Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/watres

Probabilistic parameter estimation of activated sludge processes using Markov Chain Monte Carlo

Soroosh Sharifi [a,1], Sudhir Murthy [b], Imre Takács [c], Arash Massoudieh [a,*]

[a] Civil Engineering, The Catholic University of America, 630 Michigan Ave NE, Washington, DC 20064, USA
[b] DC Water and Sewer Authority, 5000 Overlook Avenue, SW, Washington, DC 20032, USA
[c] Dynamita, 7 lieu-dit Eoupe, 26110 Nyons, France

ARTICLE INFO

Article history:
Received 15 August 2013
Received in revised form
23 November 2013
Accepted 5 December 2013
Available online 15 December 2013

Keywords:
ASM
Biological treatment
Bayesian
Markov Chain Monte Carlo
Uncertainty assessment

ABSTRACT

One of the most important challenges in making activated sludge models (ASMs) applicable to design problems is identifying the values of its many stoichiometric and kinetic parameters. When wastewater characteristics data from full-scale biological treatment systems are used for parameter estimation, several sources of uncertainty, including uncertainty in measured data, external forcing (e.g. influent characteristics), and model structural errors influence the value of the estimated parameters. This paper presents a Bayesian hierarchical modeling framework for the probabilistic estimation of activated sludge process parameters. The method provides the joint probability density functions (JPDFs) of stoichiometric and kinetic parameters by updating prior information regarding the parameters obtained from expert knowledge and literature. The method also provides the posterior correlations between the parameters, as well as a measure of sensitivity of the different constituents with respect to the parameters. This information can be used to design experiments to provide higher information content regarding certain parameters. The method is illustrated using the ASM1 model to describe synthetically generated data from a hypothetical biological treatment system. The results indicate that data from full-scale systems can narrow down the ranges of some parameters substantially whereas the amount of information they provide regarding other parameters is small, due to either large correlations between some of the parameters or a lack of sensitivity with respect to the parameters.

1. Introduction

Since its introduction in 1987, IWA's Activated Sludge Model 1 (ASM1) (Henze et al., 1987) and its successors have become extensively popular for the design and optimization of biological treatment systems (Gernaey et al., 2004; Sin et al., 2005). As mechanistic models, the main goal of ASMs is to predict the performance of biological treatment processes in removing organic matter and nutrients under different conditions. When applying ASMs to design or optimize biological treatment processes, it is important to recognize and quantify the uncertainties associated with the model outputs (Belia et al., 2009). The main sources of uncertainty in ASM

* Corresponding author.
E-mail addresses: S.Sharifi@bham.ac.uk (S. Sharifi), arashmassoudieh@gmail.com, massoudieh@cua.edu (A. Massoudieh).

[1] Current address: School of Civil Engineering, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK.
0043-1354/$ – see front matter
http://dx.doi.org/10.1016/j.watres.2013.12.010

modeling can be categorized into four main groups (Cierkens et al., 2012):

1) Model input data uncertainty, i.e., uncertainties associated with influent characterization or environmental factors such as temperature.
2) Uncertainty in model parameters.
3) Model structural error, due to the fact that the model is, at best, an idealization of the real process.
4) Uncertainty associated with the numerical methods used within the model (truncation errors).

Arguably, the most important challenge in making ASM models usable in practice is attributing the values of its many stoichiometric and kinetic parameters (Gernaey et al., 2004)—hereafter, for the sake of simplicity, referred to as "parameters"—which sometimes cannot be measured directly (Weijers and Vanrolleghem, 1997). Usually, when ASMs are used for practical purposes, the values of kinetic and stoichiometric parameters, such as biomass growth rates, yield coefficients, and half-saturation constants, are determined based on the values provided in the literature. The literature values are obtained through independent batch or other types of experiments under controlled conditions or by using previous model calibrations based on data from full-scale systems. Because different parameter values are suggested by different studies, a range of values for each parameter is often reported (Jeppsson, 1996). These ranges are sometime so wide that choosing different parameters within the range can result in drastically different predictions. Lab experiments under controlled conditions often require several series of measurements of constituents of interest under a range of other influencing factors, while keeping other factors constant (Amano et al, 2002). The values obtained under these conditions are not always applicable to full-scale bioreactors, due to broader heterogeneities and the interactions of larger numbers of components, including a more diverse set of chemical and bacterial species. On the other hand, when manual calibration is used to estimate ASM model parameters using data collected from full-scale systems, it is not guaranteed that the obtained set of parameters is the only parameter-set, resulting in reproduction of the observed data. This problem has been referred to as non-uniqueness, lack of identifiability, or equifinality (Beven and Freer, 2001). This is due to the fact that ASMs are generally over-parameterized with respect to the amount of data available for calibration, and because, under certain operational bioreactor conditions, the effluent characteristics can be insensitive to the values of some of the parameters (Cierkens et al., 2012).

Automatic and semi-automatic deterministic methods based on least-squares and maximum likelihood criteria (e.g., linearized maximum likelihood (Kabouris and Georgakakos, 1996a, b)) have been used to estimate the optimal values of ASM parameters using observed data. Gradient-based (e.g., generalized reduced gradient method (Afonso and da Conceição Cunha, 2002)) and heuristic search methods (e.g., Simplex techniques (Cierkens et al., 2012)) have been used extensively in the past to determine ASM parameters. Ayesa et al. (1991) used the extended Kalman filter to estimate ASM parameters as time-dependent parameters. Vanrolleghem and Keesman (1996) compared a number of nonlinear parameter estimation methods for identifying ASM parameters and suggested the use of Monte Carlo simulations. Sin et al. (2008) used a Monte Carlo-based search algorithm to estimate the ASM parameters. Cox (2004) compiled a few databases containing the values of the parameters of ASM and used a Bayesian approach to develop statistical distributions for them. Gradient-based methods are prone to getting trapped in a local optima (Abusam et al., 2001), and sometimes, parameter values that may not be physically interpretable end up being found (Weijers and Vanrolleghem, 1997). In addition, while deterministic methods might provide a parameter set that maximizes the chance of reproducing the observed data, they are incapable of providing any reliability measure for the estimated parameter values.

The complexity involved in the calibration of ASMs has led to a number of protocols and guidelines for manual systematic calibration of full-scale ASM systems. BIOMATH (Petersen et al., 2002; 2003; Vanrolleghem et al., 2003), STOWA (Hulsbeek et al., 2002), HSG (Langergraber et al., 2004), WERF (Melcer et al., 2003), and IWA's STR (Rieger et al., 2012) are among the most well-known protocols. These all consist of four main steps: 1) characterizing influent wastewater; 2) constructing dynamic influent loading data; 3) manual parameter estimation; and 4) model validation. A critical comparison of these methods can be found in (Sin et al., 2005).

Almost all of the approaches used for the automatic calibration of ASMs have been deterministic so far with the exception of the work of Juznic et al. (2001), who applied Bayesian inference to estimate parameter uncertainty associated with a revised version of ASM3 and showed its advantage over some deterministic linear theory methods. In deterministic parameter estimation approaches one set of parameter values often as a results minimization between some measures of misfit between the modeled and measured results is obtained using a manual or automated optimization technique. However, many sources of uncertainty and error, including observation error, model structural error, errors associated with input variables and external forcing, and possible non-uniqueness of optimum parameters or lack of sensitivity of the predicted effluent concentrations to certain parameters under some conditions, are inevitably propagated into the estimated parameters and need to be quantified. Deterministic parameter estimation approaches provide a single set of parameters, and it is not clear how much deviation from those estimated values is still acceptable and what the shape of the region of plausibility in the parameter space looks like. Regardless of what calibration method is used, parameter uncertainty is always present and eventually transmits into model output uncertainty (Morgan and Henrion, 1992). Using a single set of parameters to obtain some model outputs could result in the sub-optimal design of biological treatment systems, incorrect planning decisions, and poor effluent water quality. Therefore, to use the ASMs effectively for optimization of the operation and design of biological treatment systems, it is necessary that the effects of these uncertainties on the uncertainties of parameter estimation to be quantified.

Various uncertainty quantification approaches have been used in various scientific fields. These include local sensitivity