Evaluation of the information content of long-term wastewater characteristics data in relation to activated sludge model parameters

Jamal Alikhani, Imre Takacs, Ahmed Al-Omari, Sudhir Murthy and Arash Massoudieh

ABSTRACT

A parameter estimation framework was used to evaluate the ability of observed data from a full-scale nitrification–denitrification bioreactor to reduce the uncertainty associated with the bio-kinetic and stoichiometric parameters of an activated sludge model (ASM). Samples collected over a period of 150 days from the effluent as well as from the reactor tanks were used. A hybrid genetic algorithm and Bayesian inference were used to perform deterministic and parameter estimations, respectively. The main goal was to assess the ability of the data to obtain reliable parameter estimates for a modified version of the ASM. The modified ASM model includes methylotrophic processes which play the main role in methanol-fed denitrification. Sensitivity analysis was also used to explain the ability of the data to provide information about each of the parameters. The results showed that the uncertainty in the estimates of the most sensitive parameters (including growth rate, decay rate, and yield coefficients) decreased with respect to the prior information.

Key words | activated sludge model, Bayesian inference, nitrification/denitrification, parameter estimation, sensitivity analysis, uncertainty analysis

Jamal Alikhani
Arash Massoudieh (corresponding author)
Department of Civil Engineering,
The Catholic University of America,
630 Michigan Ave NE,
Washington,
DC 20064,
USA
E-mail: *massoudieh@cua.edu*

Imre Takacs
Dynamita,
7 Eoupe,
Nyons 26110,
France

Ahmed Al-Omari
Sudhir Murthy
DC Water and Sewer Authority,
5000 Overlook Avenue SW,
Washington,
DC 20032,
USA

INTRODUCTION

The activated sludge process is one of the most widely applied systems for the removal of natural organic matter and nutrients in municipal wastewater treatment plants (WWTPs) (Gernaey *et al.* 2004). The International Water Association's activated sludge model number 1 (ASM1) was one of the first modeling concepts to explicitly account for the interactions of multiple constituents and functional groups of microorganisms (Henze *et al.* 1987). International Water Association task groups later developed more complex versions of the ASM1: ASM2, ASM2d, and ASM3 (Henze *et al.* 2000). Other models that have been developed include the Barker and Dold model (Barker & Dold 1997), the TUDP model (Meijer *et al.* 2001), and the EAWAG Bio-P module for ASM3 (Rieger *et al.* 2001). A comprehensive review of the different ASM models can be found in Gernaey *et al.* (2004).

The ASMs provide a systematic way to represent the interaction between state variables, process rates, and stoichiometric coefficients in bioreactions, making them flexible tools for incorporating various processes in both full-scale and pilot systems (Comas *et al.* 2008; Sun *et al.* 2009; Fang *et al.* 2010; Fall *et al.* 2011; Busch *et al.* 2013). Modified versions of the ASMs have been developed to address different levels of complexity in representing processes, emerging contaminants, particulate matter forms, and functional groups of microorganisms (Ekama & Wentzel 2004; Stare *et al.* 2006; Bolong *et al.* 2009; Sun *et al.* 2009; Hao *et al.* 2011; Yang *et al.* 2013).

One of the challenges faced in using any ASM is the difficulty of calibrating a large number of bio-kinetic and stoichiometric parameters (Smets *et al.* 2003). Modeling can be performed without calibration and just by using parameter values obtained from the literature based on independent batches or pilot investigations (Henze *et al.* 2000; Cox 2004; Hauduc *et al.* 2011; Rahman *et al.* 2016a, 2016b). However, the complexities of biological reaction systems, variations in raw wastewater characteristics, design configurations, and different experimental or operating

doi: 10.2166/wst.2017.004

conditions can make it difficult to find suitable parameter values from the literature for a specific plant. Furthermore, wide ranges of values have been reported for ASM parameters in the literature, which makes choosing a single parameter set representing the condition in a specific reactor difficult. Cox (2004) reviewed a number of articles and created a database containing different ASM parameter values. As an example, 31 different values have been reported for the maximum heterotrophic growth rate in 22 references, ranging between 0.51 and 11 day^{-1}.

In many cases, an estimation of the ASM parameters through calibration using measured data provides more realistic values for existing biological treatment systems or design of a new system compared to determining the parameter values from the literature (Sharifi *et al.* 2014). Manual and automatic calibration methods have been widely used to estimate ASM parameters using measured data (Meijer *et al.* 2001; Fall *et al.* 2011). These methods are typically based on minimizing an objective function defined as the misfit between model predictions and observed data (Vanrolleghem *et al.* 1999; Machado *et al.* 2009; Kim *et al.* 2010; Keskitalo & Leiviskä 2012). Keskitalo & Leiviskä (2012) applied an evolutionary optimizer to minimize the sum of squared errors between modeled and measured data obtained from a municipal WWTP and a pulp mill WWTP. Kim *et al.* (2010) used the response surface method as a surrogate for the mechanistic ASM to estimate model parameters in an effort to reduce computational costs. Although deterministic (maximum likelihood) parameter estimation can lead to more realistic values for a particular plant than the use of default values from the literature, there remain uncertainties associated with the estimated parameters due to measurement error, model structural error, the errors associated with the representativeness of sampling, influent's flowrate and composition, and environmental factors such as temperature (Sharifi *et al.* 2014). The effect of these uncertainties on the deterministically estimated parameters needs to be quantified to assess the reliability of ASM models (Neumann & Gujer 2008; Hauduc *et al.* 2010; Mannina *et al.* 2012; Sharifi *et al.* 2014). Uncertainty analysis can help in determining the appropriate safety factors to control the capital investment and effluent quality (Bixio *et al.* 2002; Flores-Alsina *et al.* 2008; Busch *et al.* 2013). A common method to conduct uncertainty analysis is the Monte Carlo simulation – by generating random parameter sets within the preconceived range for each parameter – aiming to find the model effluent's confidence intervals. Bixio *et al.* (2002) applied Monte Carlo simulations to avoid large safety factors in the design stage of a conventional WWTP, which resulted in a 21% reduction in sizing and 43% reduction in capital investment equivalent to 1.2 million euros. Flores-Alsina *et al.* (2008) evaluated different control strategies with input uncertainty using Monte Carlo simulations. Sin *et al.* (2011) and Sin *et al.* (2009) performed Monte Carlo simulations by considering uncertainties associated with ASM1's parameters, partitioning coefficients (e.g., chemical oxygen demand (COD) content of particulate constituents), and hydraulic parameters to obtain the confidence intervals for model predictions. Alikhani *et al.* (2015) applied autoregressive moving-average model to generate random realizations of influent flowrates and composition into a biological nutrient removal system aiming at evaluating the influent uncertainty.

Different approaches have been used in different areas of water resources to perform backward uncertainty propagation. These approaches can be classified into those based on multiple linear regression (Hill & Tiedeman 2007; Foglia *et al.* 2009), global methods such as the generalized likelihood uncertainty estimator (Beven & Freer 2001; Mannina *et al.* 2012; Sathyamoorthy *et al.* 2014), and Bayesian inference (Dotto *et al.* 2011; Albrecht 2013; Sharifi *et al.* 2014; Alikhani *et al.* 2016a). Albrecht (2013) compared four different parameter estimation techniques for a collection of reaction models with artificially added noise to experimental data and concluded that for highly nonlinear models, the Markov Chain Monte Carlo (MCMC) algorithm based on Bayesian inference has better accuracy than local sensitivity-based approaches. However, the computational cost of MCMC simulations is significantly higher than local sensitivity-based methods, as they require a relatively larger number of model runs. For a discussion of the number of runs used in our work please see 'Parameter estimation using MCMC' section. The Bayesian approach provides a joint probability distribution representing the degree of confidence regarding parameters after the application of observed data (referred to as the posterior distribution). The posterior range is the parameter's degree of credibility dictated by the observed measured points through the Bayesian inference (Kaplan 1997). In the Bayesian approach, prior knowledge about parameter values, based on the literature or independent laboratory experiments, can be incorporated into estimations via prior distributions (Sharifi *et al.* 2014). The magnitude of reduction in the spread of the posterior distribution of parameters relative to their prior distributions can be used as an indicator for the amount of information gained regarding each parameter. The posterior joint probability distribution of

parameters obtained from the Bayesian inference can be applied in a Monte Carlo stochastic simulation technique for stochastic simulation of the biological treatment systems (Sin *et al.* 2011, 2009; Mannina *et al.* 2012).

Only a small number of studies have focused on the parameter estimation of the nitrification and denitrification (NDN) processes using real data from full-scale bioreactors (Andreottola *et al.* 1997; Koch *et al.* 2001; Choubert *et al.* 2009; Kaelin *et al.* 2009; Mannina *et al.* 2012). The main contribution of the presented study is to systematically quantify the information content of long-term field data from a methanol-feed NDN system in the context of the estimation of the modified ASM's bio-kinetics and stoichiometric parameters. The objectives of this study can be summarized as follows: (1) finding the level of confidence regarding parameter values and model prediction by applying Bayesian inference over the long-term wastewater characteristic data, (2) explaining the ability of the observed data to inform about each parameter through sensitivity analysis and evaluating parameter correlation, and (3) evaluating the effect of sampling intervals on the information content of the observed data with respect to model parameters.

MATERIALS AND METHODS

Modeling biological treatment processes

The plant model consists of a series of continuously stirred tank reactors (CSTRs) connected to a single clarifier (Figure 1). Transient mass balance equations for all the state variables, consisting of the concentration of each of the model constituents in each CSTR, can be expressed as a system of ordinary differential equations (ODEs):

$$\frac{d(V_k C_{i,k})}{dt} = \omega_{r,k} Q_r C_{i,r} + \omega_k Q_{in} C_{i,in} + \sum_{k' \neq k}^{m} H(Q_{k',k}) Q_{k',k} C_{i,k'} + \left(\sum_{k' \neq k}^{m} H(-Q_{k',k}) Q_{k',k} \right) C_{i,k} + V_k \sum_{l=1}^{nr} \Phi_{l,i} R_l + F_{i,k} V_k (C_i^* - C_{i,k}) + \dot{m}_{i,k} \quad (1)$$

where V, C, Q, and ω indicate the volume, concentration, flow, and flow's fraction factor (determining how much of the inflow and return flow enters each stage in a step-feed system), respectively, H is the Heaviside (unit step) function, R is the reaction rate, Φ is the stoichiometric coefficient, F is the mass transfer rate, C^* is the saturated concentration, and $\dot{m}$ is the external mass flowrate. The indices are: i indicates a constituent, k represents a tank, r indicates a return flow, 'in' shows conditions in influent, and l indicates a process (or reaction). Moreover, m is the total number of tanks, and nr is the total number of reactions. $Q_{k',k}$ is the flowrate from tank k' to tank k either due to sequential stages or through feed-back or bypass (a positive $Q_{k',k}$ value means flow into stage k). In Equation (1), the term on the left side is the rate of change in the total mass of the constituents i in tank k; the first term on the right-hand side of Equation (1) is the mass inflow due to return flow; the second term is the mass inflow of the influent; the third and fourth terms represent inner-connection flow between tanks k and k'; the fifth term is the production or disappearance of constituents due to reactions; the sixth term is the effect of rate-limited mass transfer (e.g., aeration); and the last term is the direct addition of the constituents (e.g., the addition of a carbon source for denitrification). To obtain solid particle

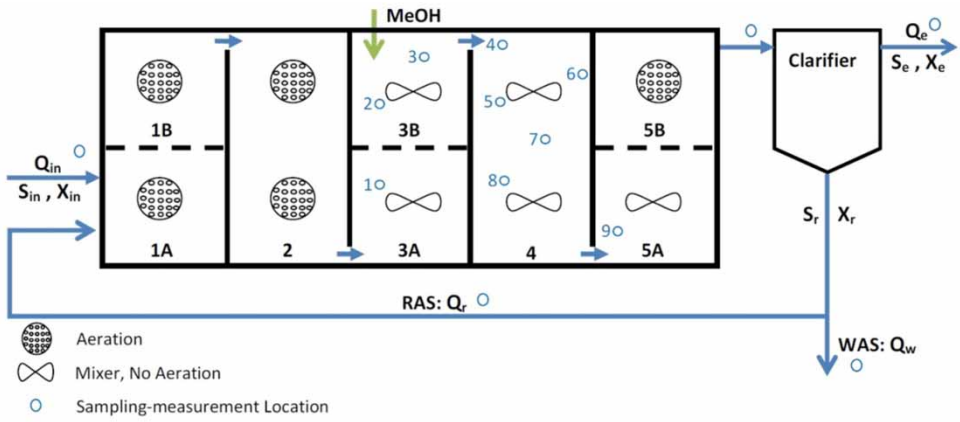

Figure 1 | Configuration of the NDN system with the location of observed points.

concentrations in the return flow ($C_{i,r}$), a dynamic clarifier model (Takács *et al.* 1991) or a quasi steady-state approximation (i.e., performing mass balance while ignoring the solid storage changes in the clarifier) can be applied.

Bayesian parameter estimation

Equation (1) can be expressed in vector form as (Sharifi *et al.* 2014):

$$\frac{d\mathbf{C}}{dt} = f(\mathbf{C}(t),\ \mathbf{U}(t),\ \boldsymbol{\Theta}) \tag{2}$$

where $\mathbf{C}(t)$ is the state variables vector containing the concentrations of all the constituents in the model. The length of the $\mathbf{C}$ vector is equal to the number of constituents times the number of CSTR tanks. $\mathbf{U}(t)$ is the external forcing input representing the influent flowrate Q_{in}, influent concentrations $C_{i,in}$, return activated sludge (RAS) flowrate Q_r, waste activated sludge (WAS) flowrate Q_w, external chemical loading rate $\dot{m}_i$, and the temperature. $\boldsymbol{\Theta}$ is the vector of the model's parameters. Function f in Equation (2) represents the bioreactor model structure, which is represented in Equation (1).

Estimation of the parameters can be performed by comparing modeled and observed concentrations. To obtain the proper model concentration consistent with the measured concentrations (e.g., COD, volatile suspended solids (VSS), and total Kjeldahl nitrogen (TKN)), the following transformation can be used:

$$\mathbf{Y}(t) = \mathbf{Z}\mathbf{C}(t) \tag{3}$$

$$\tilde{\mathbf{Y}}(t) = g^{-1}\{g(\mathbf{Y}(t)) + \boldsymbol{\varepsilon}\} \tag{4}$$

where $\mathbf{Y}$ is the vector containing the model's observed constituents, which is considered a linear combination of the model's individual concentrations. $\mathbf{Z}$ is the partitioning coefficients matrix, $\tilde{\mathbf{Y}}(t)$ represents the observed data vector, g represents the error structure function (e.g., for log-normal error structure, g is the natural logarithm function, while for Gaussian error structure, g is the identity function), and $\boldsymbol{\varepsilon}$ is a random vector containing measurement, structural, and external forcing errors, which is assumed to collectively follow a multivariate normal distribution.

The Bayes' theorem (Kaipio & Somersalo 2005) can be used to obtain the joint probability distribution of parameters given the observed data, namely the posterior distribution of the parameters as:

$$p(\boldsymbol{\varphi}|\tilde{\mathbf{Y}}) = \frac{p(\tilde{\mathbf{Y}}|\boldsymbol{\varphi})\ p(\boldsymbol{\varphi})}{p(\tilde{\mathbf{Y}})} \tag{5}$$

where $p(\tilde{\mathbf{Y}}|\boldsymbol{\varphi})$ is the likelihood function that is the probability of observing the measured data $\tilde{\mathbf{Y}}$ given a parameter set $\boldsymbol{\varphi}$, $p(\boldsymbol{\varphi})$ represents the prior knowledge about the parameters and error structure that can be obtained from literature reviews, experts' experience, or independently performed experimental results. The denominator $p(\tilde{\mathbf{Y}})$ is a normalizing factor (Lu *et al.* 2014a).

In Equation (5) $\boldsymbol{\varphi}$ contains all the parameters that are intended to be estimated using the observed data set ($\tilde{\mathbf{Y}}$). Vector of $\boldsymbol{\varphi}$ includes the bio-kinetics and stoichiometric parameters $\theta_i \in \boldsymbol{\Theta}$, the elements of the partitioning coefficient $z_i \in \mathbf{Z}$, and the elements of the variance-covariance matrix for the random error term $\boldsymbol{\varepsilon}$ ($\sigma_{ij} \in \boldsymbol{\Gamma}$). Therefore, $\boldsymbol{\varphi} = \{\theta_1, \theta_2, \ldots, \theta_{n_\theta}, z_1, z_2, \ldots, z_{n_z}, \sigma_1, \sigma_2, \ldots, \sigma_{n_\sigma}\}$, where n_θ is the total number of unknown ASM model parameters in Equation (2), n_z is the total number of unknown partitioning coefficients in Equation (3), and n_σ is the total number of elements in the variance-covariance matrix in Equation (7). Applying Bayesian inference enhances our degree of confidence about each unknown parameter any time a new evidence point (observed data) applies. In fact, each measured data point contains a potential level of information that can be determined by applying the Bayesian inference.

If the errors due to the uncertainties associated with the external forcing are considered part of the observation error, then the external forcing can be considered deterministically and thus the model outputs become a function of only the bio-kinetic and stoichiometric parameters. In this case, the error function can be rewritten as $p(\tilde{\mathbf{Y}}|\boldsymbol{\Theta}, \mathbf{Z}, \boldsymbol{\Gamma})$. If it is assumed that errors of consecutive observations of the same observed constituent are independent of each other, then the likelihood function of the observed vector can be computed as a function of the model parameters (Freni *et al.* 2009):

$$P(\tilde{\mathbf{Y}}|\boldsymbol{\varphi}) = \frac{\prod_{i=1}^{n_i}\prod_{j=1}^{n_j} g'(\tilde{y}_{ij})}{(2\pi|\boldsymbol{\Gamma}|)^{\sum_{i=1}^{n_i} n_i/2}} \times \exp\left(-\sum_{i=1}^{n_i}\sum_{j=1}^{n_j}\frac{[g(\tilde{y}_{ij}) - g(y_{ij})]\boldsymbol{\Gamma}^{-1}[g(\tilde{y}_{ij}) - g(y_{ij})]^T}{2}\right) \tag{6}$$

where $\tilde{y}_{ij} \in \tilde{\mathbf{Y}}$ is the observed concentration of constituent i, n_i is the number of different observed constituents, j indicates the time and location of the measurement, n_j is the number of

total samples of observed constituent i. The mapping function g in the likelihood function is the transformation depending on the error distribution, and g' is its derivative. For example, considering g being identity function implies a normally distributed and additive error structure, while assuming g to be logarithm function results in a log-normally distributed and multiplicative error structure.

In theory, each element of the variance-covariance matrix $\mathbf{\Gamma}$ should be estimated using the MCMC approach. However, this makes the total number of error standard deviations ($\sigma_{ij} \in \mathbf{\Gamma}$) to be estimated very large and imposes a large computational burden. Therefore, observation errors of different constituents are typically assumed to be independent, and the correlations between observed concentration errors of different constituents are neglected (Walsh & Whitney 2012). This assumption is justified when the main source of the error is the measurement error. However, when the main error component is the model structural error, there is a possibility of correlation between errors of different constituents. By making this assumption, the variance-covariance matrix becomes diagonal and Equation (6) can be simplified to:

$$P(\tilde{\mathbf{Y}}|\boldsymbol{\varphi}) = \frac{\prod_{i=1}^{n_i}\prod_{j=1}^{n_j} g'(\tilde{y}_{ij})}{\left(2\pi\prod_{i=1}^{n_i} n_i\sigma_i^2\right)^{\sum_{i=1}^{n_i} n_i/2}} \exp\left(-\sum_{i=1}^{n_i}\sum_{j=1}^{n_j}\frac{[g(\tilde{y}_{ij}) - g(y_{ij})]^2}{2\sigma_i^2}\right) \quad (7)$$

where $y_{ij} \in \mathbf{Y}$ is the predicted concentration of constituent i at time/location j, and σ_i is the standard deviation of observed constituent i.

MCMC sampling and numerical solution scheme

Due to a large number of parameters and heterogeneity of the observed data, solving Equation (5) by using the analytical methods is not possible. Therefore, an MCMC approach (Gamerman & Lopes 2006; Meyn & Tweedie 2012) was used to generate random samples according to the posterior distribution. In this study, the Metropolis–Hasting (M–H) algorithm (Metropolis *et al.* 1953) was used to obtain a sequence of random numbers from the posterior probability distribution. In the deterministic approach, the values of parameters that maximize the likelihood function in Equation (7) were obtained by applying the hybrid genetic algorithm (GA).

To evaluate the convergence of the MCMC algorithm, a fake parameter with no influence on the model prediction is defined and its posterior distribution is monitored during the Markov Chain sampling. It is expected that the posterior and prior distribution of the fake parameter be identical after a sufficient number of MCMC sampling. The practice of throwing away the initial portion of Markov Chain samples in order to avoid the effect of potentially mis-represented samples on the construction of the posterior distribution is referred to as burn-in (Meyn & Tweedie 2012). The fake parameter was also used to estimate the number of burn-in samples.

To calculate the likelihood function of Equation (6), for each set of Markov Chain samples, the system of ODEs comprising Equation (2) should be solved. In addition to the non-linearity of these ODEs, a wide range of biochemical reaction rate scales ranging from seconds (e.g., for oxygen transfer rate) to days (e.g., for microbial growth rates) results in a stiff system of ODEs (Lindberg & Carlsson 1996). The stiffness of the system of ODEs and daily fluctuations in external forcing (influent, RAS, WAS, loadings, and temperature changes) requires small computational time-steps, which in turn can result in a long simulation runtime. This is particularly a challenge due to the fact that the MCMC algorithm requires a large number of model simulations to obtain the final solution. To overcome this problem, an adaptive time-step backward differentiation formula (BDF) algorithm was developed to solve the system of ODEs. A detailed explanation of the proposed adaptive BDF algorithm for solving the system of ODEs comprising the ASM is presented in Alikhani *et al.* (2016b) and Alikhani *et al.* (2016c).

The total suspended solids (TSS) and VSS values in the effluent of the NDN system during the time of simulation show average values of 3.8 and 2.6 mg L^{-1}, respectively, indicating the very close to the ideal settling performance of the clarifier. Therefore, a quasi steady-state approximation is assumed to obtain the solid particle concentrations in the return flow ($C_{i,r}$) in Equation (1). In addition, in the biological reaction network used for modeling the reactors, the program developed for this study allows any reaction rate expression R_l and stoichiometric constant $\Phi_{l,i}$ to be specified by the user as a function of concentrations of components and/or model parameters.

Global sensitivity analysis

Local and global sensitivities can be used to rank the model parameters in terms of the significance of their impact on important state variables (Saltelli *et al.* 2008; Kalyanaraman *et al.* 2015). Sensitivity analysis can also be applied to explain parameters' identifiability (Weijers *et al.* 1996; Brun *et al.* 2002; Sin *et al.* 2005). A large sensitivity value of model output with respect to each parameter is a necessary but

not sufficient condition for the identifiability of the parameter due to the fact that the model output can change more significantly as a result of changing that parameter (Freni *et al.* 2009). Another factor that can influence the identifiability of parameters is parameter correlation (Eberly & Casella 2003; Ngo *et al.* 2015). Several approaches for estimating local and global sensitivity have been used in the literature (Saltelli *et al.* 2008; Zhan *et al.* 2013).

In this study, an aggregate local sensitivity for each time-series output is defined as the root mean square of sensitivity over the time-series:

$$s_{ij} = \sqrt{\frac{1}{(t_2 - t_1)} \int_{t_1}^{t_2} \left(\frac{\partial \ln y_j(t, \boldsymbol{\theta})}{\partial \ln \theta_i} \right)^2 dt} \tag{8}$$

where s_{ij} represents the sensitivity of the observed j^{th} constituent in the model with respect to the i^{th} parameter, and t_1 and t_2 represent the simulation period. Since in the Bayesian parameter estimation approach, the value of the parameters is expressed as a probability density, the expected value of sensitivity ($\bar{s}_{ij}$) can be expressed as:

$$\bar{s}_{ij} = \int s_{ij}(\theta_i) p(\theta_i | \tilde{\boldsymbol{Y}}) d\theta_i \tag{9}$$

where $p(\theta_i | \tilde{\boldsymbol{Y}})$ is the posterior distribution of parameter i obtained from the Bayesian inference application. The global sensitivities in this study were calculated by sampling the parameter sets provided by the MCMC analysis and performing local sensitivity analyses. The expected values of the sensitivities can then be estimated algebraically.

Equation (9) results in n_j sensitivity values for each parameter associated with the n_j measured constituents, which makes it difficult to gain an overall insight into the sensitivities associated with each parameter. One way would be to add up the values of sensitivity for each assigned parameter to a single value. However, the ability of data on a particular measured constituent to inform about a certain parameter depends on $\bar{s}_{ij}$ as well as the standard deviations of error σ_i for that constituent, as in effect $1/\sigma_i$ acts as a weighting factor on the observed data. Because the standard deviations of errors are treated as unknown parameters in Equation (7), the Bayesian inference provides a way to objectively assign weights to each group of measured data. For this reason, a scaled sensitivity factor (SSF) is defined as:

$$SSF_i = \sum_j \frac{\bar{s}_{ij}}{\sigma_i} \tag{10}$$

where SSF_i is the sum of the sensitivities of all the observed constituents with respect to the parameter i, weighted by the reciprocal of the observation error standard deviation, σ_i, obtained by applying Equation (7). In fact, SSF is an indicator of the overall sensitivity of the model output with respect to each parameter.

MODEL APPLICATION

Biological nitrogen removal process configuration

The described parameter estimation framework was applied to estimate the bio-kinetic, stoichiometric, and aeration parameters in the NDN phase of the Blue Plains Advanced WWTP in Washington, DC, USA. The system can be described as a post-denitrification with methanol addition as an external carbon source. The NDN reactor consists of 12 parallel reactors, each having an active volume of 17,500 m^3. Each reactor is compartmented into eight tanks in series, including two large and six smaller compartments (Figure 1). Nitrification occurs mainly at tanks 1A, 1B, and 2. Tank 3A is not aerated and serves as the deoxygenation stage. Methanol is added at tank 3B, and denitrification occurs in tanks 3B, 4, and 5A under anoxic conditions. The last tank is aerated again to strip nitrogen gas, which can otherwise hinder the settling of activated sludge in the clarifier. Fine bubble diffusers in the aerated tanks and hyperboloid mixers in the un-aerated tanks are used to create completely mixed conditions.

Observed data

The flowrates of influent, effluent, WAS, and RAS, and wastewater characteristics including TSS, VSS, biochemical oxygen demand (BOD), soluble chemical oxygen demand (sCOD), TKN, ammonium (NH_3), nitrite and nitrate (NO_x), dissolved oxygen, and methanol concentrations data were collected from different sampling locations (Figure 1) during February to June of 2010. Three sets of measured data were used in this study, including the following:

1. External forcing data referring to the flowrate and characteristics of the influent, WAS, RAS, and temperature (Figure 2) that were used as given inputs ($\mathbf{U}(t)$ in Equation (2)) to the model.
2. Daily-average flowrate and characteristics of effluent over 150 days of simulation, used as observed data.

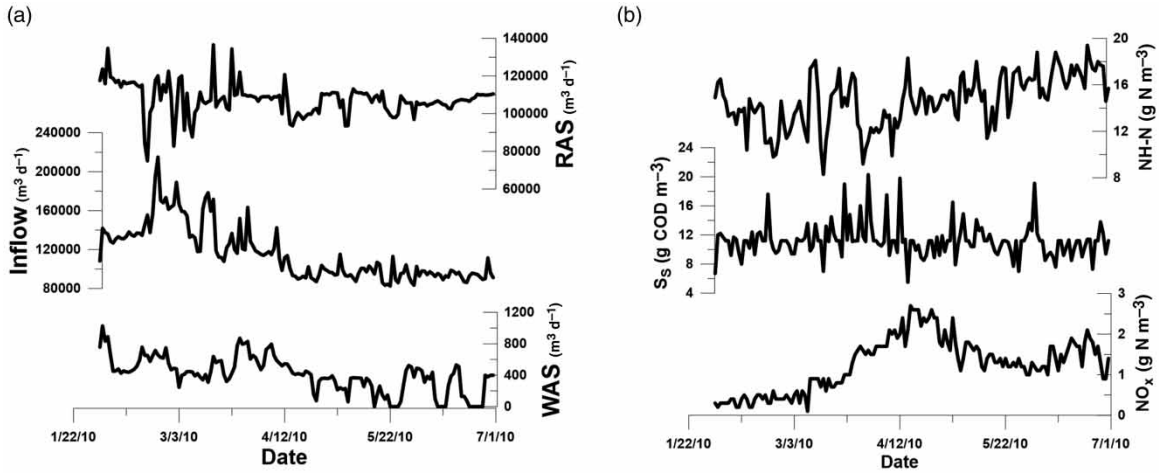

Figure 2 | (a) Daily-average flowrates of influent, RAS, and WAS. (b) Daily influent concentrations of biodegradable substrate (S_S), ammonia nitrogen (NH-N), and nitrate-nitrite (NOx).

3. Spatial profile data referring to the measured concentrations of different constituents in the bioreactors in different sampling locations used as observed data. These observed data were gathered on a weekly to bi-weekly basis as grab samples from points 1 to 9 in tanks 3A through 5A (Figure 1) to take into account the effect of spatial heterogeneity as a source of uncertainty in measured constituents.

Approximately 7,000 measured data points taken from different times and locations were used to calculate the likelihood functions. In the Blue Plains advanced WWTP, the influent to the denitrification stage is the effluent of the BOD removal phase (the phase prior to NDN), and thus contains high levels of ammonia and low sCOD (roughly 15 mg N L^{-1} of ammonia and 25 mg L^{-1} of sCOD). Inflows typically varied between 113,000 and 133,000 $m^3\,d^{-1}$ per reactor depending on the influent wastewater flowrate and the number of reactors in service. A higher inflow fluctuation occurred in the period prior to early April (Figure 2) because the Blue Plains WWTP receives combined domestic wastewater and stormwater runoff and thus faces high flowrate fluctuations, particularly in early spring. Methanol is applied at a rate of 45 liters per 1,000 m^3 of the influent and serves as the carbon source for methylotrophic denitrification. The return flowrate (RAS) ratio to the influent flowrate (Q_r/Q) varied from 0.5 to 1.3.

Reaction network

In the United States, methanol is often used as an external substrate for heterotrophic denitrification because of its availability and cost (Mokhayeri *et al.* 2008). When the methanol concentration is adequate, a new functional group of specialist heterotrophic microorganisms, collectively referred to as methylotrophs, are able to thrive (Lu *et al.* 2014b). Methylotrophs only utilize methanol as an electron donor in anoxic conditions and have different growth, yield, and decay rates than generalist heterotrophic organisms that use natural organic matter (Lee & Welander 1996; Hallin *et al.* 2006; Baytshtok *et al.* 2008; Dold *et al.* 2008; Mustakhimov *et al.* 2013; Alikhani *et al.* in press). Therefore, a modified version of ASM1 (referred to as M-ASM1), to take into account the anoxic growth and decay of methylotrophs plus the aerobic growth of heterotrophs on methanol, has been developed (Alikhani *et al.* 2014). The M-ASM1 reaction table is shown in Table 1 along with the process description rate expressions (R_l) and stoichiometric coefficients ($\Phi_{l,i}$), which are respectively introduced in Equation (1). In this modification, processes 2, 4 and 7 are added to ASM1. The process 2 was included because the unused methanol in the anoxic zone can be utilized by heterotrophs in the aerobic zone. There are also two new constituents, methanol (S_m) and methylotrophic biomass ($X_{B,M}$), added to the original 13 constituents of ASM1. Reaction rates 1 to 5 were modified by considering ammonia as a limiting factor to all the biomass growth rates using Monod kinetics (Hauduc *et al.* 2010). Processes 6 to 7 were modified by implementing the adjustable decay rates under anoxic and aerobic conditions.

Parameter estimation using MCMC

All model parameters, including the 95% prior ranges used to construct the prior distributions and temperature correction factor 'θ_T' are shown in Table 2. The prior

Table 1 | Process description, reaction and rate expressions in M-ASM1 model

J	Process	Reaction	Rate
1	Aerobic growth of heterotrophs	$\frac{1}{Y_H}S_S + \frac{1-Y_H}{Y_H}S_O + i_{XB}S_{NH} \rightarrow X_{B,H}$	$\mu_H \frac{S_S}{K_S+S_S}\frac{S_O}{K_{O,H}+S_O}\frac{S_{NH}}{K_{NH}+S_{NH}}X_{B,H}$
2	Aerobic growth of heterotrophs on methanol	$\frac{1}{Y_{H,M}}S_M + \frac{1-Y_H}{Y_H}S_O + i_{XB}S_{NH} \rightarrow X_{B,\mathrm{H}}$	$\mu_H \frac{S_M}{K_{M,H}+S_M}\frac{S_O}{K_{O,H}+S_O}\frac{S_{NH}}{K_{NH}+S_{NH}}X_{B,H}$
3	Anoxic growth of heterotrophs	$\frac{1}{Y_H}S_S + \frac{1-Y_H}{2.86\,Y_H}S_{NO} + i_{XB}S_{NH} \rightarrow X_{B,H}$	$\mu_H \frac{S_S}{K_S+S_S}\frac{K_{O,H}}{K_{O,H}+S_O}\frac{S_{NO}}{K_{NO,H}+S_{NO}}\frac{S_{NH}}{K_{NH}+S_{NH}}\eta_g X_{B,H}$
4	Anoxic growth of methylotrophs	$\frac{1}{Y_M}S_M + \frac{1-Y_M}{2.86\,Y_M}S_{NO} + i_{XB}S_{NH} \rightarrow X_{B,M}$	$\mu_M \frac{S_M}{K_{M,M}+S_M}\frac{K_{O,M}}{K_{O,M}+S_O}\frac{S_{NO}}{K_{NO,M}+S_{NO}}\frac{S_{NH}}{K_{NH}+S_{NH}}X_{B,M}$
5	Aerobic growth of autotrophs	$\frac{4.57-Y_A}{Y_A}S_O + \left(i_{XB}+\frac{1}{Y_A}\right)S_{NH} \rightarrow X_{B,A} + \frac{1}{Y_A}S_{NO}$	$\mu_A \frac{S_O}{K_{O,A}+S_O}\frac{S_{NH}}{K_{NH,A}+S_{NH}}X_{B,A}$
6	Decay of heterotrophs	$X_{B,H} \rightarrow (1-f_p)\,X_S + f_p\,X_p + (1-f_p)\,i_{XB}\,X_{ND}$	$b_H\left(\frac{S_O}{K_{O,H}+S_O}+\eta_h\frac{S_{NO}}{K_{NO,H}+S_{NO}}\frac{K_{O,H}}{K_{O,H}+S_O}\right)X_{B,H}$
7	Decay of methylotrophs	$X_{B,M} \rightarrow (1-f_p)\,X_S + f_p\,X_p + (1-f_p)\,i_{XB}\,X_{ND}$	$b_M\left(\frac{S_O}{K_{O,M}+S_O}+\eta_h\frac{S_{NO}}{K_{NO,M}+S_{NO}}\frac{K_{O,M}}{K_{O,M}+S_O}\right)X_{B,M}$
8	Decay of autotrophs	$X_{B,A} \rightarrow (1-f_p)\,X_S + f_p\,X_p + (1-f_p)\,i_{XB}\,X_{ND}$	$b_A\left(\frac{S_O}{K_{O,A}+S_O}+\eta_h\frac{S_{NO}}{K_{NO,A}+S_{NO}}\frac{K_{O,A}}{K_{O,A}+S_O}\right)X_{B,A}$
9	Ammonification of soluble organic nitrogen	$S_{ND} \rightarrow S_{NH}$	$k_a\,S_{ND}X_{Bio}\,,\quad [X_{Bio} = X_{B,H} + X_{B,M} + X_{B,A}]$
10	Hydrolysis of entrapped organics	$X_S \rightarrow S_S$	$k_H\frac{X_S/X_{Bio}}{K_X + X_S/X_{Bio}}\left(\frac{S_O}{K_{O,H}+S_O}+\eta_h\frac{S_{NO}}{K_{NO,H}+S_{NO}}\frac{K_{O,H}}{K_{O,H}+S_O}\right)X_{Bio}$
11	Hydrolysis of entrapped organic nitrogen	$X_{ND} \rightarrow S_{ND}$	$\frac{X_{ND}}{X_S}\mathrm{R}_{10}$

Table 2 | M-ASM1 parameter ranges including the posterior ranges

					Prior Range		Posterior Range	
Type	**Symbol**	**Parameter**	**Unit**	**Factor θ^a**	**Low**	**High**	**Expected Value**	**Standard Deviation**
Heterotrophic kinetics	μ_H	Maximum specific growth rate of heterotrophs	d^{-1}	1.072	2	10	1.742	0.158
	K_S	Substrate half saturation for heterotrophs	g COD m^{-3}	1.03	1	10	8.065	1.274
	$K_{M,H}$	Methanol half saturation for heterotrophs	g COD m^{-3}	1	0.1	0.5	0.254	0.046
	$K_{O,H}$	O_2 half saturation for heterotrophs	g O_2 m^{-3}	1	0.02	0.1	0.030	0.005
	$K_{NO,H}$	NOx half saturation for heterotrophs	g N m^{-3}	1	0.01	0.1	0.067	0.012
	η_g	Anoxic growth reduction for heterotrophs	–	1	0.4	0.8	0.582	0.043
	b_H	Aerobic decay rate coefficient for heterotrophs	d^{-1}	1	0.4	0.7	0.641	0.055
	K_{NH}	NHx half saturation for heterotrophs/methylotrophs	g N m^{-3}	1	0	0.03	0.014	0.004
Methylotrophic kinetics	μ_M	Maximum specific growth rate of methylotrophs	d^{-1}	1.09	0.8	2	0.720	0.053
	$K_{M,M}$	Methanol half saturation coefficient	g COD m^{-3}	1	0.1	1	0.119	0.058
	$K_{O,M}$	O_2 half saturation for methylotrophs	g O_2 m^{-3}	1	0.01	0.1	0.033	0.010
	$K_{NO,M}$	NOx half saturation for methylotrophs	g N m^{-3}	1	0.01	1	0.629	0.171
	b_M	Aerobic decay rate coefficient for methylotrophs	d^{-1}	1.03	0.04	0.1	0.098	0.008
Autotrophic kinetics	μ_A	Maximum specific growth rate of autotrophs	d^{-1}	1.072	0.7	1.2	1.059	0.058
	$K_{NH,A}$	Ammonia half saturation for autotrophs	g N m^{-3}	1	0.5	1	0.952	0.049
	$K_{NO,A}$	NOx half saturation for autotrophs	g N m^{-3}	1	0.01	0.2	0.020	0.010
	$K_{O,A}$	Oxygen half saturation for autotrophs	g O_2 m^{-3}	1	0.1	0.5	0.238	0.080
	b_A	Aerobic decay rate coefficient for autotrophs	d^{-1}	1.03	0.15	0.25	0.185	0.009
Conversion kinetics	η_h	Anoxic growth reduction for decay	–	1	0.3	0.8	0.736	0.065
	K_h	Hydrolysis rate coefficient	d^{-1}	1.03	1	3	1.088	0.172
	K_X	Hydrolysis half saturation coefficient	–	1	0.01	0.2	0.095	0.028
	K_a	Ammonification rate coefficient	d^{-1}	1.03	0.01	0.1	0.057	0.021
Stoichiometries	Y_H	Aerobic yield of heterotrophs on substrate	–	1	0.6	0.7	0.611	0.011
	Y_{HM}	Aerobic yield of heterotrophs on methanol	–	1	0.4	0.4	–	–
	Y_A	Autotroph yield	–	1	0.24	0.24	–	–
	Y_M	Methylotroph yield	–	1	0.3	0.5	0.462	0.011
	f_P	Endogenous fraction (death-regeneration)	–	1	0.08	0.08	–	–

(*continued*)

Table 2 | continued

Type	Symbol	Parameter	Unit	Factor θ^a	Prior Range Low	Prior Range High	Posterior Range Expected Value	Posterior Range Standard Deviation
Partitioning coefficients	i_{XB}	Nitrogen fraction in biomass	g N g COD^{-1}	1	0.05	0.1	0.058	0.005
	$i_{VSS,B}$	COD/VSS ratio of biomass	g COD g VSS^{-1}	1	1.42	1.42	–	–
	$i_{VSS,i}$	COD/VSS ratio of X_i	g COD g VSS^{-1}	1	1.5	2	–	–
	$i_{VSS,s}$	COD/VSS ratio of X_s	g COD g VSS^{-1}	1	1.8	1.8	–	–
	$i_{VSS,P}$	COD/VSS ratio of X_p	g COD g VSS^{-1}	1	1.42	1.42	–	–
	i_{MeOH}	COD ratio of methanol	g COD g^{-1}	1	1.5	1.5	–	–
Mass transfer	$k_{L,O2}$	Oxygen mass transfer rate	day^{-1}	1	50	250	190.446	7.780

[a]Temperature dependent parameters: $P(T) = P_{20}\theta^{(T-20)}$.

distributions of each parameter are assumed to follow a log-normal distribution, as suggested by Cox (2004). The given range for parameters was obtained from various sources (Bullock *et al.* 1996; Lee & Welander 1996; Cox 2004; Baytshtok *et al.* 2008; Dold *et al.* 2008; Hauduc *et al.* 2011; Alikhani *et al.* in press). Although the partition coefficient (components of **Z** vector in Equation (3)) can contribute to the uncertainty, in this research the main focus was on estimating the bio-kinetics and stoichiometric parameters, so the partitioning coefficients were assumed to be known with an acceptable level of certainty.

Using the M–H algorithm, 500,000 samples in 10 chains were drawn from the posterior distribution of parameters. To expedite the convergence of the MCMC algorithm, the deterministically estimated parameters using a hybrid GA with 500 generations, and a population of 50 were used as the starting point of each MCMC chain. This expedites the convergence of the M–H algorithm by decreasing the number of burn-in samples. The posterior probability of each parameter was evaluated by the Bayesian inference introduced in Equation (5), assuming a log-normal error structure. To reach stationary distributions, 100,000 of the sampled parameters were discarded due to the burn-in period (Brooks *et al.* 2011).

RESULTS AND DISCUSSION

Bayesian credible interval of parameters

Figure 3 shows the model's ability to reproduce observed data. The 95% bands have been obtained by performing 1,000 realizations of Monte Carlo simulations using the posterior distributions of parameters and error standard deviations for each observed constituent. As shown in Figure 3, the model performed well in terms of capturing the trends and magnitudes of the observed data. The spread of data points at identical times in stages 3B and 4 is an indicator of the heterogeneity of concentrations inside a single stage, which is represented as homogeneous CSTR in the model. This variability can be due to actual variations in the concentrations at small temporal and spatial scales, or errors due to sampling methods and chemical analysis. It appears that trends in effluent concentrations generally reproduced better results compared to the spatial profile data.

Figure 4 illustrates the posterior distributions for all the estimated parameters in the model, shown with solid lines, along with the prior distribution of parameters, shown as dashed curves. The expected value (mean) and standard deviation of the posterior distributions of parameters are also provided in the last two columns of Table 2. The posterior distribution represents our degree of confidence about the parameter value. By and large, the ranges of values of parameters are consistent with previous studies. For example, the estimated range for parameter b_H (0.5–0.97 d^{-1}) is consistent with the deterministically estimated value of 0.926 d^{-1} by Keskitalo & Leiviskä (2012) using the ASM1 model. The estimated 95% range for b_A (0.17–0.21 d^{-1}) is larger than the value of 0.12 d^{-1} obtained by Keskitalo & Leiviskä (2012) but consistent with the value of 0.17 d^{-1} suggested by Fall *et al.* (2011) and the value of 0.18 d^{-1} suggested by Mannina *et al.* (2012). The estimated range for K_s (4.53–11.22 g COD m^{-3}) is smaller than the value of 14.48 g COD m^{-3} found by Keskitalo & Leiviskä (2012). The ranges of $K_{NH,A}$ (0.8–1.06 g N m^{-3}) and K_{NH} (0.01–0.03 g N m^{-3}) are both smaller than the value of

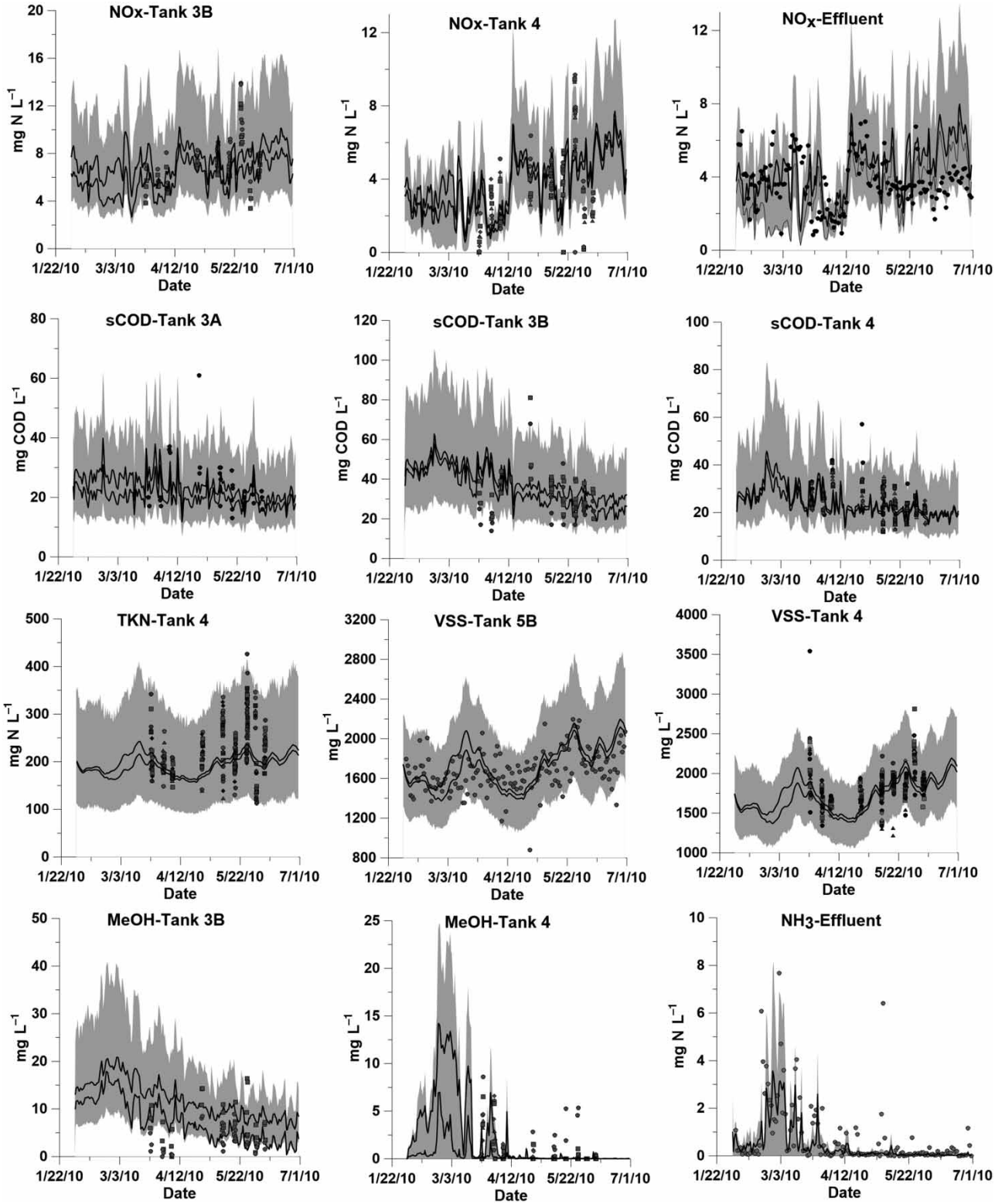

Figure 3 | Modeled 95% intervals and measured constituent concentrations at the effluent and inside the bioreactors. The solid lines represent the 95% intervals of model predictions only considering parameter uncertainty, and the gray areas represent the 95% interval considering parameter uncertainty as well as other uncertainties, including measurement errors and model structural error. Dots show the measured data points.

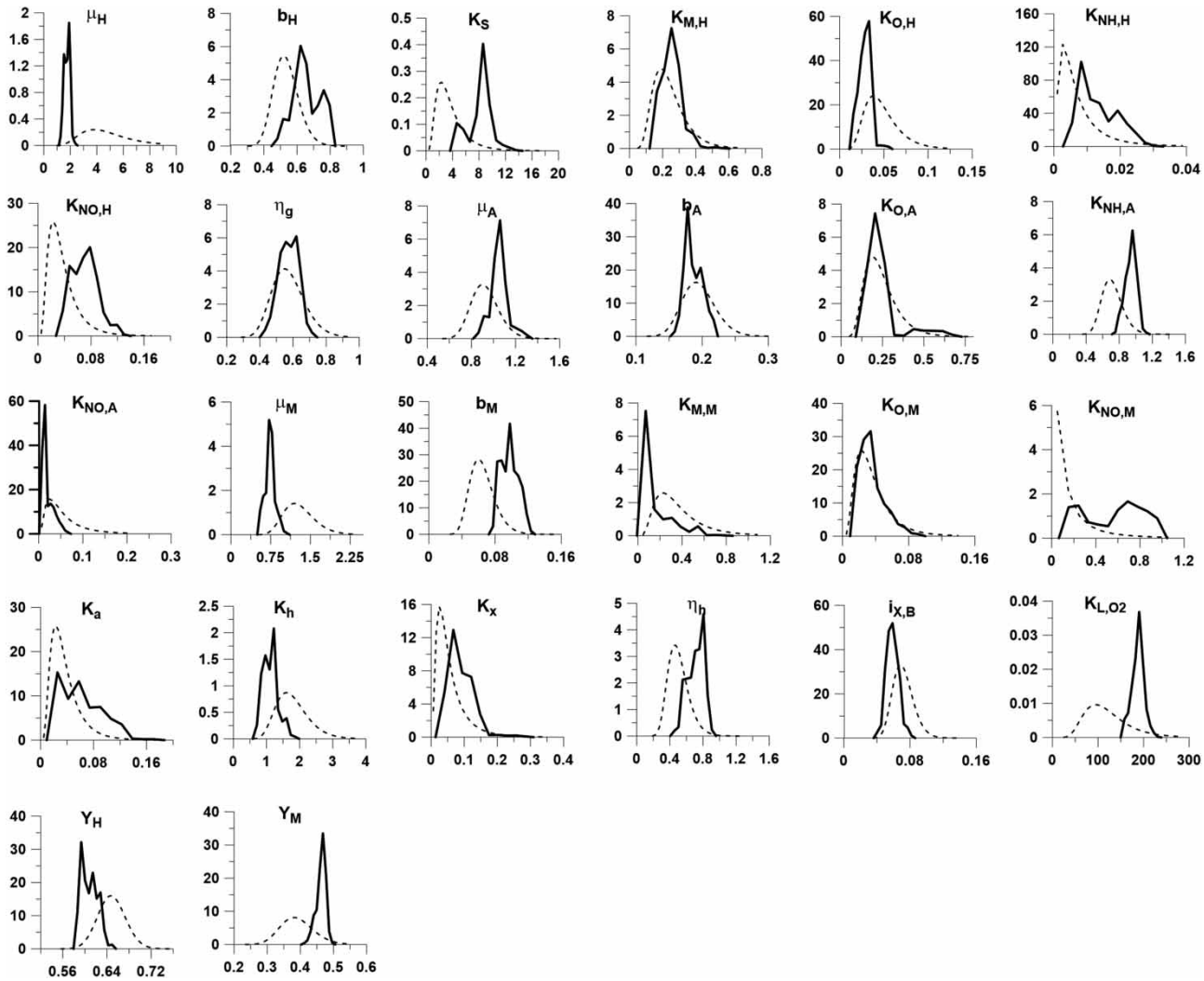

Figure 4 | Posterior distribution (solid lines) and prior distribution (dashed lines) of M-ASM1 parameters for 500,000 MCMC samples.

1.41 g N m^{-3} provided by Mannina *et al.* (2012). The inconsistency between the K_{NH} ranges obtained in this study with previous studies is due to the consideration of two different parameters for ammonia half saturation for autotrophs and heterotrophs, which was not done in ASM1. It worth noting that the presented approach does not capture the possible temporal changes in the values of the parameters that may occur as a result of aging or seasonal effects. However, the presence of such temporal variations in the values of model parameters is implicitly laid within the posterior interval. In order to explicitly capture these temporal changes, the processes leading to them should be incorporated into the model.

The capability of the data to provide information for each parameter can be assessed by comparing the entropy of posterior distribution compared to the prior distribution (Shannon 2001; Lewis *et al.* in press):

$$I_i = \int p(\theta_i|\tilde{\boldsymbol{Y}})\ln[p(\theta_i|\tilde{\boldsymbol{Y}})]d\theta_i - \int p(\theta_i)\ln[p(\theta_i)]d\theta_i \quad (11)$$

where I_i is the information content gained for parameter i (referred to as entropy improvement), the first term on the right-hand side is the entropy of posterior distribution and the second term on the right-hand side is the entropy of the prior distribution of parameter i. A positive value for I_i shows a degree of information gain for parameter i. The higher the value of I_i the higher the level of information is obtained by applying the observed data in the Bayesian inference.

To evaluate the amount of information content of the observed data two scenarios have been tested, one with higher frequency time intervals of observed data and the second one a lower frequency basis. Figure 5 illustrates the reduction of the 95% credible intervals for both scenarios due to posterior distribution by mapping the 95% equal-tail credible intervals (Mukhopadhyay 2000; Casella & Berger 2002) of each parameter – by excluding 2.5% from each tail of posterior distribution – inside the 95% confidence intervals of prior distributions. Calculated values of I_i for all model parameters for two scenarios are shown in Table 3. In Figure 5, a narrower 95% posterior range of a parameter relative to its prior indicates more information being provided by the observed data. Moreover, the value of I_i in Table 3 quantitatively shows the level of obtained information while a larger value of I_i indicates a higher level of information obtained for the parameter i by applying the observed data.

The values of the entropy improvement (I) for each parameter in Table 3 is consistent with the posterior credible intervals shown in Figure 5. In addition, the entropy improvement values (I) of the parameters are consistent with the sensitivity level of outputs with respect to them, with parameters having larger SSFs (Figure 6) typically showing a larger increase in the value of I. Table 3 shows the entropy improvement for both scenarios of high frequency and low frequency observed data. The comparison between the two scenarios clearly shows that for all parameters, except for some of the observed standard deviations, the information gained from the high frequency scenario is higher than the information obtained from the low frequency scenario. This result indicates that the information content of the observed data is a function of the number of data points and the frequency of sampling.

For the first scenario (labeled as High Frequency (daily) Obs. Data), the observed data provided a good level of a posterior confidence interval for a large number of the parameters, which can be interpreted as reducing a parameter's uncertainty. For most of the parameters presenting higher sensitivity (Figure 6), including μ_H, $K_{L,O2}$, μ_M, Y_M, $K_{M,M}$, $K_{O,H}$, and b_A, the uncertainty was reduced by more than 60% in comparison with their prior range. Among the kinetics parameters, large reductions in the uncertainty are obtained for the maximum specific growth rates of heterotrophs and methylotrophs, μ_H and μ_M, respectively, and the decay rate for autotrophs, b_A. Among the yield coefficients, both methylotrophs and heterotrophs yield coefficients, Y_M and Y_H, showed high levels of reduction in the uncertainty. Most of the half saturation coefficients represent low levels of sensitivity in comparison with growth, decay, and yield parameters, as shown in the 'Sensitivity analysis' section (Figure 6), and thus for them lower levels of uncertainty reduction are obtained. The exceptions are the oxygen half saturation of heterotrophs $K_{O,H}$ and the methanol half saturation of methylotrophs $K_{M,M}$, which showed around 70% reduction in the 95% posterior interval relative to their prior ranges. The oxygen mass transfer rate $K_{L,O2}$ also showed a large reduction in its posterior intervals, although this could be due to the lack of prior knowledge about its value. As for the parameters related to the methylotrophic process the

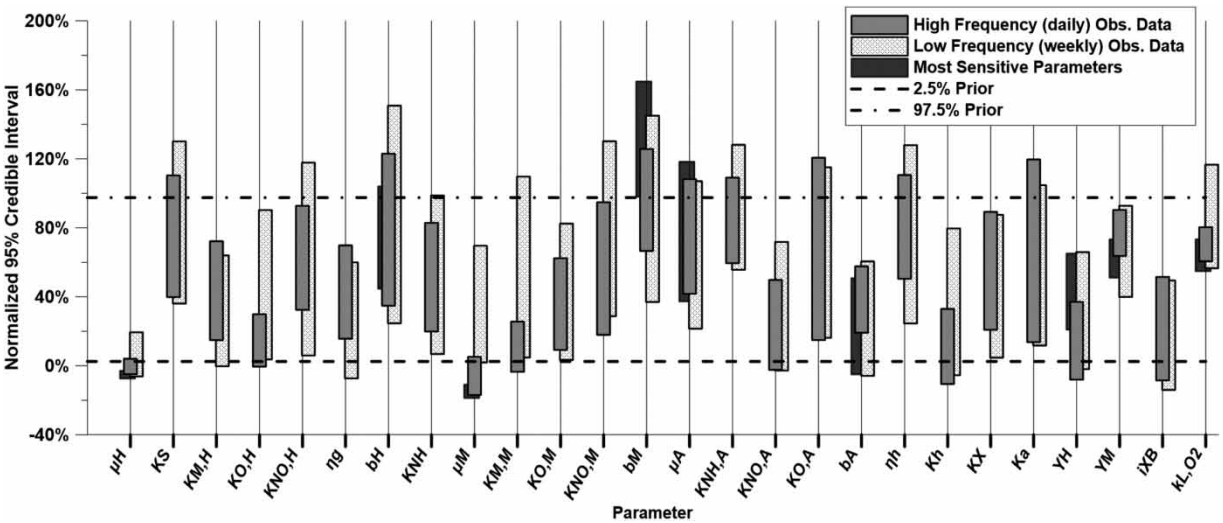

Figure 5 | 95% Bayesian credible intervals of parameters normalized by the length of the range of each parameter (floating bars) in comparison with the 95% prior credible range (dashed lines).

Table 3 | Entropy of prior and posterior distribution indicating how much information gained by applying Bayes's theorem on the observed data

			High Frequency Obs. Data		Low Frequency Obs. Data	
Type	**Symbol**	**Prior Entropy**	**Posterior Entropy**	***I***	**Posterior Entropy**	***I***
Heterotrophic kinetics	μ_H	−0.87	0.07	0.94	−0.30	0.58
	K_S	−0.87	−0.71	0.16	−1.14	−0.27
	$K_{M,H}$	0.43	0.60	0.17	0.52	0.09
	$K_{O,H}$	1.13	1.52	0.40	1.10	−0.03
	$K_{NO,H}$	1.12	1.09	−0.03	0.95	−0.17
	η_g	0.39	0.63	0.24	0.53	0.13
	b_H	0.51	0.51	0.00	0.44	−0.07
	K_{NH}	1.64	1.64	0.00	0.83	−0.80
Methylotrophic kinetics	μ_M	−0.08	0.43	0.51	0.07	0.15
	$K_{M,M}$	0.13	0.24	0.12	0.04	−0.08
	$K_{O,M}$	1.12	1.15	0.03	0.89	−0.23
	$K_{NO,M}$	0.16	0.07	−0.10	−0.62	−0.78
	b_M	1.22	1.39	0.16	1.36	0.14
Autotrophic kinetics	μ_A	0.29	0.54	0.25	0.44	0.15
	$K_{NH,A}$	0.30	0.55	0.26	0.47	0.18
	$K_{NO,A}$	0.85	1.23	0.38	1.00	0.15
	$K_{O,A}$	0.43	0.55	0.12	0.37	−0.06
	b_A	0.99	1.28	0.29	1.15	0.16
Conversion kinetics	η_h	0.30	0.43	0.12	0.35	0.04
	K_h	−0.29	0.07	0.37	−0.23	0.07
	K_X	0.85	0.81	−0.04	0.65	−0.20
	K_a	1.12	0.82	−0.30	0.79	−0.33
Stoichiometries	Y_H	0.99	1.23	0.24	1.12	0.13
	Y_M	0.69	1.25	0.56	1.08	0.39
	i_{XB}	1.30	1.52	0.22	1.44	0.15
Standard deviations	σ_0	−0.09	1.22	1.31	0.90	0.98
	σ_1	−0.09	0.56	0.65	0.67	0.76
	σ_2	−0.09	0.39	0.47	0.46	0.55
	σ_3	−0.09	1.44	1.53	1.48	1.56
	σ_4	−0.09	0.46	0.55	0.36	0.45
	σ_5	−0.09	0.58	0.67	0.53	0.62
	σ_6	−0.09	0.32	0.40	0.37	0.46
	σ_7	−0.09	1.60	1.69	1.37	1.46
Mass transfer	$k_{L,O2}$	−2.27	−1.68	0.59	−2.04	0.22

posterior ranges for μ_M, b_M, Y_M, and $K_{M,M}$ were significantly reduced with respect to the given prior range, especially for μ_M and b_M where the 95% Bayesian credible interval is almost outside the prior range, indicating poor prior knowledge about these parameters that can be enhanced by applying the Bayesian parameter estimation method.

For the second scenario (labeled as Low Frequency (weekly) Obs. Data in Figure 5), the spatial profile data were reduced to one-third of their original size (from 1,650 data points to 450 data points) and the effluent observed data were reduced to weekly intervals instead of their original daily intervals. As can be seen in Figure 5, the posterior ranges of the low frequency observed data are, as might be expected, wider for almost all of the parameters, especially for the μ_H, $K_{O,H}$, μ_M, $K_{M,M}$, Y_M, η_h, and $K_{L,O2}$. However, in some cases, the minimum to maximum ranges are slightly moved. This observation indicates that high frequency observed data carry a higher amount of information about most of the parameters.

A common practice when calibrating in an over-parameterized model such as ASM is to fix the parameters that are deemed to present lower sensitivity and only calibrate the model using the high-sensitivity parameters (Sin *et al.* 2005). This practice is helpful because it reduces the dimension of the search space and makes the calibration more computationally efficient and it also avoids the chance of becoming

trapped in local minima for an over-parameterized ASM model. To study the effect of this decision the analysis was also performed by only considering the most sensitive parameters (Figure 6). The specific growth rate and decay rate of all types of bacteria (μ_H, b_H, μ_M, b_M, μ_A, and b_A), the heterotrophic and methylotriophic yield coefficients (Y_H and Y_M), and the oxygen mass transfer coefficient rate ($K_{L,O2}$) were treated as unknown parameters. The remainder of the parameters were fixed at the maximum likelihood estimation value found by the GA. The resulting 95% intervals for this scenario are shown in Figure 5 (labeled as Most Sensitive Parameters). As shown in Figure 5, the resulting 95% posterior intervals when only the nine most sensitive parameters are estimated mostly overlap with the results when all the parameters are estimated, with the ranges being mostly narrower when only nine parameters are considered. As expected the narrow confidence brackets obtained for μ_H and μ_m and to some degree Y_m and $K_{L,O2}$, compared to the case when all parameters are included in the analysis (in the scenario of high frequency observed data), shows that when only the most sensitive parameters are considered in the uncertainty analysis the uncertainty assigned to these parameters can be under-estimated. For a few parameters, the uncertainty increases when only the sensitive parameters are considered. This can be due to the model's lower flexibility to reproduce the data, therefore introducing larger observed error standard deviations.

In all the cases, a moderate to small amount of information was obtained regarding the greater number of parameters. This is likely due to: (a) the model being

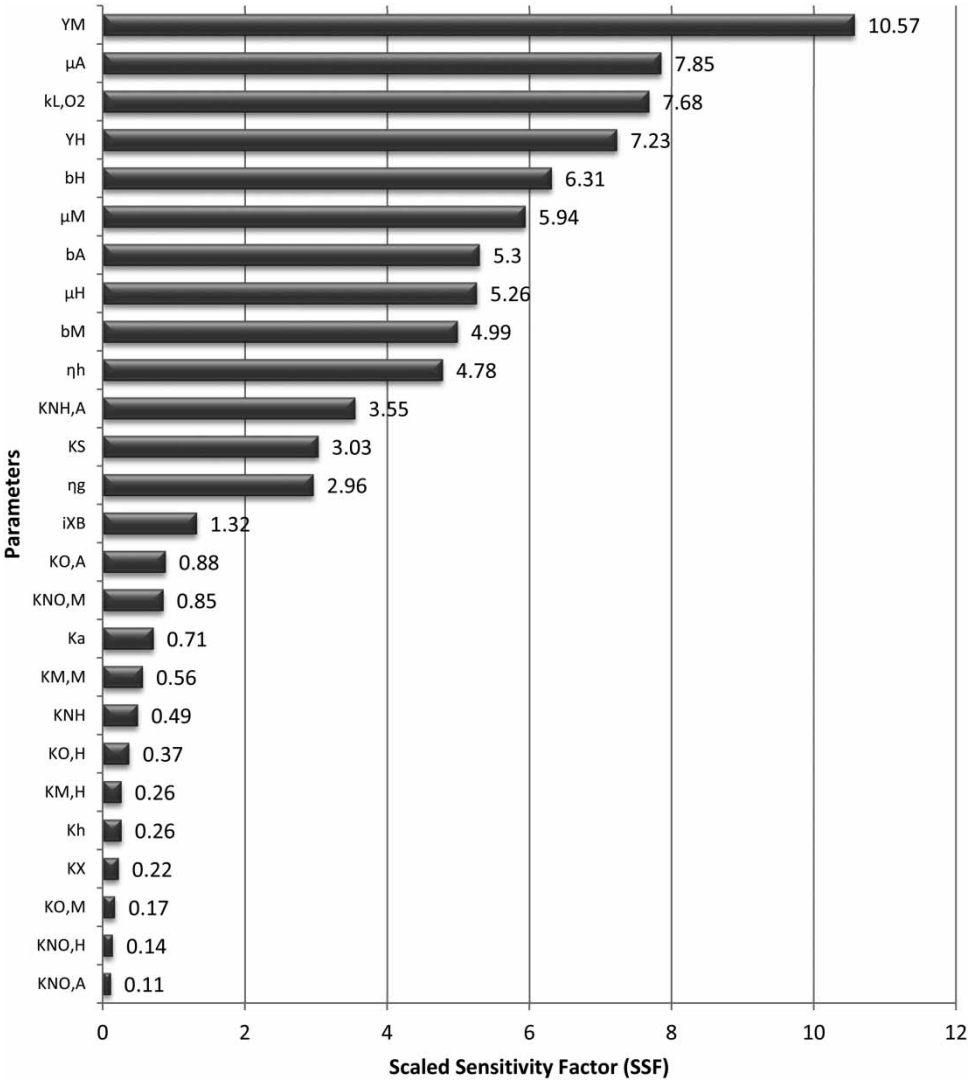

Figure 6 | Ranked SSF for the global sensitivity analysis of M-ASM1 parameters.

over-parameterized with respect to the amount and the diversity of data and (b) the fact that the operational conditions of the full-scale bioreactors are fairly constant and the observed data are non-informative regarding some of the parameters under those conditions. The ability of the observed data to inform the values of the parameters (i.e., by providing a narrower posterior interval in comparison to the prior interval) depends on several factors including: (1) the sensitivity of the measured constituents with respect to the parameters, (2) the correlation between the parameters (collinearity), and (3) the weight of the observed constituents that are most sensitive with respect to the parameter.

Sensitivity analysis

Figure 6 shows the ranked SSF for each parameter. Among the stoichiometric and the rate parameters, Y_M, Y_H, μ_A, b_A, μ_M, b_M, μ_H, and b_H showed a higher level of sensitivity compared to the other parameters. For all of the half saturation parameters, the model showed relatively lower sensitivities. The model also showed a relatively higher sensitivity with respect to the oxygen mass transfer rate coefficient than for other parameters. In general, the sensitivity analysis was in agreement with the uncertainty reduction, as the parameters with higher sensitivity achieved a higher level of uncertainty reduction. Nevertheless, sensitivity is not the only factor in the uncertainty analysis. A high to moderate level of information, as determined by the reduction of the 95% posterior interval relative to the prior was obtained for most high-sensitivity parameters, specifically μ_H, $K_{L,O2}$, and $K_{NH,A}$. The exceptions are $K_{O,H}$ and $K_{M,M}$, for which a significant level of information was obtained despite a small sensitivity factor. It is speculated that this is because the prior interval attributed to these parameters was exceptionally wide due to a lack of adequate prior knowledge from the literature about their values. The small (effectively zero) measured effluent concentrations of methanol also force $K_{M,M}$ to be small, narrowing its posterior interval.

In general, a positive correlation between SSF and reduction in the credibility interval can be seen. The global sensitivity with respect to most of the parameters for which the information gain was non-significant was found to be small. Global sensitivities of the parameters estimated here represent sensitivities under operational conditions during which the data were collected. This means that as long as the operational condition of the bioreactor is within the range experienced during the data collection period, the influence of the values of parameters with small information gain (lower sensitivity) will not be substantial. However, evaluating the impact of other operational conditions on the value of these parameters requires further study.

The global sensitivity results are fairly consistent with previous studies. In other studies, the maximum growth rates, decay rates, and biomass yields have been reported as high-sensitivity parameters in the ASMs. Kim *et al.* (2010) reported i_{Xp}, η_g, k_{NH}, b_A, μ_A, k_x, η_h, and f_p as parameters with large effects on the overall output, while Sharifi *et al.* (2014) found Y_A, b_H, Y_H, and μ_A to be highly sensitive parameters for ASM1 based on a local sensitivity analysis at the optimal value of the parameters. Based on a local sensitivity analysis, Fang *et al.* (2010) reported that effluent COD and ammonia/ammonium (NH) concentrations in ASM3 are more sensitive with respect to μ_A, μ_H, Y_A, and Y_H.

Correlation between parameters

Posterior results can be used to estimate the posterior correlation between parameters (or parameter collinearity). A high correlation between two parameters means that ways exist to change both of them without seeing a change in the model output or its fitness in terms of the overall ability to reproduce the data. The correlation matrix is provided in the supplementary information for the sake of brevity (available with the online version of this paper). Significant posterior correlations exist between μ_H and b_H, μ_H and K_s, b_H and b_M, and μ_M and $K_{M,M}$. Overall, posterior correlations were smaller than the case where synthetic (model generated) data were used as the observed data (Sharifi *et al.* 2014). This may be due to greater heterogeneity and structural errors associated with the real observed data, widening the range of the parameters, which reduces the correlation factor.

CONCLUSIONS

In this study, a stochastic parameter estimation method was applied to measured data collected from the NDN reactors at the Blue Plains advanced WWTP in Washington, DC. Wastewater characteristics, including sCOD, VSS, nitrite/nitrate, ammonia, and methanol concentration, were measured inside the reactor in different tanks, locations, and times, as well as in the effluent wastewater. The goal of this study was to evaluate whether effluent and spatial profile data obtained from a full-scale plant could provide information about the values of model parameters and to quantify the level of confidence they can provide for each parameter relative to its prior knowledge. The following general conclusions were obtained from the presented study.

1. The model was generally able to capture the magnitude and trends of the observed data.
2. The data were able to significantly improve the level of confidence for some of the model parameters as determined by the increase in the entropy of posterior relative to the prior distribution of parameters.
3. Model prediction's uncertainty reduced by applying the posterior distributions of parameters instead of using the prior ranges.
4. Backward uncertainty assessment reveals the correlations between the parameters, representing how the combined variation of two or more parameters can result in the similar reproduction of observed data.
5. By and large, the overall sensitivity of the model output with respect to each parameter as measured by an SSF can explain the ability of data to narrow down the ranges of each parameter.
6. The posterior range of methylotrophic parameters was found to be narrowed down substantially, which reflects the poor prior knowledge about these parameters.
7. The number and the frequency of observed data can influence the level of information gained by the parameters. The results showed that lower frequency in observed data sampling (daily vs. weekly) can narrow down the posterior distribution.

ACKNOWLEDGEMENTS

This project was supported by the District of Columbia Water and Sewer Authority and partially by the DC Water Resources Research Institute.

REFERENCES

Albrecht, J. 2013 Estimating reaction model parameter uncertainty with Markov Chain Monte Carlo. *Computers & Chemical Engineering* **48**, 14–28.

Alikhani, J., Takacs, I., Omari, A. A., Murthy, S., Mokhayeri, Y. & Massoudieh, A. 2014 Inverse modeling of nitrification-denitrification processes: a case study on the Blue Plains wastewater treatment plant in Washington, DC. *Proceedings of the Water Environment Federation* **2014** (19), 4556–4567.

Alikhani, J., Stewart, H. A., Takacs, I., Al-Omari, A., Murthy, S. & Massoudieh, A. 2015 Time series analysis for uncertainty evaluation of influent fluctuation of a wastewater treatment plant. *Proceedings of the Water Environment Federation* **2015** (8), 3519–3526.

Alikhani, J., Deinhart, A. L., Visser, A., Bibby, R. K., Purtschert, R., Moran, J. E., Massoudieh, A. & Esser, B. K. 2016a Nitrate vulnerability projections from Bayesian inference of multiple groundwater age tracers. *Journal of Hydrology* **543** (Part A), 167–181.

Alikhani, J., Shoghli, B., Bhowmik, U. K. & Massoudieh, A. 2016b An adaptive time-step backward differentiation algorithm to solve stiff ordinary differential equations: application to solve activated sludge models. *American Journal of Computational Mathematics* **6**, 298–312.

Alikhani, J., Massoudieh, A. & Bhowmik, U. K. 2016c GPU-accelerated solution of Activated Sludge Model's system of ODEs with a high degree of stiffness. In: *Proceedings of the 2016 International Conference on Computational Science and Computational Intelligence*.

Alikhani, J., Al-Omari, A., De Clippeleir, H., Murthy, S., Takacs, I. & Massoudieh, A. in press Assessment of the endogenous respiration rate and the observed biomass yield for methanol-fed denitrifying bacteria under anoxic and aerobic conditions. *Water Science and Technology* DOI: 10.2166/wst.2016.486.

Andreottola, G., Bortone, G. & Tilche, A. 1997 Experimental calibration and validation of a simulation model for nitrification/denitrification in SBRs. *Water Science and Technology* **35**, 113–120.

Barker, P. S. & Dold, P. L. 1997 General model for biological nutrient removal activated-sludge systems: model presentation. *Water Environment Research* **69** (5), 969–984.

Baytshtok, V., Kim, S., Yu, R., Park, H. & Chandran, K. 2008 Molecular and biokinetic characterization of methylotrophic denitrification using nitrate and nitrite as terminal electron acceptors. *Water Science and Technology* **58** (2), 359–365.

Beven, K. & Freer, J. 2001 Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of Hydrology* **249** (1–4), 11–29.

Bixio, D., Parmentier, G., Rousseau, D., Verdonck, F., Meirlaen, J., Vanrolleghem, P. A. & Thoeye, C. 2002 A quantitative risk analysis tool for design/simulation of wastewater treatment plants. *Water Science and Technology* **46** (4–5), 301–307.

Bolong, N., Ismail, A. F., Salim, M. R. & Matsuura, T. 2009 A review of the effects of emerging contaminants in wastewater and options for their removal. *Desalination* **239** (1–3), 229–246.

Brooks, S., Gelman, A., Jones, G. & Meng, X. L. 2011 *Handbook of Markov Chain Monte Carlo*. CRC Press, Boca Raton, FL, USA.

Brun, R., Kühni, M., Siegrist, H., Gujer, W. & Reichert, P. 2002 Practical identifiability of ASM2d parameters – systematic selection and tuning of parameter subsets. *Water Research* **36** (16), 4113–4127.

Bullock, C. M., Bicho, P. A., Zhang, Y. & Saddler, J. N. 1996 A solid chemical oxygen demand (COD) method for determining biomass in waste waters. *Water Research* **30** (5), 1280–1284.

Busch, J., Elixmann, D., Kühl, P., Gerkens, C., Schlöder, J. P., Bock, H. G. & Marquardt, W. 2013 State estimation for large-

scale wastewater treatment plants. *Water Research* **47** (13), 4774–4787.

Casella, G. & Berger, R. L. 2002 *Statistical Inference.* Duxbury, Pacific Grove, CA, USA.

Choubert, J.-M., Stricker, A.-E., Marquot, A., Racault, Y., Gillot, S. & Héduit, A. 2009 Updated activated sludge model number 1 parameter values for improved prediction of nitrogen removal in activated sludge processes: validation at 13 full-scale plants. *Water Environment Research* **81** (9), 858–865.

Comas, J., Rodríguez-Roda, I., Gernaey, K. V., Rosen, C., Jeppsson, U. & Poch, M. 2008 Risk assessment modelling of microbiology-related solids separation problems in activated sludge systems. *Environmental Modelling & Software* **23** (10–11), 1250–1261.

Cox, C. D. 2004 Statistical distributions of uncertainty and variability in activated sludge model parameters. *Water Environment Research* **76** (7), 2672–2685.

Dold, P., Takacs, I., Mokhayeri, Y., Nichols, A., Hinojosa, J., Riffat, R., Bott, C., Bailey, W. & Murthy, S. 2008 Denitrification with carbon addition–kinetic considerations. *Water Environment Research* **80** (5), 417–427.

Dotto, C. B. S., Kleidorfer, M., Deletic, A., Rauch, W., McCarthy, D. T. & Fletcher, T. D. 2011 Performance and sensitivity analysis of stormwater models using a Bayesian approach and long-term high resolution data. *Environmental Modelling & Software* **26** (10), 1225–1239.

Eberly, L. E. & Casella, G. 2003 Estimating Bayesian credible intervals. *Journal of Statistical Planning and Inference* **112** (1–2), 115–132.

Ekama, G. A. & Wentzel, M. C. 2004 A predictive model for the reactor inorganic suspended solids concentration in activated sludge systems. *Water Research* **38** (19), 4093–4106.

Fall, C., Espinosa-Rodriguez, M. A., Flores-Alamo, N., van Loosdrecht, M. C. & Hooijmans, C. M. 2011 Stepwise calibration of the activated sludge model no. 1 at a partially denitrifying large wastewater treatment plant. *Water Environment Research* **83** (11), 2036–2048.

Fang, F., Ni, B.-J., Xie, W.-M., Sheng, G.-P., Liu, S.-G., Tong, Z.-H. & Yu, H.-Q. 2010 An integrated dynamic model for simulating a full-scale municipal wastewater treatment plant under fluctuating conditions. *Chemical Engineering Journal* **160** (2), 522–529.

Flores-Alsina, X., Rodríguez-Roda, I., Sin, G. & Gernaey, K. V. 2008 Multi-criteria evaluation of wastewater treatment plant control strategies under uncertainty. *Water Research* **42** (17), 4485–4497.

Foglia, L., Hill, M. C., Mehl, S. W. & Burlando, P. 2009 Sensitivity analysis, calibration, and testing of a distributed hydrological model using error-based weighting and one objective function. *Water Resources Research* **45** (6), W06427.

Freni, G., Mannina, G. & Viviani, G. 2009 Identifiability analysis for receiving water body quality modelling. *Environmental Modelling & Software* **24** (1), 54–62.

Gamerman, D. & Lopes, H. F. 2006 *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference,* 2nd edn. Taylor & Francis, Chichester, UK.

Gernaey, K. V., van Loosdrecht, M. C. M., Henze, M., Lind, M. & Jørgensen, S. B. 2004 Activated sludge wastewater treatment plant modelling and simulation: state of the art. *Environmental Modelling & Software* **19** (9), 763–783.

Hallin, S., Throback, I. N., Dicksved, J. & Pell, M. 2006 Metabolic profiles and genetic diversity of denitrifying communities in activated sludge after addition of methanol or ethanol. *Applied and Environmental Microbiology* **72** (8), 5445–5452.

Hao, X., Wang, Q., Cao, Y. & van Loosdrecht, M. C. M. 2011 Evaluating sludge minimization caused by predation and viral infection based on the extended activated sludge model No. 2d. *Water Research* **45** (16), 5130–5140.

Hauduc, H., Rieger, L., Takacs, I., Heduit, A., Vanrolleghem, P. A. & Gillot, S. 2010 A systematic approach for model verification: application on seven published activated sludge models. *Water Science and Technology* **61** (4), 825–839.

Hauduc, H., Rieger, L., Ohtsuki, T., Shaw, A., Takacs, I., Winkler, S., Heduit, A., Vanrolleghem, P. A. & Gillot, S. 2011 Activated sludge modelling: development and potential use of a practical applications database. *Water Science and Technology* **63** (10), 2164–2182.

Henze, M., Leslie Grady Jr, C. P., Gujer, W., Marais, G. V. R., Matsuo, T., Henze, M., Leslie Grady Jr, C. P., Gujer, W., Marais, G. V. R. & Matsuo, T. 1987 A general model for single-sludge wastewater treatment systems. *Water Research* **21** (5), 505–515.

Henze, M., Gujer, W., Mino, T. & van Loosdrecht, M. 2000 *Activated Sludge Models ASM1, ASM2, ASM2d and ASM3*, IWA Publishing London.

Hill, M. C. & Tiedeman, C. R. 2007 *Effective Calibration of Groundwater Models, with Analysis of Data, Sensitivities, Predictions, and Uncertainty.* John Wiley, New York.

Kaelin, D., Manser, R., Rieger, L., Eugster, J., Rottermann, K. & Siegrist, H. 2009 Extension of ASM3 for two-step nitrification and denitrification and its calibration and validation with batch tests and pilot scale data. *Water Research* **43** (6), 1680–1692.

Kaipio, J. & Somersalo, E. 2005 *Statistical and Computational Inverse Problems.* Springer, New York.

Kalyanaraman, J., Fan, Y., Labreche, Y., Lively, R. P., Kawajiri, Y. & Realff, M. J. 2015 Bayesian estimation of parametric uncertainties, quantification and reduction using optimal design of experiments for CO_2 adsorption on amine sorbents. *Computers & Chemical Engineering* **81**, 376–388.

Kaplan, S. 1997 The words of risk analysis. *Risk Analysis* **17** (4), 407–417.

Keskitalo, J. & Leiviskä, K. 2012 Application of evolutionary optimisers in data-based calibration of Activated Sludge Models. *Expert Systems with Applications* **39** (7), 6609–6617.

Kim, Y. S., Kim, M. H. & Yoo, C. K. 2010 A new statistical framework for parameter subset selection and optimal parameter estimation in the activated sludge model. *Journal of Hazardous Materials* **183** (1–3), 441–447.

Koch, G., Kühni, M. & Siegrist, H. 2001 Calibration and validation of an ASM3-based steady-state model for activated sludge systems – part I: prediction of nitrogen removal and sludge production. *Water Research* **35** (9), 2235–2245.

Lee, N. M. & Welander, T. 1996 The effect of different carbon sources on respiratory denitrification in biological wastewater treatment. *Journal of Fermentation and Bioengineering* **82** (3), 277–285.

Lewis, P. O., Chen, M.-H., Kuo, L., Lewis, L. A., Fučíková, K., Neupane, S., Wang, Y.-B. & Shi, D. in press Estimating Bayesian phylogenetic information content. *Systematic Biology* DOI: 10.1093/sysbio/syw042.

Lindberg, C. F. & Carlsson, B. 1996 Adaptive control of external carbon flow rate in an activated sludge process. *Water Science and Technology* **34** (3–4), 173–180.

Lu, D., Ye, M., Hill, M. C., Poeter, E. P. & Curtis, G. P. 2014a A computer program for uncertainty analysis integrating regression and Bayesian methods. *Environmental Modelling & Software* **60** (0), 45–56.

Lu, H., Chandran, K. & Stensel, D. 2014b Microbial ecology of denitrification in biological wastewater treatment. *Water Research* **64**, 237–254.

Machado, V. C., Tapia, G., Gabriel, D., Lafuente, J. & Baeza, J. A. 2009 Systematic identifiability study based on the Fisher Information Matrix for reducing the number of parameters calibration of an activated sludge model. *Environmental Modelling & Software* **24** (11), 1274–1284.

Mannina, G., Cosenza, A. & Viviani, G. 2012 Uncertainty assessment of a model for biological nitrogen and phosphorus removal: application to a large wastewater treatment plant. *Physics and Chemistry of the Earth, Parts A/B/C* **42–44**, 61–69.

Meijer, S. C. F., van Loosdrecht, M. C. M. & Heijnen, J. J. 2001 Metabolic modelling of full-scale biological nitrogen and phosphorus removing wwtps. *Water Research* **35** (11), 2711–2723.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. 1953 Equations of state calculations by fast computing machines. *Journal of Chemical Physics* **21** (6), 1087–1092.

Meyn, S. P. & Tweedie, R. L. 2012 *Markov Chains and Stochastic Stability*. Springer Science & Business Media, New York.

Mokhayeri, Y., Riffat, R., Takacs, I., Dold, P., Bott, C., Hinojosa, J., Bailey, W. & Murthy, S. 2008 Characterizing denitrification kinetics at cold temperature using various carbon sources in lab-scale sequencing batch reactors. *Water Science and Technology* **58** (1), 233–238.

Mukhopadhyay, N. 2000 *Probability and Statistical Inference*. Marcel Dekker, New York.

Mustakhimov, I., Kalyuzhnaya, M. G., Lidstrom, M. E. & Chistoserdova, L. 2013 Insights into denitrification in *Methylotenera mobilis* from denitrification pathway and methanol metabolism mutants. *Journal of Bacteriology* **195** (10), 2207–2211.

Neumann, M. B. & Gujer, W. 2008 Underestimation of uncertainty in statistical regression of environmental models: influence of model structure uncertainty. *Environmental Science and Technology* **42** (11), 4037–4043.

Ngo, V. V., Michel, J., Lucas, L., Latifi, A. & Simonnot, M.-O. 2015 Sensitivity, estimability and correlation of parameters describing equilibrium and nonequilibrium transports of bromide tracer in the field lysimeter. *European Journal of Environmental and Civil Engineering* **19** (4), 445–466.

Rahman, A., Riffat, R., Okogi, S., Takacs, I., Al-Omari, A. & Murthy, S. 2016a Evaluation of anoxic heterotrophic yield using multiple calculation methods. *International Journal of Environmental Research* **10** (2), 255–264.

Rahman, A., Okogi, S., Riffat, R., Al-Omari, A. & Murthy, S. 2016b Characterizing denitrification kinetics in lab scale reactors for longer time ethanol dosage. *Journal of Water and Environment Technology* **14** (5), 372–385.

Rieger, L., Koch, G., Kühni, M., Gujer, W. & Siegrist, H. 2001 The EAWAG Bio-P module for activated sludge model no. 3. *Water Research* **35** (16), 3887–3903.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M. & Tarantola, S. 2008 *Global Sensitivity Analysis: the Primer*. John Wiley, Chichester, UK.

Sathyamoorthy, S., Vogel, R. M., Chapra, S. C. & Ramsburg, C. A. 2014 Uncertainty and sensitivity analyses using GLUE when modeling inhibition and pharmaceutical cometabolism during nitrification. *Environmental Modelling & Software* **60**, 219–227.

Shannon, C. E. 2001 A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review* **5** (1), 3–55.

Sharifi, S., Murthy, S., Takács, I. & Massoudieh, A. 2014 Probabilistic parameter estimation of activated sludge processes using Markov Chain Monte Carlo. *Water Research* **50**, 254–266.

Sin, G., Van Hulle, S. W. H., De Pauw, D. J. W., van Griensven, A. & Vanrolleghem, P. A. 2005 A critical comparison of systematic calibration protocols for activated sludge models: a SWOT analysis. *Water Research* **39** (12), 2459–2474.

Sin, G., Gernaey, K. V., Neumann, M. B., van Loosdrecht, M. C. M. & Gujer, W. 2009 Uncertainty analysis in WWTP model applications: a critical discussion using an example from design. *Water Research* **43** (11), 2894–2906.

Sin, G., Gernaey, K. V., Neumann, M. B., van Loosdrecht, M. C. M. & Gujer, W. 2011 Global sensitivity analysis in wastewater treatment plant model applications: prioritizing sources of uncertainty. *Water Research* **45** (2), 639–651.

Smets, I. Y., Haegebaert, J. V., Carrette, R. & Van Impe, J. F. 2003 Linearization of the activated sludge model ASM1 for fast and reliable predictions. *Water Research* **37** (8), 1831–1851.

Stare, A., Hvala, N. & Vrečko, D. 2006 Modeling, identification, and validation of models for predictive ammonia control in a wastewater treatment plant – a case study. *ISA Transactions* **45** (2), 159–174.

Sun, P., Wang, R. & Fang, Z. 2009 Fully coupled activated sludge model (FCASM): model development. *Bioresource Technology* **100** (20), 4632–4641.

Takács, I., Patry, G. G. & Nolasco, D. 1991 A dynamic model of the clarification-thickening process. *Water Research* **25** (10), 1263–1271.

Vanrolleghem, P. A., Spanjers, H., Petersen, B., Ginestet, P. & Takacs, I. 1999 Estimating (combinations of) activated sludge model no. 1 parameters and components by respirometry. *Water Science and Technology* **39** (1), 195–214.

Walsh, S. & Whitney, P. 2012 A graphical approach to diagnosing the validity of the conditional independence assumptions of a Bayesian network given data. *Journal of Computational and Graphical Statistics* **21** (4), 961–978.

Weijers, S., Kok, J., Preisig, H., Buunen, A. & Wouda, T. 1996 Parameter identifiablity in the IAWQ model no. 1 for modelling activated sludge plants for enhanced nitrogen removal. *Computers & Chemical Engineering* **20**, S1455–S1460.

Yang, M., Sun, P., Wang, R., Han, J., Wang, J., Song, Y., Cai, J. & Tang, X. 2013 Simulation and optimization of ammonia removal at low temperature for a double channel oxidation ditch based on fully coupled activated sludge model (FCASM): a full-scale study. *Bioresource Technology* **143**, 538–548.

Zhan, C.-S., Song, X.-M., Xia, J. & Tong, C. 2013 An efficient integrated approach for global sensitivity analysis of hydrological model parameters. *Environmental Modelling & Software* **41**, 39–52.

First received 21 August 2016; accepted in revised form 14 December 2016. Available online 2 January 2017