Activated Sludge Model 2d calibration with full-scale WWTP data: comparing model parameter identifiability with influent and operational uncertainty.

Vinicius Cunha Machado, Javier Lafuente, Juan Antonio Baeza*

Department of Chemical Engineering, Universitat Autònoma de Barcelona, ETSE, 08193 Bellaterra (Barcelona), Spain. Phone: +34935811587. FAX: +34935812013.

E-mails: vinicius.cunhamachado@gmail.com, javier.lafuente@uab.cat, juanantonio.baeza@uab.cat

*Corresponding Author

Post-print of: Cunha, V; Lafuente, FJ and Baeza, JA "Activated sludge model 2d calibration with full-scale WWTP data: comparint model parameter identifialibility with influent and operational uncertainty" in Bioprocess and biosystems engineering (Ed. Springer), vol. 37, issue 7 (July 2014), p. 1271-1287. The final versión is available at DOI 10.1007/s00449-013-1099-8

Abstract

The present work developed a model for the description of a full-scale WWTP (Manresa, Catalonia, Spain) for further plant upgrades based on the systematic parameter calibration of the ASM2d model using a methodology based on the Fisher Information Matrix (FIM). The influent was characterized for the application of the ASM2d and the confidence interval of the calibrated parameters was also assessed. No expert knowledge was necessary for model calibration and a huge available plant database was converted into more useful information. The effect of the influent and operating variables on the model fit was also studied using these variables as calibrating parameters and keeping the ASM2d kinetic and stoichiometric parameters, which traditionally are the calibration parameters, at their default values. Such an "inversion" of the traditional way of model fitting allowed evaluating the sensitivity of the main model outputs regarding to the influent and to the operating variables changes. This new approach is able to evaluate the capacity of the operational variables used by the WWTP feedback control loops to overcome external disturbances in the influent and kinetic/stoichiometric model parameters uncertainties. In addition, the study of the influence of operating variables on the model outputs provides useful information to select input and output variables in decentralized control structures.

Keywords: ASM2d, EBPR, FIM, full-scale WWTP, calibration, influent characterization, modelling.

Nomenclature

A^2/O	Anaerobic, Anoxic and Aerobic (WWTP configuration)
ASM	Activated Sludge Models
BOD5	Biological Oxygen Demand (5 days)
CCF	Calibration Cost Function
COD	Chemical Oxygen Demand
DO	Dissolved Oxygen
EBPR	Enhanced Biological Phosphorus Removal
FIM	Fisher Information Matrix
GAO	Glycogen Accumulating Organisms
IWA	International Water Association
PAO	Phosphorus Accumulating Organisms
PCCF	Preliminary Calibration Cost Function
PID	Proportional-Integral-Derivative controller
SRT	Sludge Retention Time
TKN	Total Kjeldahl Nitrogen
TN	Total Nitrogen
TSS	Total Suspended Solids
VCF	Validation Cost Function
WWTP	Wastewater Treatment Plant
WERF	Water Environment Research Foundation

1. Introduction

Modelling wastewater treatment plants (WWTP) is the fundamental stone to improve WWTP performance through identifying bottlenecks and proposing modifications of existent plants or to design a completely new one. Besides the experimental knowledge, mathematical models are a set of tools for predicting plant behaviour under different conditions from the ordinary outlook of the WWTP or under unexpected operational scenarios [1]. The models are also useful for changing process concepts and developing new plant configurations [2]. The operation of WWTPs is based on the behaviour of different microorganisms, which are responsible for biological nutrient (nitrogen and phosphorus) and organic matter (carbon) removal. Such processes are well described by the IWA models ASM1, ASM2, ASM2d and ASM3, even though other models have been used and accepted in practical and scientific media as the TUD-P model [3–5] or the ASM3 EAWAG Bio-P [6]. ASM2d model is being used in many researches concerning WWTP due to including the most important biological processes of ordinary heterotrophic biomass, heterotrophic PAO biomass and nitrifiers. Ferrer et al. [7] used this model to fit full-scale WWTP data and then to evaluate different configurations for improving nutrient removal. Ingildsen et al. [8] calibrated the ASM2d model for the Avedøre WWTP (Denmark) to support a control strategy for maintaining the enhanced biological phosphorus removal (EBPR) process activated for long periods. Xie et al. [9] also used ASM2d to simulate and optimize a full-scale Carrousel WWTP. García-Usach et al. [10] or Machado et al. [11] successfully used ASM2d for describing EBPR process at pilot scale.

WWTP models are also useful for studying and proposing several control strategies in order to guarantee the effluent quality with or without external disturbances (storm events, peaks of pollutants in the influent…). The effluent quality is the main goal of the control structures, where ammonium, nitrate and phosphorus are the main pollutants that should be kept at lower values to avoid the eutrophication effect. Nevertheless, dissolved oxygen (DO) and the sludge residence time (SRT) are the inventory variables that should be controlled first [12]. To control ammonium concentration, a cascade controller which calculates the DO setpoint in the aerobic basin using the error between the desired ammonium concentration and the real measurement in the effluent is designed [13]. An ammonium feedback-feedforward controller also could be implemented if the ammonium influent load is estimated or measured [14]. Nitrate removal is accomplished by the denitrification processes which depend on the readily organic matter available in an anoxic zone and the nitrate concentration. Two ways of controlling the nitrate concentration at the effluent is adding external carbon source and changing the nitrate recycle from the aerobic basin to the anoxic one in most of WWTP [15, 16]. It is worth noticing that the measured and the manipulated variables also have uncertainties, like recycling flow measurements and dissolved oxygen concentrations. All the abovementioned control applications using WWTP models should be preceded by a correlation analysis of the available manipulated variables not to add internal disturbances to control the effluent quality.

Despite all the cited models are essential tools for improving many aspects of the wastewater treatment, they are structured on kinetic and stoichiometric parameters that should be identified for better accuracy. Besides, their state variables are not exactly the same as the information obtained from laboratory analysis periodically performed in the WWTP. Therefore, it is necessary, first, to convert some daily plant measurements of the influent into model states and, second, to calibrate parameters using plant data and lab assays (batch tests with the plant biomass). In the literature it is possible to find a methodology to accomplish the first task before mentioned [17], although the influent identifiability linked to its variability has not received much attention. The parameter calibration could be performed using protocols reported in the literature [18, 19] as the protocols developed by STOWA [20], BIOMATH [21], WERF [22], HSG [23] or Mannina et al. [24, 25]. All these protocols are good at posing well the goals of the calibration, systematically treat the plant data gathered and have a validation step with different data from those used to calibrate the model. On the other hand, only BIOMATH, WERF or Mannina et al. protocols pay attention to the parameter subset selection to maximize the information mined from the plant data. Machado et al. [11], developed an alternative calibration methodology, called the “seeds methodology”, using criteria derived from the Fisher Information Matrix (FIM) to avoid overfitting. Although the hydraulics modelling and a detailed biomass characterization are not emphasized in this last method as in the HSG and BIOMATH protocols, respectively, the usage of a large amount of available plant data combined with a systematic procedure to find the most identifiable parameters subset, without testing all the possible parameters combinations, are the strengths of the “seeds methodology”. Unfortunately, the performance of all the abovementioned protocols is affected by uncertainties from different sources during the modelling task. Refsgaard et al. [26] pointed out that several error sources affect the quality of model simulation results: (i) context and framing; (ii) input uncertainty; (iii) model structure uncertainty; (iv) parameter uncertainty and (v) model technical uncertainty. Sin et al. [27] deepened in the uncertainty analysis, concluding that both biokinetic/stoichiometric/influent fractionation related parameters as well as hydraulics/mass transfer related parameters induced significant uncertainty in the predicted performance of WWTP. Moreover, Cierkens et al. [28] studied the effect of the influent data frequency on the calibration quality and output uncertainty of the WWTP model fit.

Uncertainty assessment of kinetic and stoichiometric model parameters of ASM1 and ASM2 has been applied for full-scale WWTP as in Mannina et al. [29], who evaluated the model reliability identifying the crucial aspects where higher uncertainty rely and more efforts should be provided in terms of both data gathering and modelling practises. The uncertainty associated to operation and design parameters of WWTP have also been studied [30] showing that they are the most sensitive parameters for some benchmarking studies. Finally, Belia et al. [31] pointed out that identifying and quantifying the uncertainties involved in a new design or plant upgrade becomes crucial because WWTP are required to operate with increased energy efficiency and close to their limits. They also note the need for the development of a protocol to include uncertainty evaluations in model-based design and optimisation projects.

To consider some kind of those commented uncertainties on the modelling task and concentrating effort at the calibration step, the present work developed an AS model for the Manresa WWTP (Manresa, Catalonia, Spain) based on the systematic parameter calibration of the ASM2d model using the “seeds methodology” for further plant upgrades, as the insertion of the EBPR and the design of a new control structure for the plant. The influent was characterized as required by the ASM2d and the parameters were selected, calibrated and their confidence intervals were assessed as stated in the “seeds methodology”. The calibration parameters were divided into three groups: the traditional kinetic/stoichiometric parameter group (K group); the influent factors representing errors/uncertainties of the influent characterization (I group) and the operational variable factors (O group), considering errors/uncertainties on the measurement of the operational variables. The procedure assessed, in addition to the conventional calibration of the K group, influent and operational variables uncertainties in two additional calibrations: i) influent vector of states in the ASM2d model (I group) was used as calibration parameters while parameters of K and O groups were kept at their default values and ii) the O group was used as calibrating parameters while K and I groups kept constant. Such “inversion” of the traditional model fit procedure allowed to evaluate the quality of the influent characterization and to observe the set of operating variables which has the less number of uncorrelated variables amongst themselves for better designing a decentralized control structure for the WWTP.

2. Material and Methods

2.1 Brief description of the Manresa WWTP

The average flow rate of Manresa WWTP is 27,000 m^3/d. This WWTP (Figure 1) consists of a pre-treatment (gross and grit removal), primary treatment with a clarifier, a secondary stage (biological removal) and a possible tertiary stage (chlorination). There are two main treatment lines in the secondary stage (Figure 2). Each line has three anoxic reactors (1460 m^3) and one aerobic reactor made up by two parts of 3391 m^3. Each reactor has approximately 7 m of depth. After passing through the anoxic zone, the bulk liquid is mixed and is again divided to feed the aerobic zone. Air is bubbled from the bottom of the aerobic tanks with membrane diffusers, allowing biological oxidation of the organic matter and ammonium. An internal recycle pipe connects the aerobic zone to the anoxic one in order to bring the nitrate to be denitrified in the anoxic zone. At the end of the secondary stage two settlers separate the biomass from the treated effluent. Settled biomass returns to the entrance of the anoxic reactor by an Archimedes screw. The excess of sludge is anaerobically digested and sent to a composting plant. The effluent, after leaving the secondary settler, can be chlorinated and it is disposed to the environment at the Cardener River.

It is worth noticing that experimentally is observed preferential flux of the inlet mass stream to one of the main treatment lines. The presence of DO (0.5-1.0 mg/L) at the end of the anoxic reactors indicates that the denitrification is not occurring at the maximum intensity because there is a lack of organic matter to improve the nitrate reduction and a poor mixing is taking place. Also, a non-homogeneous spatial distribution of DO was observed along the aerobic reactors, not only along the influent path but also in depth.

Daily analyses of COD, BOD5, total suspended solids (TSS), NH_4^+, NO_3^-, PO_4^{3-}, total Kjeldahl nitrogen (TKN) and total nitrogen (TN) are performed at the influent and the effluent of the secondary treatment. The daily composite samples are collected from the full-scale WWTP by sampling every 2 hours. The only system variable measured in each reactor of the secondary treatment is the TSS concentration. .

The air supply system is composed by 4 air blowers with 100,000 Nm^3/d of capacity, whose motor speed are controlled by a single DO feedback controller in the aerobic basins. The aerobic zone of each water line has two on-line DO sensors, one of them placed at the 25% of the path along the zone and the other one placed at 75% of the aerobic zone. The DO PI controller uses a weighted average of the four DO concentrations as the measured variable, and compares it to a DO setpoint, usually equal to 2.0 mg/L. Once computed the error between the setpoint and the averaged DO, the new setpoint speed of the blowers is calculated by the PI algorithm and sent to the devices. Physically, the air is moved to a primary header after being discharged by the blowers. Then, the air flow rate is divided into two branches. The right branch feeds the middle part of the two aerobic zones while the left branch feeds the entrance and the end of the two aerobic zones.

The main operation costs are electrical energy for aeration and pumping, sludge treatment (anaerobic digestion and composting) and chemical products for P precipitation.

2.2 Influent composition and patterns

Influent composition and its variability is key information for plant modelling and description of changes along the year due to seasonal patterns. Table 1 shows influent properties (averages) straightforward linked to the wastewater composition in winter and summer months for the Manresa WWTP. Considering the effluent limits of COD (125 mg O_2/L), BOD5 (25 mg O_2/L), total N (10 mg/L), ammonium (4 mg/L) and total P (1 mg/L), defined by the local water agency (Agència Catalana de l'Aigua, ACA), the Manresa WWTP, with average effluent flow rate of 27,000 m^3/day, could deliver an effluent load of 3375 kg/d, 675 kg/d, 270 kg/d, 108 kg/d and 27 kg/d, respectively for these pollutants. The total P discharge load was kept at the limit of 27 kg/d, which means an average value of 1 mg/L of P, with large usage of $FeCl_3$ in 2008, 2009 and 2010. Such chemical precipitation represents a cost around 50,000 €/y, but allows meeting the legal discharge level of the EC directive.

On summer months, contaminant loads are considerably lower than in winter months, probably also due to the people moves from Manresa to vacation locations. These qualitatively recognized patterns can be mathematically analysed looking for daily, weekly or monthly profiles that could help to improve the tuning of feed-forward controllers, for refusing external variations whose pure feedback controllers do not deal easily, as well as to promote a time-scheduling load profile for dosing extra COD source for denitrification and $FeCl_3$ for chemical P removal.

2.3 Model structure

The kinetic model implemented for modelling COD, N and P removal was the IWA ASM2d model [5]. It has 19 state variables and 21 processes, which include nitrification and denitrification and the PHA (poly-hydroxyalkanoates) accumulation process, the latter fundamental for EBPR.

The settler model adopted was the 10 layer Takács model [32]. The wastewater entrance is at the fifth layer. At the end of the process, the effluent leaves the settler from the upper part (the collector, layer 1) and the settled biomass is recycled from the bottom of the settler (layer 10) to the feed of the biological treatment. The recycled biomass (external recycle, Q_{RAS}) is reincorporated to the process, being mixed to new influent of the biological treatment. The soluble components of the wastewater leave the settler with a concentration calculated considering CSTR behaviour for these compounds. The settleability of particulate states is linked to the settling velocity which is calculated by a double exponential function (Equation 1).

$$v_s = v_0 \cdot e^{-r_h(X_i - f_{ns} \cdot X_{IN})} - v_0 \cdot e^{-r_p(X_i - f_{ns} \cdot X_{IN})} \qquad \text{Eq. 1}$$

Where:

v_0 is the settling velocity if the *Stokes' Law* could be applied to the wastewater, [m/h];

f_{ns} is the fraction of non-settleable solids;

X_{IN} is the inlet solid concentration, [g TSS/m^3];

X_i is the solid concentration of the layer *i*, [g TSS/m^3];

r_h and r_p are weights for modelling the effect of the size of the particles in the settling velocity.

Parameter v_s is compared to a maximum settling velocity, $v_{s,max}$, which is experimentally determined. *Xt* is another model parameter required as a threshold value that indicates an upper limit in the settler capacity to prevent an overflow of solids in the equipment. The default values of the adopted model were:

v_0:	500 m/h	r_p:	$2.86 \cdot 10^{-3}$	f_{ns}:	$2.28\ 10^{-3}$
v_{smax}:	250 m/h	r_h:	$5.76 \cdot 10^{-4}$	X_t:	3000 g TSS/m^3

2.4 Influent characterization according to the model states

Although daily analysis of the influent is performed as detailed in section 2.1, additional experimental data was needed to obtain the specific characterization required for ASM2d. Therefore, some experiments were performed with wastewater leaving the primary clarifier following the methodology described by Orhon et al. [33] as detailed in Montpart [34]. The determined influent stream characteristics were S_I = 0.080 COD, X_I = 0.055 COD, X_S = 0.450 COD and S_F = 0.410 COD and these ratios were assumed constant. See supplementary information S1 for details of this characterization.

The values of the influent variables X_{TSS}, S_{NH4}, S_{NO3}, S_{PO4} were assumed to be equal to the experimental observations (analysis of daily composite samples). The variables S_A, X_{PHA}, X_{PAO}, X_{PP}, S_{N2}, S_{O2}, X_A, X_{MEP} were assumed to be zero. Hence, the inlet heterotrophic biomass was calculated by Equation 2:

$$X_H = COD - (S_I + S_A + S_F + X_I + X_S + X_A) \qquad \text{Eq. 2}$$

The variable X_{MEOH} was not considered zero due to the presence of chemical phosphorus precipitant agent and its value along the time was defined in the steady state calibration, when the phosphorus behaviour in the effluent was evaluated. Finally, S_{ALK} (the plant influent alkalinity) was assumed to be 7 moles of HCO_3^-/m^3.

3. Results and Discussion

3.1 Preliminary steady-state calibration

Model calibration was performed in two steps: a steady-state calibration and a dynamic calibration. The former step was useful to minimize structural discrepancies between the plant model and plant data. By its turn, the dynamic calibration involved not only the determination of kinetic and stoichiometric parameters, but also an estimative of the useful volumes of reactors and settlers and the necessities of P chemical precipitant agent and extra load of biodegradable COD required for denitrification. Figure 3 shows a simplified scheme of the overall calibration / validation process used in this work.

Preliminary calibration aims to reduce structural discrepancies between the model and the experimental variables, especially to reduce the main differences between experimental TSS and TSS model predictions. Experimental data were averaged (influent values and operational parameters like DO and flow rates) and the resultant values were used as inputs to the simulation model (constant inputs). A period of 1200 days was simulated with the default ASM2d parameters and the steady-state values were used as initial values for all the simulations performed afterwards. TSS concentrations in the effluent and in the wastage purge stream were used as output variables to calibrate the following parameters:

a) r_p and f_{ns} (settling model parameters), to decrease the differences between TSS in the effluent and the model predictions for this output.

b) f_{Qw} and f_{QRAS}, in order to adjust the model TSS in the effluent and in the purge (and consequently in the solids inside the aerobic reactors).

A preliminary calibration cost function (PCCF, Equation 3) was employed to perform the preliminary steady-state calibration of the WWTP model.

$$PCCF = \sum_{k=1}^{2} q_k \sqrt{\cdot \sum_{r=1}^{m} \left(y_{Exp\,k,r} - y_{Model\,k,r} \right)^2} \qquad \text{Eq. 3}$$

Where:

- k is related to each output variable
- r is related to each experimental data (each day). The whole period studied had m = 1200 days.
- q_k is the weight to normalize the output variables since their values are very different. Ammonium was used as the reference value for the normalization, and hence the weights were calculated as the ratio of the average of ammonium concentration at the effluent to the average of the other output variable (TSS in the effluent and in the external recycle) as shown in equation 4. The weights used for the TSS in the effluent and in the external recycle were, respectively, $3.637\ 10^{-4}$ and $2.404\ 10^{-4}$.
- $y_{Exp\ k,r}$ is the experimental data of variable k at day r.
- $y_{Model\ k,r}$ is the model output of variable k at day r.

$$q_k = \frac{\frac{1}{m}\sum_{r=1}^{m} y_{NH_4,r}}{\frac{1}{m}\sum_{r=1}^{m} y_{k,r}} \qquad \text{Eq. 4}$$

Where $y_{k,r}$ is the data of the other output variables (i = X_{TSS} at the effluent and X_{TSS} at the purge) and m is the total number of experimental data (m = 1200).

In addition, X_{MeOH} in the influent was manipulated to adjust the phosphate concentrations in the effluent. The calibrated values of the parameters were: $r_P = 1.036\cdot 10^{-2}$, $f_{ns} = 2.566\cdot 10^{-3}$, $f_{Qw} = 0.1736$, $f_{QRAS} = 1.911$ and $f_{XMeOH} = 1.237$. These calibrated parameters were considered constant and were maintained in these values during the dynamic calibration procedure. The values of the calibrated parameters f_{Qw} and f_{QRAS} were also used as initial guesses in the dynamic calibration of the Operational Variables group

3.2 Development of the cost function for dynamic calibration

Data from seven effluent variables were available for model calibration of Manresa WWTP: ammonium, nitrate, phosphorus, TSS, COD, BOD5 and TKN. These variables were considered the output variables of interest. Data period used for model calibration was from October 2007 to May 2008. Due to its daily oscillation, COD and BOD5 were used only for model validation. In the dynamic calibration step, equation 5 was used as cost function.

$$CCF = \sum_{i=1}^{5} w_i \sqrt{\cdot \sum_{j=1}^{n} \left(y_{Exp\,i,j} - y_{Model\,i,j}\right)^2} \qquad \text{Eq. 5}$$

Where:

- i is related to each output variable
- j is related to each experimental data (each day). The whole period studied had n = 251 days.
- w_i is like q_k a weight to normalize all the output variables, which have different units and values, using ammonium (w = 1) as a common base. Hence, the weights were calculated as the ratio of the average of ammonium concentration at the effluent to the average of the other output variable (NO_3^-, PO_4^{3-}, TSS and TKN) as shown in equation 6. The weights calculated for nitrate, phosphorus, TSS and TKN were 0.235, 1.124, 0.091 and 0.532, respectively.
- $y_{Exp\,i,j}$ is the experimental data of variable i at day j.
- $y_{Model\,i,j}$ is the model output of variable i at day j.

$$w_i = \frac{\frac{1}{n}\sum_{j=1}^{n} y_{NH_4,j}}{\frac{1}{n}\sum_{j=1}^{n} y_{i,j}} \quad \text{Eq. 6}$$

Where $y_{i,j}$ is the data of the other output variables (i = NO_3^-, X_{TSS}, N_{TKN} or PO_4^{3-}) and n is the total number of experimental data (n = 251).

The CCF value calculated with the original model prediction (with default parameters) was 83.46, but after the preliminary calibration step (optimization of PCCF) it was reduced to 67.68 (18.9% improvement).

The validation cost function (VCF) was calculated also with equation 5, but using experimental results of years 2008 to 2010.

Due to the associated uncertainty of full-scale WWTP, operational variables, as the plant flow rates and the DO in the aerobic basins could also be used as parameters to calibrate. The internal recycle, external recycle and purge flow rates data observed by the WWTP personnel probably contain uncertainties (no reliable flowmeters are usually available) and hence, some multiplying factors were created to consider these uncertainties. These factors were f_{QW} for the purge flow rate, f_{QRINT} for the internal recycle flow rate and f_{QRAS} for the external flow rate. In the case of the uncertainties of the DO sensors, the multiplying factor was the *DO_Gain*.

As the influent concentrations of each model variable neither are perfectly determined, additional influent factors were adopted for further adjustments in the inlet concentration of these variables.

3.3 Parameter grouping for dynamic calibration

The full plant model has about 90 model parameters, but only the 24 most sensitive parameters were studied. This set of 24 parameters was divided into three subsets: the kinetic/stoichiometric parameters (group K, with 10 parameters), the influent parameters (group I, with 10 parameters) and the operational parameters (group O, with 4 parameters). In fact, only parameters of the kinetic/stoichiometric macro-group were used for real model calibration. The macro-groups I and O were used to obtain additional information for process control and data quality.

The subset of kinetic/stoichiometric parameters was made up of the growth and decay parameters, yields and saturation constants of all the involved biomasses (autotrophic, heterotrophic and PAO). When calibrating the model with this group, it was assumed that the influent composition during all the calibration period was completely known, as well as the operational parameters. This assumption was not strictly correct since on-line measurements of all the ASM2d states are never available. On the other hand, using the subset of influent parameters, it was assumed that all the default kinetic/stoichiometric ASM2d parameters were perfectly correct, as well as the operational parameters. As determining on-line all the ASM2d variables in the influent stream would be a very difficult and expensive task, the group I calibration was used for obtaining additional information about the influent data quality and to determine which variables in the influent could be easily modified in order to adjust the model. At last, using the group of operational parameters, both kinetic/stoichiometric parameters and the influent composition were considered perfectly fitting the biological processes rates and the incoming pollutant loads, respectively. Amongst all the parameters, group O was used for process control in the normal plant operation. Therefore, it was determined the parameters of this group that more easily provided fast plant response to reject external disturbances to the control system. This knowledge was obtained using the same calibration methodology of the group K to the group O.

3.4 Sensitivity Analysis

Table 2 presents the overall sensitivity, calculated as the sum of relative sensitivity for ammonium, phosphate, nitrate, TKN and TSS in the effluent (see the equations used in the supplementary information S2). The advantage of using the relative sensitivity for calculating the overall sensitivity is that all the output variables have the same importance.

The parameters of each macro-group that most affect the model outputs were ranked in descending order in this table. In the case of the K group, the heterotrophic biomass growth yield, the nitrification and the phosphorus chemical precipitation are well represented by the ranked parameters. K_{PRE} and K_{RED} have almost the same impact on the model outputs, but their impacts are less important than the N removal processes.

Regarding the influent group, the inlet X_S, P-related processes and the inlet ammonium concentration were the most important calibrating parameters. It is observed that PO_4^{3-} or MeOH inlet concentrations are more important that the own kinetic precipitation parameters K_{PRE} and K_{RED} of K group. These results indicate that chemical P-precipitation and P-redissolution processes are kinetically limited due to the low phosphate and MeOH concentration in the biological reactors.

In the case of the operational parameters, the purge flow rate and the DO have the most influence on the model outputs. Nevertheless, all the parameters of this group would have to change considerably to affect the outputs in the same quantity as the kinetic/stoichiometric or the influent parameters. Table 2 also shows that inlet MeOH concentration, which could be used to control P chemical precipitation, produces more impact on the outputs that the process control variables considered in group O. Regarding S_F inlet concentration, which could be used for controlling denitrification, it would affect the outputs in the same extent of the best parameter of the group O, the purge flow rate.

The previous sensitivity analysis was used to select the possible calibration parameters for applying the "seeds" methodology. The K group has 10 elements that most affect the model outputs. No more kinetic or stoichiometric parameters were included since the 10th parameter of the sensitivity list (Table 2) of this group ($\eta_{NO3,D}$) only affects the model output less than 10% the 1st parameter. The I group has all the influent states that commonly could affect the model output. It is important to remember that this group could be of size 19, the 19 state variables of the ASM2d, but the results of Table 2 show that only the first nine affected the outputs. Finally, the O group has all the 4 variables that are commonly used to control de WWTP processes. In case of adding external readily organic matter to improve denitrification or phosphorus removal, such group of parameters would have size of 5.

3.5 Dynamic calibration methodology

Dynamic calibration was performed following the methodology of the "*seeds*" [11] and starting from the results obtained by the preliminary calibration and the sensitivity analysis. This is the first reported application of this method using full-scale plant data. The procedure uses the RDE criteria calculated from the Fisher Information Matrix (FIM) as the ratio of normalized D to modified E criteria (RDE). From the sensitivity ranking, the best-ranked parameters are named as ‘‘seeds’’, since each one serves for growing a parameter subset for model calibration. The subset generation process adds to the seed subset a parameter that presents the highest RDE among the combination between the current seed subset and all the other remaining best-ranked parameters of the sensitivity rank. The process of generation of parameter subsets is automated, independent of the user and exclusively based on mathematical tools, which was considered a necessary improvement of model calibration techniques pointed out by Sin et al. [18]. The “seed” methodology allows generating subsets with the maximum capacity to explain plant behaviour with the less possible correlation amongst its parameters. All the subsets generated could systematically be compared with each other. The process of parameter addition repeats until the RDE decreases from the current iteration to the previous one, for each seed. After that, the subset with the highest RDE criterion is elected and the parameters values are already changed to the calibrated values during the “seed” growth.

3.6 Dynamic calibration results

Tables 3, 4 and 5 present the results of applying the abovementioned calibration methodology using parameters of groups K, I and O, respectively.

3.6.1 Calibration of kinetic parameters

The 10 best subsets were selected from the tested seeds. See Table 3 for details. The common subset size is of 4 parameters. Nevertheless the highest RDE value was calculated for a subset of 6 parameters (subset of seed $\eta_{NO3,D}$). This subset produces the lowest CCF and VCF, resulting in the most suitable subset for model calibration even though the confidence interval of one parameter is considerably high. As the current plant is an A/O WWTP, no parameters related to the biological P-removal appear in the 10 most impacting seeds. On the other hand, in all the subsets appears K_{PRE} or K_{RED}, parameters linked to the P-chemical precipitation. Y_H and b_H are present in all the subsets, with high values of parameter confidence interval, which indicate less reliable calibrating values. Parameter $\eta_{NO3,D}$ is the parameter that provides more information about the plant behaviour (lowest CCF and VCF when this parameter is inside the calibration set), despite its lower value (0.0296) and more than 50% of confidence interval (default ASM2d value is 0.80). Such value indicates that a poor denitrification process is occurring in the plant caused by a lack of carbon source and some amount of DO transported from the aerobic zone to the anoxic one. It would be recommendable to add extra carbon source to the influent stream to increase the efficiency of the nitrogen removal processes.

Considering that the influent composition determined by lab test and using plant data is perfectly known along the years of calibration and validation data, the best subset obtained following the methodology of Machado et al. [11] amongst the kinetic group is the subset obtained from the parameter $\eta_{NO3,D}$. The full subset is composed by the parameters $\{\eta_{NO3,D}, K_{PRE}, b_A, Y_H, K_{O2,A}, b_H\}$ with values [0.0296, 1.005, 0.2203, 0.4181, 0.1130, 0.0829]. See Figure 4 for comparisons between the model prediction and the plant data. In this subset, a calibrated value of 0.4181 for Y_H means that more COD is consumed for maintenance of the heterotrophic biomass than the consumed for promoting the growth of the microorganisms. It was not expected this low value for this parameter, since the default value of Y_H is 0.625 [5]. However, similar values for Y_H around 0.45 were obtained in the other subsets from the rest of seeds. Such an unexpected result, probably, is derived from a lack of knowledge on the influent composition and from the optimized values for sedimentation parameters obtained in the static calibration. Nevertheless, $\eta_{NO3,D}$ subset showed the best compromise between explaining the plant behaviour and avoiding parameters correlations, with lower CCF and VCF values.

Gross modelling errors could be corrected in the preliminary calibration step. Nevertheless, poor BOD5 and ammonium predictions in the effluent could be an indication that a false denitrification rate is occurring, probably because a lack of easily biodegradable COD is not being captured. Figure 5 compares the model predictions to the validation data, which is a completely different dataset from the calibration data. In Figure 5, the parameters subset of the best seed of Table 3 makes the model suitable for predicting correctly nitrate, phosphate, solids, TKN and COD in the effluent stream and the solids in Q_{RAS} stream and inside the basins. The model predicts a very low ammonium and BOD5 concentration in the effluent. Such results also could indicate dead volumes in aerobic basins not modelled as well as a spatial gradient of DO, ignored in the current model. As a consequence, not all the regions of the aerobic basins operate with a reasonable DO concentration (2-3 mg/L). Figures 4 and 5 clearly show that events with fast dynamics are not well captured, since some plant measurements that made up calibration and validation data subsets have their sample time equal to one day and the samples are integrated (each 2 hours a volume of wastewater is hold to compose a final sample before chemical and biochemical analysis). Besides, the plant data presents abrupt changes which bring additional difficulty to estimate model parameter errors.

3.6.2 Calibration of influent parameters

Although the parameters of the influent group would not be used to make a real fit of the model as in a conventional calibration procedure, some useful information can be extracted from these results (Table 4). The optimized values of parameters are factors that multiply the influent vectors for each variable of the influent. Therefore, a value of 1.414 of f_{SNH4} of the f_{SI} seed means that the ammonium vector of original plant data increased 41.4% in order to minimize the cost function.

The most common subset size is 5 or 6 parameters. Parameters f_{SALK}, f_{XMeOH}, f_{SNH4} and f_{SPO4} are present in almost all the subsets, which indicate that each variable is explaining the model and is not interdependent amongst all of them. This information is also useful to decide the influent variables where the sampling and measuring efforts should be focused for a reliable optimization of kinetic parameters.

Table 4 also brings some other relevant remarks. The influent parameter group could achieve good values of CCF and VCF in most of the tested subsets compared to the subsets of the kinetic group. Thereby, if the weight of the influent parameter group (new approach) is stronger than the kinetic one (traditional way) on the model prediction, the variability of the influent composition and the error concerned to the characterization procedure could explain the deviation between the current model and the standard model ASM2d predictions. Therefore, these results demonstrate that the confidence of the influent characterization is a key factor to consider before fitting any parameter of a given model. In this sense, the importance of uncertainties associated to the influent characterization that induce significant uncertainty in the model predictions have been already highlighted in the literature [27, 28].

Comparing the results of f_{XTSS} and f_{XS} seeds it is observed that the result of f_{XTSS} seed explains better the outputs than the result of f_{XS} seed, although the inclusion of f_{SF} in the former subset increases correlation among parameters. In addition, the calibrating methodology did not allow the simultaneous presence of f_{XS} and f_{XTSS} in any calibration subset, probably due to the high correlation between these variables.

Finally, nitrate data are correlated to the S_F data, since in both created subsets where f_{SNO3} appears (seeds f_{SNO3} and f_{SI}), high parameter confidence interval values are reported. The existence of such correlation is clearly realized in the subset created by the f_{SNO3} seed, which is made up only by f_{SNO3} and f_{SF}.

3.6.3 Calibration of operational variables

Considering the operational variables, only two different subsets could be created (see Table 5), which means that almost all the variables help to explain the experimental observations without correlation. Nevertheless, when inserting the biomass recycle flow rate (f_{QRAS}) into a parameter calibration subset, a strong correlation to the internal recycle flow rate was added. It indicates that in a possible control structure for controlling simultaneously N, P and COD removal, the biomass recycle flow rate and the internal recycle flow rate could not be changed at the same time or their modifications should be done in different magnitudes to avoid its interaction.

Table 5 also shows that operational variables could improve model fit, i.e., the observed variability with respect ASM2d prediction with default parameters could be explained considering that the operational variables were not well measured. This is an important problem in any model fit using full-scale WWTP data, where there are gradients and time variability of operational variables, which do not have the same homogeneity and reliability than in a controlled pilot WWTP.

3.7 Remarks

The "seeds" methodology applied to different group of parameters, not only the traditional kinetic and stoichiometric ones, is a novel approach and allows:

- To automate the parameter subset selection, an improvement in the model calibration techniques, pointed out by Sin et al. [18]. The usage of the sensitivity analysis is similar to that found in BIOMATH protocol [21]. The "seed" methodology searches for the minimal number of parameters that explains the plant data with the less possible correlation amongst the calibration parameters.

The utilization of a higher number of parameters as in other works [24, 36] provides a good model fit, but it is not usually supported by a study of its correlation, which weakens its mathematical validity, as it is likely disregarding overfitting problems that could reduce the model predictive capacity.

- To measure, in some extent, the influent states with higher uncertainties, which aid to concentrate efforts in programming specific experiments to better characterize these input variables (load disturbances). Such an uncertainty measurement is in agreement to the philosophy of BIOMATH [21], STOWA [20] and WERF [22] protocols, which are supported, amongst other premises, on an excellent influent characterization.
- To identify the most correlated operational variables not to add them together inside a control structure with decentralized controllers (e.g. PID controllers), to avoid internal conflicts with the different control loops. Also, observing the CCF and the confidence intervals of the best subsets of K and O groups, it is possible to infer if some control structure designed based on the group O will be able to compensate kinetic/stoichiometric uncertainties, since the industrial controllers are model-based controllers, which means that the controllers performance are dependent of the model accuracy. In the studied case, the operational variables of Manresa WWTP are able to keep the plant under a stable operating point since the CCF of subsets of the O group are lower than the K group as well as the confidence intervals.

4. Conclusions

The ASM2d model was calibrated for the Manresa WWTP (Catalonia, Spain) using the “seeds” methodology, which permits to calibrate models with the lowest number of parameters, avoiding the correlation among the parameters optimized. As a novel approach in ASM model calibration, the uncertainty on the influent characterization could be evaluated fixing the kinetic and operational variables at their default/common values and varying multipliers of the influent vector until reach the best objective function value and lower correlation amongst the calibration parameters (multipliers). One of the advantages of this novel approach was to identify what influent states should be better characterized. In terms of process control, the applied methodology was able to identify the most correlated operational variables, aiding to build decentralized control structures with less internal conflicts amongst all the WWTP feedback loops.

5. Acknowledgements

The authors greatly acknowledge to Ricard Tomas and Ana Lupón (Aigües de Manresa S.A.) all the support provided in conducting this work. Vinicius Cunha Machado has received a Pre-doctoral scholarship of the AGAUR (Agència de Gestió d'Ajuts Universitaris i Recerca - Catalonia, Spain), inside programs of the European Community Social Fund. This work was supported by the Spanish Ministerio de Economía y Competitividad (CTM2010-20384). The authors are members of the GENOCOV research group (Grup de Recerca Consolidat de la Generalitat de Catalunya, 2009 SGR 815).

6. References

1. Jeppsson U (1996) Modelling Aspects of Wastewater Treatment Processes. PhD thesis. Available from http://www.iea.lth.se/publications. Lund Institute of Technology, Sweden

2. Yuan Z, Bogaert H, Leten J, Verstraete W (2000) Reducing the size of a nitrogen removal activated sludge plant by shortening the retention time of inert solids via sludge storage. Water Res 34:539–549

3. Henze M, Grady Jr. CPL, Gujer W, Marais GR, Matsuo T (1987) Activated sludge model No. 1, IAWQ Scientific and technical report No.1. IWAQ, London

4. Meijer SCF, van Loosdrecht MCM, Heijnen JJ (2002) Modelling the start-up of a full-scale biological phosphorous and nitrogen removing WWTP. Water Res 36:4667–4682

5. Henze M, Gujer W, Mino T, van Loosdrecht MCM (2000) Activated sludge models ASM1, ASM2, ASM2d and ASM3: Scientific and technical report No. 9. IWA Publishing, London

6. Trutnau M, Petzold M, Mehlig L, Eschenhagen M, Geipel K, Müller S, Bley T, Röske I (2011) Using a carbon-based ASM3 EAWAG Bio-P for modelling the enhanced biological phosphorus removal in anaerobic/aerobic activated sludge systems. Bioprocess Biosyst Eng 34:287–295

7. Ferrer J, Morenilla JJ, Bouzas A, García-Usach F (2004) Calibration and simulation of two large wastewater treatment plants operated for nutrient removal. Water Sci Technol 50:87–94

8. Ingildsen P, Rosen C, Gernaey KV, Nielsen MK, Guildal T, Jacobsen BN (2006) Modelling and control strategy testing of biological and chemical phosphorus removal at Avedøre WWTP. Water Sci Technol 53:105–113

9. Xie W-M, Zhang R, Li W-W, Ni B-J, Fang F, Sheng G-P, Yu H-Q, Song J, Le D-Z, Bi X-J, Liu C-Q, Yang M (2011) Simulation and optimization of a full-scale Carrousel oxidation ditch plant for municipal wastewater treatment. Biochem Eng J 56:9–16

10. García-Usach F, Ferrer J, Bouzas A, Seco A (2006) Calibration and simulation of ASM2d at different temperatures in a phosphorus removal pilot plant. Water Sci Technol 53:199–206

11. Machado VC, Tapia G, Gabriel D, Lafuente J, Baeza JA (2009) Systematic identifiability study based on the Fisher Information Matrix for reducing the number of parameters calibration of an activated sludge model. Environ Model Softw 24:1274–1284

12. Olsson G (2006) Instrumentation, control and automation in the water industry--state-of-the-art and new challenges. Water Sci Technol 53:1–16

13. Copp JB, Spanjers H, Vanrolleghem PA (2002) Respirometry in Control of the Activated Sludge Process: Benchmarking Control Strategies. Scientific and Technical Report N^{o} 11. 160

14. Vrecko D, Hvala N, Stare A, Burica O, Strazar M, Levstek M, Cerar P, Podbevsek S (2006) Improvement of ammonia removal in activated sludge process with feedforward-feedback aeration controllers. Water Sci Technol 53:125–132

15. Samuelsson P, Carlsson B (2001) Feed-forward control of the external carbon flow rate in an activated sludge process. Water Sci Technol 43:115–122

16. Ayesa E, De La Sota A, Grau P, Sagarna JM, Salterain A, Suescun J (2006) Supervisory control strategies for the new WWTP of Galindo-Bilbao: the long run from the conceptual design to the full-scale experimental validation. Water Sci Technol 53:193–201

17. Orhon D, Artan N, Ates E (1994) A description of three methods for the determination of the initial inert particulate chemical oxigen demand of wastewater. J Chem Technol Biotechnol 61:73–80

18. Sin G, Van Hulle SWH, De Pauw DJW, van Griensven A, Vanrolleghem PA (2005) A critical comparison of systematic calibration protocols for activated sludge models: a SWOT analysis. Water Res 39:2459–2474

19. Ruano M V, Ribes J, De Pauw DJW, Sin G (2007) Parameter subset selection for the dynamic calibration of activated sludge models (ASMs): experience versus systems analysis. Water Sci Technol 56:107–115

20. Hulsbeek JJW, Kruit J, Roeleveld PJ, van Loosdrech MCM (2002) A practical protocol for dynamic modelling of activated sludge systems. Water Sci Technol 45:127–136

21. Vanrolleghem PA, Insel G, Petersen B, Sin G, De Pauw D, Nopens I, Dovermann H, Weijers S, Gernaey K (2003) A comprehensive model calibration procedure for activated sludge. WEFTEC 76th Annu. Tech. Exhib. Conf. Water Environment Federation, Los Angeles, California, pp 210–237

22. Melcer H, Dold PL, Jones RM, Bye CM, Takacs I, Stensel HD, Wilson AW, Sun P, Bury S (2003) Methods for Wastewater Characterization in Activated Sludge Modelling. Water Environment Research Foundation (WERF), Alexandria, VA, USA

23. Langergraber G, Rieger L, Winkler S, Alex J, Wiese J, Owerdieck C, Ahnert M, Simon J, Maurer M (2004) A guideline for simulation studies of wastewater treatment plants. Water Sci Technol 50:131–138

24. Cosenza A, Mannina G, Neumann MB, Viviani G, Vanrolleghem PA (2013) Biological nitrogen and phosphorus removal in membrane bioreactors: model development and parameter estimation. Bioprocess Biosyst Eng 36:499–514

25. Mannina G, Cosenza A, Vanrolleghem PA, Viviani G (2011) A practical protocol for calibration of nutrient removal wastewater treatment models. J Hydroinformatics 13:575–595

26. Refsgaard JC, van der Sluijs JP, Højberg AL, Vanrolleghem PA (2007) Uncertainty in the environmental modelling process – A framework and guidance. Environ Model Softw 22:1543–1556

27. Sin G, Gernaey K V, Neumann MB, Van Loosdrecht MCM, Gujer W (2009) Uncertainty analysis in WWTP model applications: a critical discussion using an example from design. Water Res 43:2894–2906

28. Cierkens K, Plano S, Benedetti L, Weijers S, de Jonge J, Nopens I (2012) Impact of influent data frequency and model structure on the quality of WWTP model calibration and uncertainty. Water Sci Technol 65:233–242

29. 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. Phys Chem Earth, Parts A/B/C 42-44:61–69

30. Benedetti L, Batstone DJ, De Baets B, Nopens I, Vanrolleghem PA (2012) Uncertainty analysis of WWTP control strategies made feasible. Water Qual Res J Canada 47:14

31. Belia E, Amerlinck Y, Benedetti L, Johnson B, Sin G, Vanrolleghem PA, Gernaey K V, Gillot S, Neumann MB, Rieger L, Shaw A, Villez K (2009) Wastewater treatment modelling: dealing with uncertainties. Water Sci Technol 60:1929–1941

32. Takács I, Patry GG, Nolasco D (1991) A dynamic model of the clarification-thickening process. Water Res 25:1263–1271

33. Orhon D, Artan N (1994) Modelling of activated sludge systems. 589p

34. Montpart N (2010) Redesign of a Dissolved Oxygen Control System in an Urban WWTP. Master in Environmental Studies. Universitat Autònoma de Barcelona, Barcelona, Catalonia, Spain.

35. Sin G, Gernaey K V, Neumann MB, Van Loosdrecht MCM, Gujer W (2009) Uncertainty analysis in WWTP model applications: a critical discussion using an example from design. Water Res 43:2894–2906

36. Vangsgaard AK, Mutlu AG, Gernaey K V, Smets BF, Sin G (2013) Calibration and validation of a model describing complete autotrophic nitrogen removal in a granular SBR system. J Chem Technol Biotechnol 88:2007–2015

Fig. 1 Scale map of the Manresa WWTP

Fig. 2 Monitored variables of the Manresa WWTP secondary treatment

Fig. 3 Simplified scheme of the overall calibration / validation process

Fig. 4 Model predictions using the best seed (subset from the seed $\eta_{NO3,D}$) and plant data (calibration data). For checking the parameter values used in this simulation, see Table 3

Fig. 5 Model predictions using the best subset (from seed $\eta_{NO3,D}$) and the validation data (plant data)

Table 1: Average influent composition.

Property	Winter (Average Temperature = 13°C)	Summer (Average Temperature = 27°C)
pH	7.9	7.6
NH_4^+ [mg N/L]	33	20
BOD5 [mg/L]	290	170
COD [mg/L]	600	460
Total N [mg N/L]	53	33
NO_3^- [mg N/L]	3.5	2.0
Total P, [mg P/L]	8.0	5.5
TKN [mg N/L] (Kjeldahl nitrogen)	48	33
Zn [mg Zn/L]	0.8	0.5

Table 2: Relative sensitivity of the weighted sum of ammonium, phosphate, nitrate, TKN and TSS in the effluent, for all the three groups of parameters.

Kinetic / Stoichiometric Group (K group)				
Order	**Parameter**	**Short Description**	**Related biomass or process**	**Sensitivity**
1	Y_H	Yield coefficient for X_H.	Heterotrophic	756
2	μ_A	Maximum growth rate of X_A	Autotrophic	678
3	b_A	Rate for lysis of X_A	Autotrophic	634
4	$K_{NH4,A}$	Saturation coefficient of substrate NH_4^+ for nitrification on S_{NH4}	Autotrophic	412
5	K_{PRE}	Precipitation constant	Chemical phosphate precipitation	150
6	$K_{O2,A}$	Saturation coefficient of O_2 for nitrification on S_{NH4}	Autotrophic	149
7	K_{RED}	Solubilisation constant	Chemical phosphate precipitation	148
8	b_H	Rate for lysis of X_H	Heterotrophic	97
9	$K_{ALK,A}$	Saturation coefficient of alkalinity for nitrification on S_{NH4}	Autotrophic	73
10	$\eta_{NO3,D}$	Reduction factor for denitrification	Heterotrophic	51
Influent Group (I group)				
Order	**Parameter**	**Short Description**	**Related biomass or process**	**Sensitivity**
1	f_{XS}	Multiplying factor of X_S representing an uncertainty on the estimated inlet X_S fraction	Influent characterization	670
2	f_{XTSS}	Multiplying factor of the inlet X_{TSS} vector.	Influent characterization	555
3	f_{XMeOH}	Multiplying factor of the inlet X_{MeOH} vector.	Influent characterization	439
4	f_{SPO4}	Multiplying factor of the inlet S_{PO4} vector.	Influent	429

			characterization	
5	f_{SNH4}	Multiplying factor of the inlet S_{NH4} vector.	Influent characterization	393
6	f_{SF}	Multiplying factor of the inlet S_F vector.	Influent characterization	247
7	f_{SALK}	Multiplying factor of the inlet S_{ALK} vector.	Influent characterization	169
8	f_{SI}	Multiplying factor of the inlet S_I vector.	Influent characterization	160
9	f_{SNO3}	Multiplying factor of the inlet S_{NO3} vector.	Influent characterization	87
10	f_{SA}	Multiplying factor of the inlet S_A vector.	Influent characterization	0
Operational Group (O group)				
Order	**Parameter**	**Short Description**	**Related biomass or process**	**Sensitivity**
1	f_{QW}	Multiplying factor of Q_W representing an uncertainty on the measured value of Q_W.	Process control	297
2	DO_Gain	Multiplying factor of DO concentration on the aerobic basins representing an uncertainty on the measured value of DO.	Process control	180
3	f_{QRINT}	Multiplying factor of Q_{RINT} representing an uncertainty on the measured value of Q_{RINT}.	Process control	135
4	f_{QRAS}	Multiplying factor of Q_{RAS} representing an uncertainty on the measured value of Q_{RAS}.	Process control	116

Table 3: Results of the calibration methodology for the kinetic Group K.

Items	Seeds									
	Y_H	μ_A	b_A	$K_{NH4,A}$	K_{PRE}	$K_{O2,A}$	K_{RED}	b_H	$K_{ALK,A}$	$\eta_{NO3,D}$
Parameters	Y_H b_A K_{PRE} b_H	μ_A Y_H K_{PRE} b_H	b_A Y_H K_{PRE} b_H	$K_{NH4,A}$ K_{PRE} Y_H b_H	K_{PRE} μ_A Y_H b_H	$K_{O2,A}$ K_{PRE} Y_H b_H b_A	K_{RED} μ_A Y_H b_H	b_H K_{RED} μ_A Y_H	$K_{ALK,A}$ K_{PRE} Y_H b_H	$\eta_{NO3,D}$ K_{PRE} b_A Y_H $K_{O2,A}$ b_H
Optimized Values	0.452 0.168 1.045 0.104	0.908 0.448 1.013 0.102	0.168 0.452 1.045 0.104	1.616 1.011 0.457 0.108	1.013 0.908 0.448 0.102	0.089 1.008 0.4105 0.0786 0.2277	0.593 0.908 0.448 0.101	0.101 0.593 0.908 0.448	0.895 1.011 0.449 0.103	0.0296 1.005 0.2203 0.4181 0.1130 0.0829
Parameter Confidence Interval (%)	22 3 9 59	3 26 9 64	3 22 9 59	6 9 21 48	9 3 26 64	68 9 30 71 5	9 3 27 66	66 9 3 27	16 9 25 61	52 9 9 22 114 52
Norm of Parameter Confidence Interval (%)	64	70	64	53	70	103	72	72	68	138
normD	$1.58\cdot10^{14}$	$4.72\cdot10^{12}$	$1.58\cdot10^{14}$	$5.46\cdot10^{11}$	$4.72\cdot10^{12}$	$1.81\cdot10^{16}$	$1.02\cdot10^{13}$	$1.02\cdot10^{13}$	$1.45\cdot10^{11}$	$9.40\cdot10^{21}$
modE	393.41	62.61	393.41	46.37	62.61	491.80	69.09	69.09	69.56	1420.93
RDEc	$4.03\cdot10^{11}$	$7.55\cdot10^{10}$	$4.03\cdot10^{11}$	$1.18\cdot10^{10}$	$7.55\cdot10^{10}$	$3.68\cdot10^{13}$	$1.47\cdot10^{11}$	$1.47\cdot10^{11}$	$2.09\cdot10^{9}$	$6.61\cdot10^{18}$
CCF	66.3	66.3	66.3	65.1	66.3	65.5	66.3	66.3	66.4	63.5
VCF	172.1	172.1	172.1	170.4	172.1	171.2	172.1	172.1	172.3	167.7
Janus	1.288	1.288	1.288	1.294	1.288	1.292	1.288	1.288	1.288	1.295

Table 4: Results of the calibration methodology for the Group I.

Items	Seeds									
	f_{XS}	f_{XTSS}	f_{XMeOH}	f_{SPO4}	f_{SNH4}	f_{SF}	f_{SALK}	f_{SI}	f_{SNO3}	f_{SA}
Parameters	f_{XS} f_{SNH4} f_{SPO4} f_{SALK} f_{XMeOH}	f_{XTSS} f_{SF} f_{SNH4} f_{SALK} f_{SPO4} f_{XMeOH}	f_{XMeOH} f_{SNH4} f_{SALK} f_{XS} f_{SPO4}	f_{SPO4} f_{SNH4} f_{SALK} f_{XS} f_{XMeOH}	f_{SNH4} f_{SPO4} f_{SALK} f_{XS} f_{XMeOH}	f_{SF} f_{XTSS} f_{SNH4} f_{SALK} f_{SPO4} f_{XMeOH}	f_{SALK} f_{SF} f_{XTSS} f_{SNH4} f_{SPO4} f_{XMeOH}	f_{SI} f_{SPO4} f_{SNH4} f_{SALK} f_{XS} f_{XMeOH} f_{SF} f_{SNO3}	f_{SNO3} f_{SF}	-
Optimized Values	1.038 1.116 0.758 0.949 0.936	0.537 2.861 1.433 1.126 0.708 1.223	0.936 1.116 0.949 1.038 0.758	0.758 1.116 0.949 1.038 0.936	1.116 0.758 0.949 1.038 0.936	2.861 0.537 1.433 1.126 0.708 1.223	1.126 2.861 0.537 1.433 0.708 1.223	6.835 0.706 1.414 1.266 1.361 1.229 2.472 0.144	1.009 0.929	-
Parameter Confidence Interval (%)	9 4 10 6 10	26 16 5 6 12 11	10 4 6 9 10	10 4 6 9 10	4 10 6 9 10	16 26 5 6 12 11	6 16 26 5 12 11	7 12 5 13 912 18 96	35 9	-
Norm of Parameter Confidence Interval (%)	18	35	18	18	18	35	35	101	36	-
normD	$1.336\cdot 10^{16}$	$2.635\cdot 10^{16}$	$1.336\cdot 10^{16}$	$1.336\cdot 10^{16}$	$1.336\cdot 10^{16}$	$2.635\cdot 10^{16}$	$2.635\cdot 10^{16}$	$9.148\cdot 10^{18}$	16598	-
modE	99.320	1480.73	99.320	99.320	99.320	1480.73	1480.73	1138.80	18.66	-
RDEc	$1.345\cdot 10^{14}$	$1.779\cdot 10^{13}$	$1.345\cdot 10^{14}$	$1.345\cdot 10^{14}$	$1.345\cdot 10^{14}$	$1.779\cdot 10^{13}$	$1.779\cdot 10^{13}$	$8.033\cdot 10^{15}$	889	-
CCF	66.1	63.6	66.1	66.1	66.1	63.6	63.6	55.8	67.6	-
VCF	170.9	168.4	170.8	170.8	170.8	168.4	168.4	162.3	172.3	-
Janus	1.289	1.311	1.289	1.289	1.289	1.311	1.311	1.371	1.278	-

Table 5: Results of the calibration methodology for the Group O.

Items	Seeds			
	fQ_w	DO_Gain	fQ_{rint}	fQ_{RAS}
Parameters	fQ_w fQ_{rint} DO_Gain	DO_Gain fQ_w fQ_{rint}	fQ_{rint} fQ_w DO_Gain	fQ_{RAS} DO_Gain fQ_{rint} fQ_w
Optimized Values	0.344 0.389 0.931	0.931 0.344 0.389	0.389 0.344 0.931	2.781 0.925 0.122 0.388
Parameter Confidence Interval (%)	8 18 11	11 8 18	18 8 11	15 11 97 9
Norm of Parameter Confidence Interval (%)	23	22	22	99
normD	$1.61 \cdot 10^9$	$1.61 \cdot 10^9$	$1.61 \cdot 10^9$	$3.26 \cdot 10^{10}$
modE	13.78	13.78	13.78	193.77
RDEc	$1.17 \cdot 10^8$	$1.17 \cdot 10^8$	$1.17 \cdot 10^8$	$1.680 \cdot 10^8$
CCF	62.3	62.3	62.3	62.2
VCF	168.9	168.9	168.9	168.9
Janus	1.322	1.322	1.322	1.323

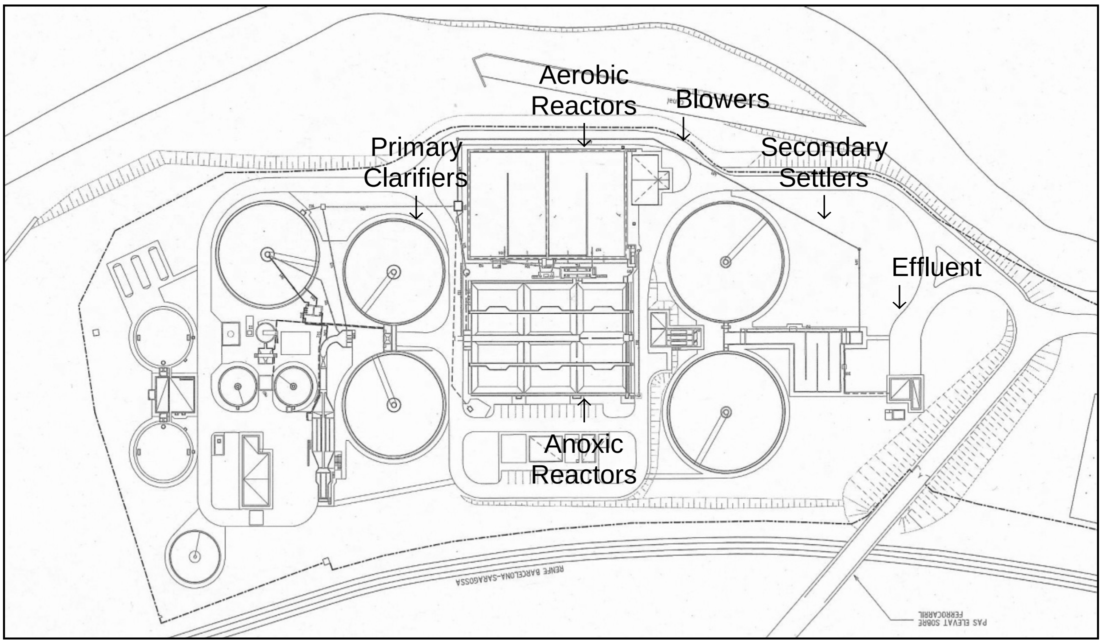

Fig. 1 Scale map of the Manresa WWTP

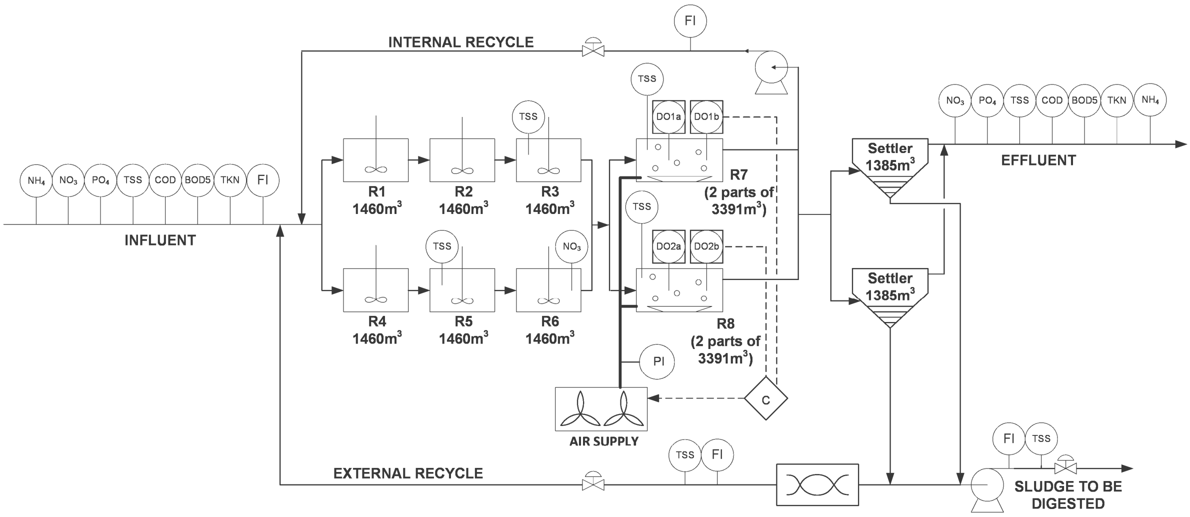

Fig. 2 Monitored variables of the Manresa WWTP secondary treatment

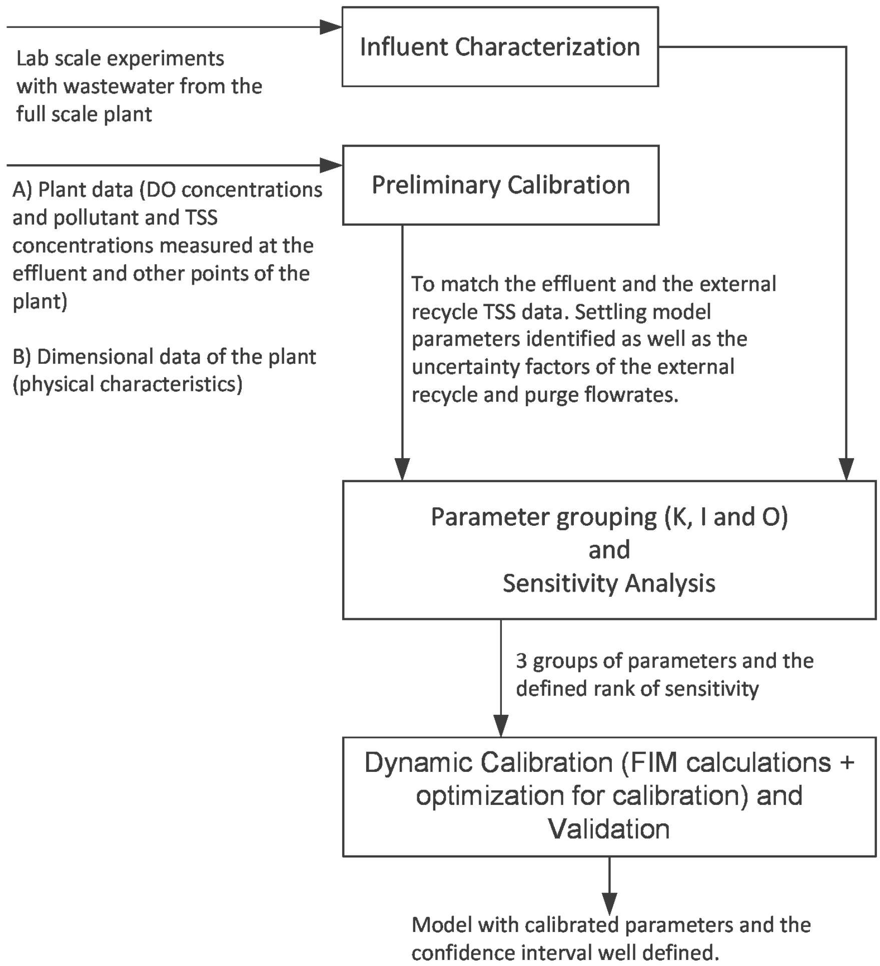

Fig. 3 Simplified scheme of the overall calibration / validation process

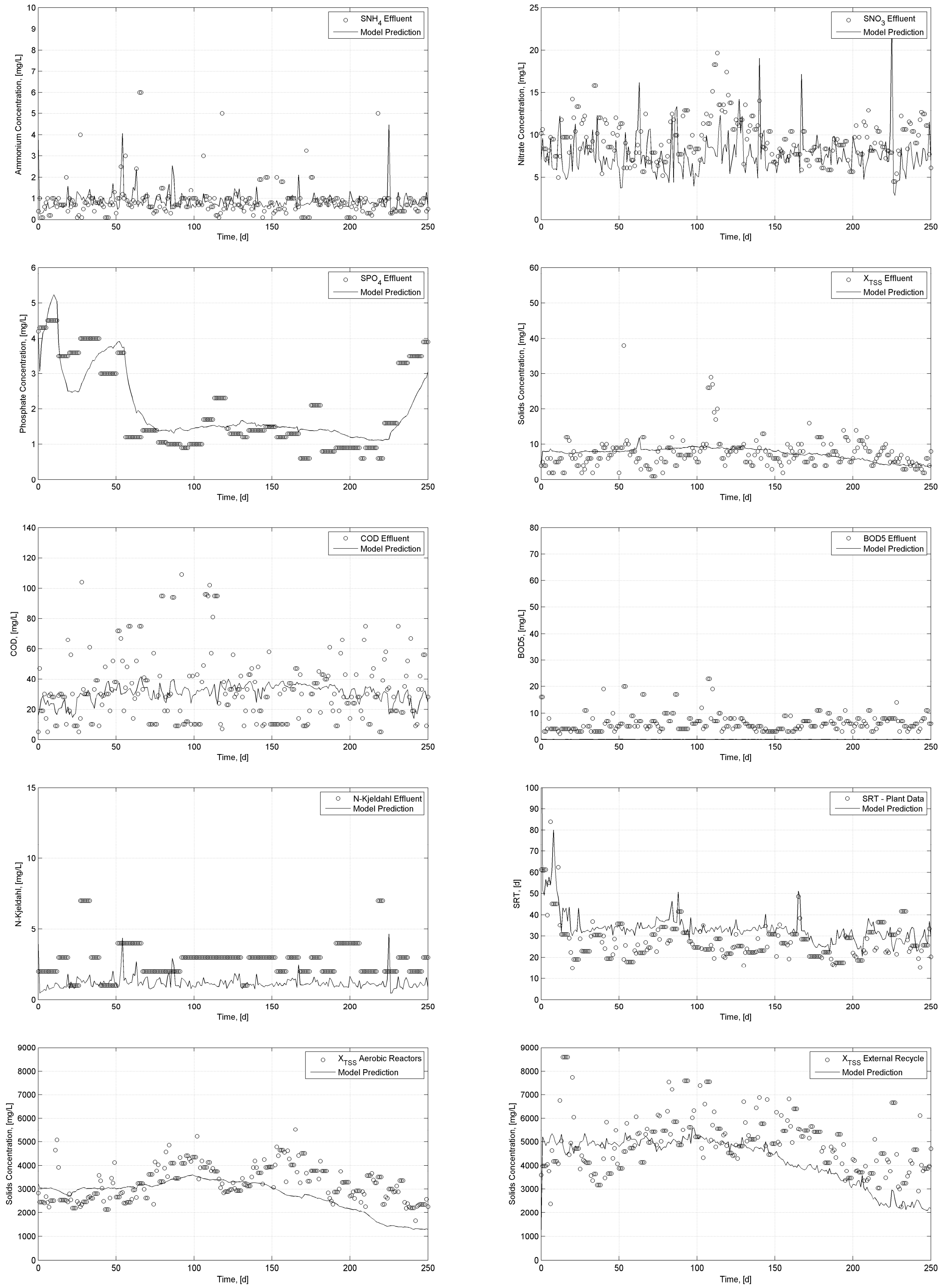

Fig. 4 Model predictions using the best seed (subset from the seed $\eta_{NO3,D}$) and plant data (calibration data). For checking the parameter values used in this simulation, see Table 3

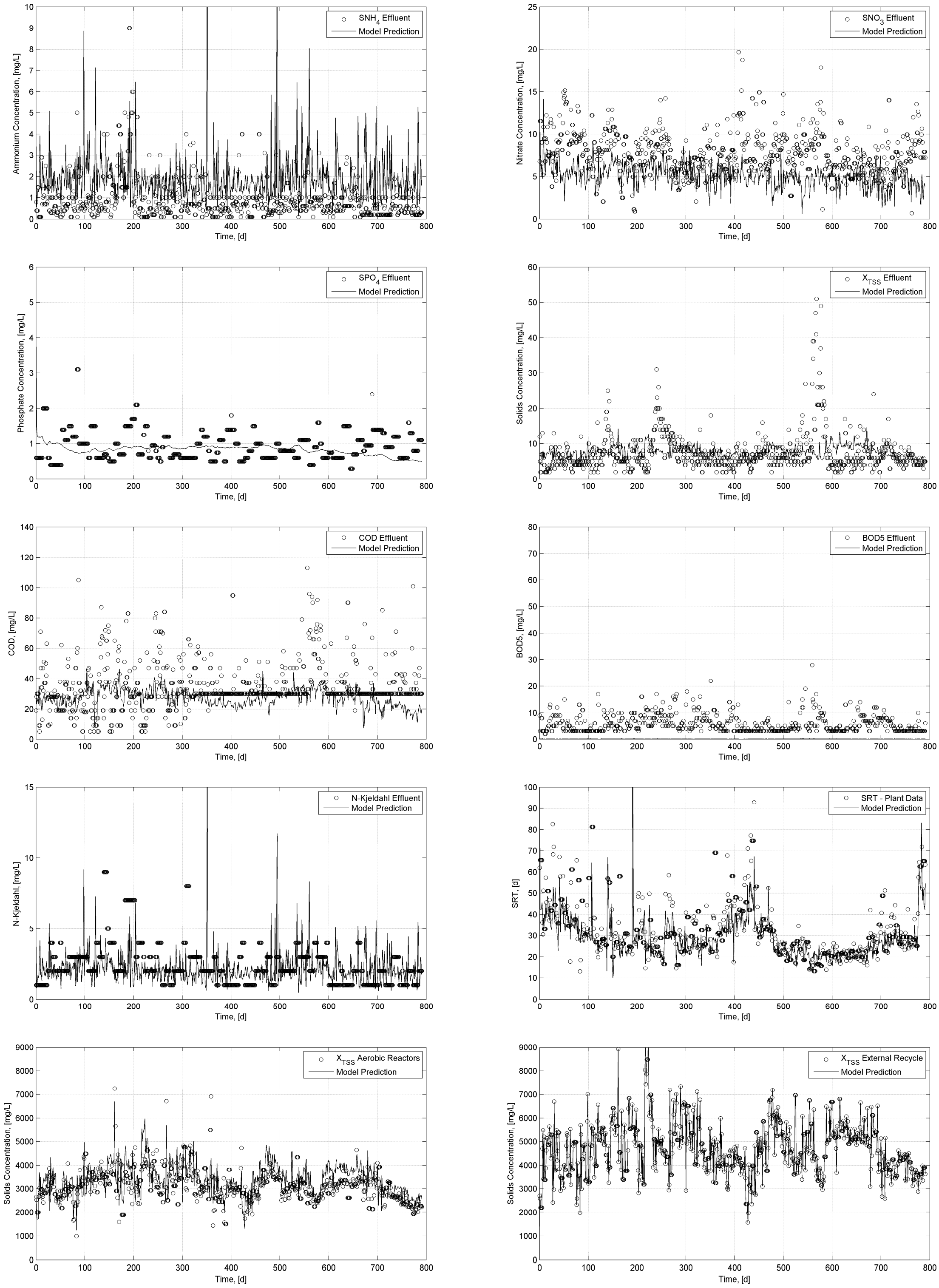

Fig. 5 Model predictions using the best subset (from seed $\eta_{NO3,D}$) and the validation data (plant data)

Supplementary information

Activated Sludge Model 2d calibration with full-scale WWTP data: comparing model parameter identifiability with influent and operational uncertainty.

Vinicius Cunha Machado, Javier Lafuente, Juan Antonio Baeza*

Department of Chemical Engineering, Universitat Autònoma de Barcelona, ETSE, 08193 Bellaterra (Barcelona), Spain. Phone: +34935811587. FAX: +34935812013.

E-mails: vinicius.cunhamachado@gmail.com, javier.lafuente@uab.cat, juanantonio.baeza@uab.cat

*Corresponding Author

S1. Influent Characterization Procedure

Orhon *et al*. [1] developed a method to determine the values of S_I, X_I, X_S and S_F (ASM2d states) in the effluent, using the well-know measurement of the COD. X variables are the particulate variables while S variables indicate soluble variables. Such method allows making an interface between the COD and ASM2d state variables.

The experimental determination of S_I and X_I is performed in two parallel CSTR reactors, one of them fed with raw WWTP influent and the other one fed with filtered WWTP influent. Both reactors operate as long as all the biological reactions have been ceased and daily analysis of total COD and the soluble COD are performed. At a sufficient time, both values of COD of the two systems will be approximately constant. At the end of the experiment, the relationship between the initial and final values of total COD and soluble COD of both systems will help to estimate S_I and X_I.

X_S is present at the beginning of the experiment for reactor 1 (with raw influent, without filtering) and it is not for reactor 2 (with filtered WW). At the end of the experiment, in both systems X_S and S_F no longer exist, differently of S_P and X_P that are produced by the microorganisms along the experiment time. S_P and X_P are, respectively, soluble and particulate residual biodegradable matter, product of microorganism activity. X_I is present at the end of the experiment only in reactor 1 (no filtered WW). With these observations, it is possible to write a system of equations as follows:

Reactor 1 (Fed with raw wastewater)	Reactor 2 (Fed with filtered wastewater)
$C_{T0} = S_{F0} + X_{S0}$ Eq. S.1	$C_{T0} = S_{T0}$ Eq. S.4
$C_{T1} = X_{I1} + S_{I1} + X_{P1} + S_{P1}$ Eq. S.2	$C_{T2} = X_{I2} + S_{I2} + X_{P2} + S_{P2}$ Eq. S.5
$S_{T1} = S_{I1} + S_{P1}$ Eq. S.3	$S_{T2} = S_{I2} + S_{P2}$ Eq. S.6

Variable C_T means the total substrate concentration in reactors. S_T means total soluble substrate. The lowercase "0" in equations S.1 and S.4 means "initial value" for variables in reactor 1 and 2, respectively. In equations S.2 and S.3 the lowercase "1" means the values at the end of the experiment in reactor 1. The same notation is used for reactor 2, in equations S.5 and S.6. For a better understanding of the whole experiment, Figure S.1 shows an illustration of the evolution of total COD and total soluble COD.

Using the equations S.1 to S.6, X_I is determined with equation S.7.

$$X_I = (C_{T1} - S_{T1}) - \left\{ [C_{T2} - S_{T2}] \cdot \frac{[C_{T0} - C_{T1}]}{[S_{T0} - C_{T2}]} \right\} \qquad \text{Eq. S.7}$$

A similar procedure is performed to determine S_I.

$$S_I = S_{T1} - \left\{ \frac{S_{T1} - S_{T2}}{1 - \left[\frac{S_{T0} - C_{T2}}{C_{T0} - C_{T1}} \right]} \right\} \qquad \text{Eq. S.8}$$

S_F value can be obtained by taking the value of total soluble COD of reactor 2 at the beginning of the experiment for determining X_I and S_I and subtracting the value of S_I (obtained by Eq. S.8).

$$S_F = COD_{\text{Soluble (filtered WW)}} - S_I \qquad \text{Eq. S.9}$$

Finally, X_S is determined by using measures of total COD in reactor 1.

$$X_S = DQO_{\text{total}} - (S_A + S_F + S_I + X_I) \qquad \text{Eq. S.10}$$

In Eq. S.10, S_A should be considered null (no conditions of fermenting X_S to produce S_A in the urban sewage system) and the rest of variables were already determined.

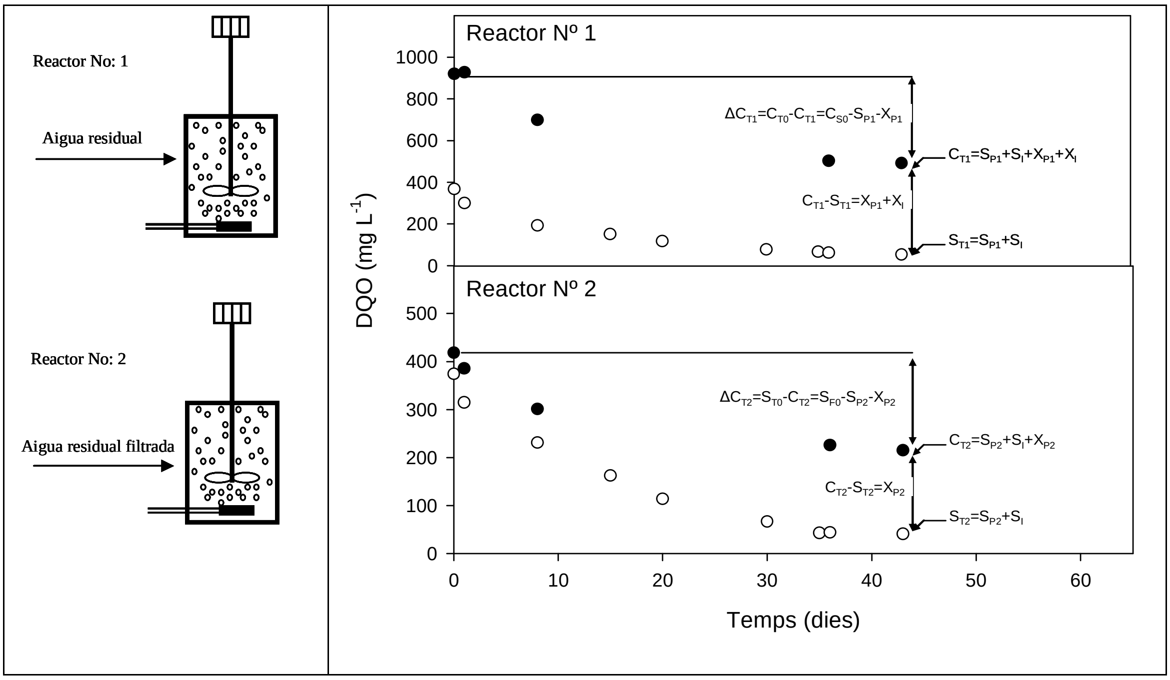

Figure S.1: Illustration of the lab scale reactors, total COD and total soluble COD data for determining S_I and X_I fractions in the secondary stage influent in a WWTP (• Total COD, ○ Total soluble COD).

S.2. Sensitivity Analysis

Sensitivity analysis allows making a ranking of the most important parameters that affect the outputs. Relative sensitivity of an output *i* (y_i) respect a parameter *j* (θ_j) is defined as [2],

$$S_{ij} = \frac{\theta_j}{y_i} \frac{dy_i}{d\theta_j} \quad \text{Eq. S.11}$$

Norton [3] proposed the utilization of algebraic sensitivity analysis because the numerical value of sensitivity applies only for a specific change from a specific value of θ_j, while the former provides algebraic relations. Numerical values of sensitivity are generally much less informative than an algebraic relation, but algebraic sensitivity analysis is not feasible if the equations of the model are complicated as in ASM2d. Therefore, the derivatives of equation S.11 were determined numerically by the finite differences method. The central difference approach with 10^{-4} (0.01%) as perturbation factor was used for the sensitivity calculations of each tested parameter around the default ASM2d value. This perturbation factor was selected because it produced equal derivative values with forward and backward finite differences [4].

The overall sensitivity of a parameter was calculated by adding absolute values of individual sensitivities. In our case, 5 output variables were declared (phosphate, ammonium, nitrate, TSS and TKN concentrations at the effluent). Hence, the overall sensitivity value of a parameter *j* (OS_j) was calculated with equation S.12.

$$OS_j = \left|S_{j,PO_4}\right| + \left|S_{j,NH_4}\right| + \left|S_{j,NO_3}\right| + \left|S_{j,XTSS}\right| + \left|S_{j,TKN}\right| \quad \text{Eq. S.12}$$

S.3. The Fisher Information Matrix and Parameter Confidence Interval

The FIM summarizes the importance of each model parameter over the outputs, since it measures the variation of output variables caused by a variation of model parameters [5, 6]. Algebraically, the FIM is represented by equation S.13.

$$FIM = \sum_{k=1}^{N} Y_\theta(k) \cdot Q_k^{-1} \cdot Y_\theta^T(k) \qquad \text{Eq. S.13}$$

For a FIM calculated for *r* output variables and p parameters, it is a *p* x *p* matrix, where *k* represents each sampling data point, Q_K is the *r* x *r* covariance matrix of the measurement noise, θ is the vector of *p* parameters, *N* is the total number of samples and Y_θ is the *p* x *r* output sensitivity function matrix, expressed by equation S.14.

$$Y_\theta^T(t) = \left[\frac{\partial y(t, \theta_0)}{\partial \theta^T} \right]_{\theta_0} \qquad \text{Eq. S.14}$$

where θ_0 is the complete model parameter vector used for calculating the derivatives and θ_T is the transposed parameter vector, which its elements are being studied. In the present study, the derivative shown in equation S.14 was numerically obtained by finite differences using a perturbation factor of 10^{-4} as in the sensitivity calculations. Mathematically was proved that the FIM provides a lower bound of the parameter error covariance matrix [7] as shown by equation S.15.

$$\mathrm{cov}(\theta_0) \geq FIM^{-1} \qquad \text{Eq. S.15}$$

This FIM property was used for calculating the confidence interval $\Delta\theta_j$ with equation S.16 for a given parameter θ_j [8].

$$\Delta\theta_j = t_{\alpha, N-p} \sqrt{\mathrm{cov}(\theta_j)} \qquad \text{Eq. S.16}$$

where t is the statistical t-student with α = 95% of confidence and N-p degrees of freedom (number of experimental data points minus p parameters), and cov(θ_j) was assumed as FIM^{-1}_{jj}.

As can be observed, the calculation of the parameter error covariance matrix using the FIM involves its inversion. To be invertible, the FIM should have a determinant different from zero and should not be ill-conditioned. To match these requirements any pair of matrix columns should not be very similar. As each column of the matrix represents a parameter, the determinant and the condition number of the FIM provides a reasonable measurement of the correlation of a set of parameters. Hence, parameters less correlated will easily provide a diagonal-dominant matrix. The FIM determinant (D criterion) and the ratio between the highest and the lowest FIM eigenvalue (modE criterion) can be used as criteria for parameter subset selection. A modE criterion value close to the unity indicates that all the involved parameters independently affect the outputs while the shape of the confidence region is similar to a circle (2 parameters) or a sphere (3 parameters) and not ellipses and ellipsoids as occur with correlated parameters. A high D criterion value means lower values of the diagonal elements of the covariance matrix, and as a consequence, lower confidence intervals of the parameters. As the D criterion is dependent on the magnitude of the involved parameters, this criterion was normalized (normD) according to Equation S.17.

$$normD = D \cdot \left\| \theta_P \right\|^2 \qquad \text{Eq. S.17}$$

where $||\theta_P||$ is the Euclidean norm of the parameter vector. Such normalization works as a scaling factor and allows comparisons among subsets with the same size but with different parameters.

From the system engineering point of view, it is important to include in the parameter subset those parameters that maximize the D criterion and minimize the modE criterion.

Hence, the ratio between the normD and the modE criteria (RDE criterion) was proposed [9] as an interesting index to define subsets of parameters for calibration. The RDE criterion (Equation S.18) establishes the capacity of a parameter subset to explain experimental data coupled to low uncertainty in the estimated parameters.

$$RDE = \frac{normD}{modE} \qquad \text{Eq. S.18}$$

S.4. References

1. Orhon D, Artan N, Ates E (1994) A description of three methods for the determination of the initial inert particulate chemical oxigen demand of wastewater. J Chem Technol Biot 61:73–80

2. Reichert P, Vanrolleghem PA (2001) Identifiability and uncertainty analysis of the river water quality model no. 1 (RWQM1). Water Sci Technol 43:329–338

3. Norton JP (2008) Algebraic sensitivity analysis of environmental models. Environ Model Softw 23:963–972

4. De Pauw DJW (2005) Optimal Experimental Design for Calibration of Bioprocess Models: A Validated Software Toolbox. PhD thesis in Applied Biological Sciences. Available from: http://biomath.ugent.be/publications/download/. University of Gent, Belgium

5. Dochain D, Vanrolleghem PA (2001) Dynamical modelling and estimation in wastewater treatment processes. IWA Publishing, London

6. Guisasola A, Baeza JA, Carrera J, Sin G, Vanrolleghem PA, Lafuente J (2006) The influence of experimental data quality and quantity on parameter estimation accuracy. Education for Chemical Engineers 1:139–145

7. Söderström T, Stoica P (1989) System identification. Prentice-Hall, Englewood Cliffs, New Jersey

8. Seber GAF, Wild CJ (1989) Nonlinear regression. Wiley, New York

9. Machado VC, Tapia G, Gabriel D, Lafuente J, Baeza JA (2009) Systematic identifiability study based on the Fisher Information Matrix for reducing the number of parameters calibration of an activated sludge model. Environ Model Softw 24:1274–1284