calibration procedure is not only simple to fit the trend of historical data by manually selecting a parameter subset, but should be competent for optimal parameter subset determination, with the aids of efficient parameter estimation algorithms. The proposed NOAP in this study could be a promising alternative to fulfill the described demands. Since the global sensitivity analysis possesses the ability to evaluate uncertainty impact of the concerning factors on the model outputs scientifically. In fact, through factors fixing, parameter

**Table 1 | Parameter estimation results of each model calibrations after global optimization using Genetic Algorithm**

Subset suggested by reference			Subset selected by NOAP					
Parameter	Value		Parameter	Value	Parameter	Value	CSTR scenario	
	SBR scenario	CSTR scenario					First calibration of SBR scenario	Re-calibration of SBR for improvement
$\mu_{AOB}$	0.78 d <sup>-1</sup>	0.38 d <sup>-1</sup>	$Y_{STO,O_2}$	0.53 g COD/g COD	$Y_{STO,O_2}$	0.53 g COD/g COD	$Y_{STO,NO_3}$	0.50 g COD/g COD
$\mu_{NOB}$	0.73 d <sup>-1</sup>	0.34 d <sup>-1</sup>	$Y_{STO,NO_3}$	0.44 g COD/g COD	$Y_{STO,NO_3}$	0.44 g COD/g COD	$Y_{STO,NO_2}$	0.36 g COD/g COD
$\eta_{H,NO_3}$	0.11	0.08	$Y_{STO,NO_2}$	0.24 g COD/g COD	$Y_{STO,NO_2}$	0.24 g COD/g COD	$\eta_{H,NO_3}$	0.42
$\eta_{H,NO_2}$	0.99	0.94	$k_{STO}$	4.37 d <sup>-1</sup>	$k_{STO}$	4.37 d <sup>-1</sup>	$\eta_{H,NO_2}$	0.52
$\eta_{H,end,NO_3}$	0.86	0.05	$\mu_{AOB}$	1.02 d <sup>-1</sup>	$\mu_{AOB}$	1.02 d <sup>-1</sup>	$K_{H,O_2,inh}$	0.73 g O <sub>2</sub> m <sup>-3</sup>
$\eta_{H,end,NO_2}$	0.95	0.98	$\mu_{NOB}$	0.76 d <sup>-1</sup>	$\mu_{NOB}$	0.76 d <sup>-1</sup>	$K_{H,NO_3}$	5.69 g N m <sup>-3</sup>
$(K_{H,O_2,inh})$	(18.44 g O <sub>2</sub> m <sup>-3</sup> )	-	$\eta_{H,NO_3}$	0.78	$K_{H,O_2,inh}$	2.85 g O <sub>2</sub> m <sup>-3</sup>	$K_{AOB,O_2}$	3.11 g O <sub>2</sub> m <sup>-3</sup>
-	-	-	-	-	$\eta_{H,NO_2}$	0.78	$K_{NOB,O_2}$	2.47 g O <sub>2</sub> m <sup>-3</sup>



estimation is performed for top uncertainty introducers, which would reduce outputs' uncertainty in maximization. Consequently, through parameter estimation of subset determined by the global sensitivity analysis, the possibility of successful model calibration and prediction can be maximized. Furthermore, parameter subset optimized by numerical global parameter correlation analysis would enhance the success of calibration and validation. Simultaneously, the proposed NOAP method can quickly capture the shifts of system states through continuous updating of the known Factors' values and unknown Factors' boundary variations. Thus, the proposed NOAP is a promising and useful tool for the efficient calibration of ASMs and could potentially apply to other ordinary differential equations models.

Currently, the proposed NOAP procedure is a decision-helping tool, rather than an automatic protocol. In fact, it is expected to develop a fully automatic calibration procedure in future. Firstly, the automatic calibration procedure should recognize the optimal parameter subset for any models, and organize efficient parameter estimation automatically and robustly, through efficient numerical or symbolic algebra calculation approaches. Secondly, automatic methods are available for optimal experiments design and data collection, when uncertainty analysis of the parameter estimated is necessary in case of improving confidence intervals. Moreover, this calibration procedure can provide implementation procedure for automatic modification, selection and extension of model structure, when both model calibration and validation failed after data optimization. Most importantly, the calibrated model cannot only monitor and predict the overall process dynamics, but also facilitate the operators to achieve optimal control of a target system.

## Methods

**Parameter estimation, parameter sensitivity and correlation analysis.** The parameter estimation is a critical step of the model calibration process. Stochastic global optimization algorithms can find the global minimum of the objective function given by equation (1)<sup>48,49</sup>, where  $\mathbf{y}_0 = [COD(0) S_{NH4}(0) S_{NO2}(0) S_{NO3}(0)]^T$  as the example. Fitting goodness criterion function for each observable is given by equation (2). Characterizations and explanations of the symbols presented in this section can be found in the Table S1 (Supporting Information, SI).

$$E(\theta) = \sum_{k=1}^{N_t} [\mathbf{y}^s(t_k) - \mathbf{y}^o(t_k)]^T \omega^{-1}(t_k) [\mathbf{y}^s(t_k) - \mathbf{y}^o(t_k)], \quad (1)$$

where

$$\frac{d\mathbf{y}}{dt} = f(\mathbf{y}, \mathbf{x}, \mathbf{u}, \theta),$$

$$\mathbf{y}(0) = \mathbf{y}_0.$$

$$A_i(B_i) = 1 - \frac{\sum_{k=1}^{N_t} w^{-1}(\mathbf{y}_i^s(t_k) - \mathbf{y}_i^o(t_k))^2}{\sum_{k=1}^{N_t} w^{-1}(\bar{\mathbf{y}}_i^o(t_k) - \mathbf{y}_i^o(t_k))^2} \quad (2)$$

GA possesses the advantages of easy implementation and mature codes to reuse compared to other resembled technologies<sup>50-52</sup>. In this study, the MATLAB R2010a (Global Optimization Toolbox) is referred as the numerical function implementation of GA (The Mathworks Inc. USA).

To realize factor fixing, a Derivative based Global Sensitivity Measures (DGSM) method is introduced to perform the global sensitivity analysis<sup>53,54</sup>. By comparisons among DGSM, Morris and Sobol's method, it indicates that: a) DGSM shows much higher convergence rate and more accurate than Morris method for non-monotonic functions; b) there is a link between DGSM and Sobol' global sensitivity indices, but the computational time required for numerical evaluation of DGSM measures is many orders of magnitude lower<sup>55,56</sup>. Essentially, the DGSM method is based on the local sensitivity measure, but perform an average of the local sensitivity measure throughout the parameter space by introducing Quasi Monte Carlo sampling methods. The relative time varying sensitivity matrix is described as the following equation (3).

$$\bar{S}_{ij} = \frac{\theta_j}{\mathbf{y}_i} \left( \frac{\partial \mathbf{y}_i}{\partial \theta_j} \right)_{\mathbf{y} = \mathbf{y}(t, \bar{\theta}), \theta = \bar{\theta}} \quad (3)$$

Average  $\bar{S}_{ij}$  over the parameter space using Quasi Monte Carlo sampling methods, a measure can be defined as the equation (4).

$$\bar{M}_{ij} = \int_{H^{N_\theta}} \bar{S}_{ij} d\theta \quad (4)$$

The numerical computation format can be expressed as the equation (5).

$$\bar{M}_{ij}^* = \frac{1}{N_\theta} \sum_{n=1}^{N_\theta} |\bar{S}_{ij}| d\theta \quad (5)$$

To overcome the time varying character, the global sensitivity analysis indices of each parameter are defined as equation (6).

$$\bar{M}_j^* = \frac{1}{N_y} \frac{1}{N_t} \sum_{i=1}^{N_y} \sum_{k=1}^{N_t} \bar{M}_{ij}^*(t_k) \quad (6)$$

About the global Correlation Analysis, a pseudo-global correlation matrix is introduced<sup>54</sup>. The local Fisher Information Matrix (FIM) is described as equation (7).

$$FIM = \sum_{k=1}^{N_t} \left[ \frac{\partial \mathbf{y}^s(t_k)}{\partial \theta} \right]^T \omega_i^{-1}(t_k) \left[ \frac{\partial \mathbf{y}^s(t_k)}{\partial \theta} \right] \quad (7)$$

The derivative covariance matrix is an approximation of the inverse of the FIM as equation (8).

$$C = FIM^{-1} = \left[ \sum_{k=1}^{N_t} \left[ \frac{\partial \mathbf{y}^s(t_k)}{\partial \theta} \right]^T \omega_i^{-1}(t_k) \left[ \frac{\partial \mathbf{y}^s(t_k)}{\partial \theta} \right] \right]^{-1} \quad (8)$$

To introduce the pseudo-global covariance matrix, the local covariance matrix needs to be averaged throughout the parameter space like DGSM done with each objective function's value as the weight as equation (9).

$$\bar{C}_{ij} = \frac{1}{\sum_{n=1}^{N_\theta} E(\theta_n)} \sum_{n=1}^{N_\theta} C_{ij}(\theta_n) E(\theta_n) \quad (9)$$

According to the pseudo-global covariance matrix, the correlation matrix is defined as equation (10).

$$\begin{cases} R_{ij} = \frac{\bar{C}_{ij}}{\sqrt{\bar{C}_{ii}\bar{C}_{jj}}}, i \neq j; \\ R_{ij} = 1, i = j. \end{cases} \quad (10)$$

Based on the equations (6) and (10), parameter sensitivity ranking order and correlation relationships would be produced systematically.

**Activated sludge model and experimental data for NOAP testing.** An extended ASM3 for two-step nitrification and denitrification<sup>6</sup> is used for verifying the proposed procedure. The model inherits the basic mechanism settings of ASM3, in the frame of "Hydrolysis - Storage - Growth - Respiration", nitrification and denitrification are extended to meet current need of description for main intermediate product - nitrite. The kinetic equations and stoichiometric matrix are presented in Table S2 and Table S3 (SI), respectively. Model structure and parameter settings are kept as the original for possibility of results comparison.

In addition, basic stoichiometric and kinetic parameters related information is presented in the Table S4 (SI), as well as parameter boundaries for GSA and parameter estimation. The validity and applicability of the approach is confirmed using experimental data obtained with two independent wastewater treatment systems, including SBR<sup>45</sup> and CSTR<sup>46</sup>, respectively.

Experimental data of the SBR related scenario were collected from two reactors with a working volume of 14 L. Both reactors were seeded with the same inoculum from a full-scale municipal WWTP, but operated in different modes<sup>45</sup>. One was operated with the complete nitrification mode, while the other one was operated with the partial nitrification mode. The complete nitrification mode was operated in the aerobic-anoxic scheme with extensive aeration. Each cycle of the aerobic-anoxic scheme consisted of 3 min feeding, aeration, anoxic phase, 1 h settling, 6 min decanting, and 1 min idling. Aeration was still provided for another 0.5 h after the ammonium has been completely oxidized to nitrite, which would offer ideal environment for the nitrite-oxidizing bacteria to oxidize the nitrite successively to nitrate completely. The data from the complete nitrification reactor are used for the preliminary model calibration because the kinetic properties of the microorganisms in the system can be properly captured by these data series. In addition, the reactor with partial nitrification mode was also operated in the aerobic-anoxic scheme, but aeration duration was controlled through a real-time control system. The data from the partial nitrification reactor with obvious nitrite accumulation are applied to validate the preliminary model calibration results<sup>45</sup>.

The lab-scale CSTR was set up to achieve partial nitrification. The reactor had an effective reaction volume of 4 L, followed by an clarifier with a working volume of 3.5 L. Sludge retention time (SRT) was kept at 12 days by wasting sludge from the secondary clarifier. Experimental data of the CSTR related scenario illustrated an obvious effect of DO on the nitrite accumulation in the CSTR system, which could be used to identify and estimate DO related switching function parameters. DO play an important role in biological nitrogen removal processes. Controlling DO at a proper level can not only reduce energy consumption, but also favor the partial nitrification





for nitrogen removal via nitrite<sup>57</sup>. During the simulation in this case, experimental data from 113-day operation of the CSTR are divided into two groups. One group (0<sup>th</sup> to 75<sup>th</sup> day) is used for the preliminary calibration, while the other group (75<sup>th</sup> to 113<sup>th</sup> day) is used for the validation.

- Hauduc, H. *et al.* Critical review of activated sludge modeling: State of process knowledge, modeling concepts, and limitations. *Biotechnol and Bioeng* **110**, 24–46 (2013).
- Hauduc, H. *et al.* Activated sludge modelling in practice: an international survey. *Water Sci Technol* **60**, 1943–1951 (2009).
- Gernaey, K. V., van Loosdrecht, M. C. M., Henze, M., Lind, M. & Jorgensen, S. B. Activated sludge wastewater treatment plant modelling and simulation: state of the art. *Environ Modell Softw* **19**, 763–783 (2004).
- Rieger, L., Koch, G., Kuhn, M., Gujer, W. & Siegrist, H. The Eawag Bio-p Module for Activated Sludge Model NO. 3. *Water Res* **35**, 3887–3903 (2001).
- Iacopozzi, I., Innocenti, V., Marsili-Libelli, S. & Giusti, E. A modified Activated Sludge Model No.3 (ASM3) with two-step nitrification-denitrification. *Environ Modell Softw* **22**, 847–861 (2007).
- Kaelin, D., Manser, R., Rieger, L., Eugster, J., Rottermann, K. & Siegrist, H. Extension of ASM3 for two-step nitrification and denitrification and its calibration and validation with batch tests and pilot scale data. *Water Res* **43**, 1680–1692 (2009).
- Garcia-Usach, F., Ribes, J., Ferrer, J. & Seco, A. Calibration of denitrifying activity of polyphosphate accumulating organisms in an extended ASM2d model. *Water Res* **44**, 5284–5297 (2010).
- Ni, B. J., Ruscalleda, M., Pellicer-Nacher, C. & Smets, B. F. Modeling nitrous oxide production during biological nitrogen removal via nitrification and denitrification: Extensions to the general ASM models. *Environ Sci Technol* **45**, 7768–7776 (2011).
- Nopens, I. *et al.* Model-based optimisation of the biological performance of a sidestream MBR. *Water Sci Technol* **56**, 135–143 (2007).
- Houweling, D., Wunderlin, P., Dold, P., Bye, C., Joss, A. & Siegrist, H. N<sub>2</sub>O Emissions: Modeling the Effect of Process Configuration and Diurnal Loading Patterns. *Water Environ Res.* **83**, 2131–2139 (2011).
- Koch, G., Kuhn, M., Gujer, W. & Siegrist, H. Calibration and Validation of Activated Sludge Model no. 3 for Swiss Municipal Wastewater. *Water Res* **34**, 3580–3590 (2000).
- Penya-Roja, J. M., Seco, A., Ferrer, J. & Serralla, J. Calibration and validation of Activated Sludge Model No.2d for Spanish municipal wastewater. *Environ Technol* **23**, 849–862 (2002).
- Ludwig, T. *et al.* An advanced simulation model for membrane bioreactors: development, calibration and validation. *Water Sci Technol* **66**, 1384–1391 (2012).
- Wett, B. *et al.* Models for nitrification process design: one or two AOB populations? *Water Sci Technol* **64**, 568–578 (2011).
- Dochain, D. & Vanrolleghem, P. A. *Dynamical Modelling and Estimation in Wastewater Treatment Processes*. IWA Publishing (2001).
- Petersen, B., Gernaey, K., Devisscher, M., Dochain, D. & Vanrolleghem, P. A. A simplified method to assess structurally identifiable parameters in Monod-based activated sludge models. *Water Res* **37**, 2893–2904 (2003).
- Zhang, T., Zhang, D. J., Li, Z. L. & Cai, Q. Evaluating the structural identifiability of the parameters of the EBPR sub-model in ASM2d by the differential algebra method. *Water Res* **44**, 2815–2822 (2010).
- Checchi, N., Giusti, E. & Marsili-Libelli, S. PEAS: A toolbox to assess the accuracy of estimated parameters in environmental models. *Environ Modell Softw* **22**, 899–913 (2007).
- Nopens, I., Hopkins, L. N. & Vanrolleghem, P. A. An overview of the posters presented at Watermatex 2000. III. Model selection and calibration/optimal experimental design. *Water Sci Technol* **43**, 387–389 (2001).
- Sharifi, S., Murthy, S., Takacs, I. & Massoudieh, A. Probabilistic parameter estimation of activated sludge processes using Markov Chain Monte Carlo. *Water Res* **50**, 254–266 (2014).
- Brun, R., Kuhn, M., Siegrist, H., Gujer, W. & Reichert, P. Practical identifiability of ASM2d parameters - systematic selection and tuning of parameter subsets. *Water Res* **36**, 4113–4127 (2002).
- Ruano, M. V., Ribes, J., De Pauw, D. J. W. & Sin, G. Parameter subset selection for the dynamic calibration of activated sludge models (ASMs): experience versus systems analysis. *Water Sci Technol* **56**, 107–115 (2007).
- Kim, Y. S., Kim, M. H. & Yoo, C. K. A new statistical framework for parameter subset selection and optimal parameter estimation in the activated sludge model. *J Hazard Mater.* **183**, 441–447 (2010).
- Mannina, G., Cosenza, A., Vanrolleghem, P. A. & Viviani, G. A practical protocol for calibration of nutrient removal wastewater treatment models. *J Hydroinform* **13**, 575–595 (2011).
- Sin, G. *et al.* Modelling nitrite in wastewater treatment systems: a discussion of different modelling concepts. *Water Sci Technol* **58**, 1155–1171 (2008).
- Sin, G., Vanhulle, S., Depauw, D., Vangriensven, A. & Vanrolleghem, P. A critical comparison of systematic calibration protocols for activated sludge models: A SWOT analysis. *Water Res* **39**, 2459–2474 (2005).
- Gillot, S., Ohtsuki, T., Rieger, L., Shaw, A., Takacs, I. & Winkler, S. Development of a unified protocol for good modeling practice in activated sludge modeling. *Influents* **4**, 70–72 (2009).
- Rieger, L. *et al.* Guidelines for Using Activated Sludge Models. 25–26 (IWA Publishing, London, 2013).
- Machado, V. C., Tapia, G., Gabriel, D., Lafuente, J. & Baeza, J. A. Systematic identifiability study based on the Fisher Information Matrix for reducing the number of parameters calibration of an activated sludge model. *Environ Modell Softw* **24**, 1274–1284 (2009).
- Makinia, J. & Wells, S. A. A general model of the activated sludge reactor with dispersive flow - II. Model verification and application. *Water Res* **34**, 3997–4006 (2000).
- Makinia, J., Rosenwinkel, K. H. & Spering, V. Long-term simulation of the activated sludge process at the Hanover-Gummerwald pilot WWTP. *Water Res* **39**, 1489–1502 (2005).
- Gujer, W. *Systems Analysis for Water Technology*. [8–10] (Springer, Verlag Berlin Heidelberg, 2008).
- Saltelli, A. *et al.* *Global Sensitivity Analysis. The Primer*. [10–39] (John Wiley & Sons, West Sussex, 2008).
- Neumann, M. B. Comparison of sensitivity analysis methods for pollutant degradation modelling: A case study from drinking water treatment. *Sci Total Environ* **433**, 530–537 (2012).
- Cosenza, A., Mannina, G., Vanrolleghem, P. A. & Neumann, M. B. Global sensitivity analysis in wastewater applications: A comprehensive comparison of different methods. *Environ Modell Softw* **49**, 40–52 (2013).
- Morris, M. D. Factorial sampling plans for preliminary computational experiments. *Technometrics* **33**, 161–174 (1991).
- Cukier, R. I., Fortuin, C. M., Shuler, K. E., Petschek, A. G. & Schaibly, J. H. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I. Theory. *J Chem Phys* **59**, 3873–3878 (1973).
- Schaibly, J. H. & Shuler, K. E. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. II. Applications. *J Chem Phys* **59**, 3879–3888 (1973).
- Cukier, R. I., Schaibly, J. H. & Shuler, K. E. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients.3. Analysis of the approximations. *J Chem Phys* **63**, 1140–1149 (1975).
- Saltelli, A., Tarantola, S. & Chan, K. P. S. A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics* **41**, 39–56 (1999).
- Sobol, I. M. Sensitivity estimates for nonlinear mathematical models. *Math Modelling Comput Experiment* **1**, 407–414 (1993).
- Chis, O., Banga, J. R. & Balsa-Canto, E. GenSSI: a software toolbox for structural identifiability analysis of biological models. *Bioinformatics* **27**, 2610–2611 (2011).
- Vajda, S., Godfrey, K. & Rabitz, H. Similarity transformation approach to identifiability analysis of nonlinear compartmental models. *Math Biosci* **93**, 217–248 (1989).
- Chis, O. T., Banga, J. R. & Balsa-Canto, E. Structural Identifiability of Systems Biology Models: A Critical Comparison of Methods. *PLoS ONE* **6**, e27755 (2011).
- Yang, Q., Liu, X. H., Peng, C. Y., Wang, S. Y., Sun, H. W. & Peng, Y. Z. N<sub>2</sub>O production during nitrogen removal via nitrite from domestic wastewater- main sources and control method. *Environ Sci Technol* **43**, 9400–9406 (2009).
- Peng, Y. Z., Guo, J. H., Horn, H., Yang, X. & Wang, S. Y. Achieving nitrite accumulation in a continuous system treating low-strength domestic wastewater: switchover from batch start-up to continuous operation with process control. *Appl Microbiol Biot* **94**, 517–526 (2012).
- Corominas, L. *et al.* New framework for standardized notation in wastewater treatment modelling. *Water Sci Technol* **61**, 841–857 (2010).
- Moles, C. G., Mendes, P. & Banga, J. R. Parameter estimation in biochemical pathways: A comparison of global optimization methods. *Genome Res* **13**, 2467–2474 (2003).
- Chou, I. C. & Voit, E. O. Recent developments in parameter estimation and structure identification of biochemical and genomic systems. *Math Biosci* **219**, 57–83 (2009).
- Kim, S., Lee, H., Kim, J., Kim, C., Ko, J. & Woo, H. Genetic algorithms for the application of Activated Sludge Model No.1. *Water Sci Technol* **45**, 405–411 (2002).
- Fang, F., Ni, B. J. & Yu, H. Q. Estimating the kinetic parameters of activated sludge storage using weighted non-linear least-squares and accelerating genetic algorithm. *Water Res* **43**, 2595–2604 (2009).
- Keskitalo, J. & Leiviska, K. Application of evolutionary optimisers in data-based calibration of Activated Sludge Models. *Expert Syst Appl* **39**, 6609–6617 (2012).
- Kucherenko, S., Rodriguez-Fernandez, M., Pantelides, C. N. S. Monte carlo evaluation of derivative based global sensitivity measures. *Reliab Eng Syst Safe* **94**, 1135–1148 (2009).
- Rodriguez-Fernandez, M. & Banga, J. R. SensSB: A software toolbox for the development and sensitivity analysis of systems biology models. *Bioinformatics* **26**, 1675–1676 (2010).
- Kucherenko, S., Rodriguez-Fernandez, M., Pantelides, C. & Shah, N. Monte Carlo evaluation of derivative-based global sensitivity measures. *Reliab Eng Syst Safe* **94**, 1135–1148 (2009).



56. Sobol, I. M. & Kucherenko, S. Derivative based global sensitivity measures and their link with global sensitivity indices. *Math Comput Simulat* **79**, 3009–3017 (2009).
57. Guo, J. H., Peng, Y. Z., Wang, S. Y., Zheng, Y. A., Huang, H. J. & Wang, Z. W. Long-term effect of dissolved oxygen on partial nitrification performance and microbial community structure. *Bioresour Technol* **100**, 2796–2802 (2009).

## Acknowledgments

This study was financially supported by Natural Science Foundation of China (51208009 and 21177005) and Natural Science Foundation of Beijing (8132008). We also acknowledge the support from Specialized Research Fund for the Doctoral Program of Higher Education (20121103120010) and National High Technology Research and Development Program (863 Program) of China (2011AA060903-02). J. Guo acknowledges the support of the Australian Research Council Discovery Early Career Researcher Award (DE13010140). B.-J. Ni acknowledges the support of the Australian Research Council Discovery Project (DP130103147).

## Author contributions

J.G. and Y.P. conceived and designed the experiments; A.Z. and J.G. performed the modelling and analyzed the data; S.W., B.-J.N., Q.Y. contributed materials/analysis tools; A.Z., J.G. and B.-J.N. wrote the paper. All authors reviewed the manuscript.

## Additional information

**Supplementary information** accompanies this paper at <http://www.nature.com/scientificreports>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Zhu, A. *et al.* A Novel Protocol for Model Calibration in Biological Wastewater Treatment. *Sci. Rep.* **5**, 8493; DOI:10.1038/srep08493 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>