Wastewater Stabilisation Ponds System: Global Sensitivity Analysis on Network Design

Maria Paz Ochoa, Vanina Estrada, Patricia M. Hoch*

Planta Piloto de Ingeniería Química – PLAPIQUI –CONICET
p.hoch@plapiqui.edu.ar

In this work, a wastewater treatment network (WWTN) consisting of a system of ponds is modelled rigorously, taking into account dynamic mass balances for the main groups of bacteria, together with different types of organic load, algae biomass, nutrients, etc.

To obtain the optimal configuration, we had first formulated a superstructure embedding these rigorous models as a mixed integer non-linear programming (MINLP) problem, with the objective to synthesise and design the WWTN that minimises the total annual costs, subject to environmental regulations. Then, a global sensitivity analysis (GSA) is performed on this kinetic dynamic model for the obtained optimal configuration of three stabilisation ponds (two aerobic ponds in series followed by a facultative one) to determine the most influential parameters of the model considering the whole range of parameters variation, as well as parameter ranking. GSA is implemented using Sobol's method, a variance based technique. The technique is implemented within gPROMS platform, a differential algebraic equation oriented environment where stochastic simulations are performed. Temporal profiles for the first order, total order and interactional sensitivity indices are obtained for the main differential and algebraic state variables.

Numerical results provide useful information about the complex relationships between technological, economic and environmental variables of the processes in the WWTN design optimisation.

1. Introduction

In a previous work, a WWTN embedding rigorous models of different types of stabilisation ponds was developed with synthesis and design purposes by Ochoa et al. (2016a). The model was formulated as a MINLP problem to establish optimal configuration system and pond dimensions necessary to conform to environmental discharge regulations. In addition, different scenarios of inlet concentration of organic load were explored to take into account process uncertainty leading to the conclusion that the model was highly influenced by initial organic load.

In this paper, another type of uncertainty is considered, related to time-invariant model parameters such as kinetic parameters, whose uncertainty results from measurement errors or the impossibility of modelling exactly the physical behaviour of a system. The values of state variable can be greatly influenced by this type of uncertainty. For this reason, it is important to identify the parameters to which model state variables are most sensitive, which is achieved in general by a sensitivity analysis (SA). Techniques for sensitivity analysis can be classified into local and global. Local techniques are based on Taylor series expansion, requiring linearity and additivity. Global sensitivity analysis (GSA) aims to quantify the relative importance of input variables or factors in determining the value of an output variable. GSA method should be used when the model is nonlinear and various input variables are affected by uncertainties of different orders of magnitude, because they take into account the influence of the whole range of variation and the form of the probability density function in the input. GSA method computes the effect of factor x_i while all others x_j, $j \neq i$, are varied as well. Temporal profiles of the influence of factors and input variables are obtained when dynamic models are analysed, leading to a great insight on the sources of uncertainty during the time horizon (Saltelli et al., 2008).

The wastewater stabilisation pond network model and the GSA methodology were implemented in an equation oriented environment with a differential algebraic equation solver in gPROMS (PSEnterprise, 2014). The implemented GSA strategy is variance-based (Sobol', 1993) and allows the determination and

classification of model parameters, according to their sensitivity indices. Temporal profiles of first order effect sensitivity indices, total sensitivity indices and those due to interactions with other model parameters have been calculated for parameters in a wastewater stabilisation pond system model. Numerical results show that the higher computational cost of global sensitivity analysis is thoroughly justified, where not only first order effects due to each parameter can be captured, but also due to interaction with other model parameters.

2. Methodology

2.1 Wastewater stabilisation pond model

A differential algebraic equation system represents three stabilisation ponds in series for biological wastewater treatment. Differential equations include dynamic mass balance for the prevailing groups of bacteria: heterotrophic, autotrophic, fermenting, acetotrophic sulphate reducing and acetotrophic methanogenic bacteria; algae biomass, the different sources of organic matter, main nutrients involved in the process and methane emissions. Whereas algebraic equations describe pond interconnections, forcing functions and reaction rates of the processes. Kinetic expressions of the reaction rates and nominal value of stoichiometric coefficients were taken from Sah et al. (2011).

Pond mass balances

$$\frac{dV_i}{dt} = Q_{in,i} - Q_{out,i} + Q_{rain,i} - Q_{evap,i} \tag{1}$$

$$\frac{dC_{i,j,m}}{dt} = \frac{C_{in,i,j} \cdot Q_{in,i} - C_{out,i,j,m} \cdot Q_{out,i}}{V_i} + r_{i,j,m} - \frac{k_m A_i (C_{i,j,m} - C_{i,j,m+1})}{\Delta h \, V_{i,m}} - C_{i,j,m} \cdot \frac{dV_{i,m}}{dt} \qquad m = 1 \tag{2a}$$

$$\frac{dC_{i,j,2}}{dt} = r_{i,j,m} + \frac{k_m A_i (C_{i,j,m-1} - C_{i,j,m})}{\Delta h \, V_{i,m}} - C_{i,j,m} \cdot \frac{dV_{i,m}}{dt} \qquad m = 2 \tag{2b}$$

Generation and consumption

$$r_{i,j,m} = \sum_k \nu_{j,k} \cdot \rho_{k,i,m} \tag{3}$$

$Q_{in,i}$, $Q_{out,i}$, $Q_{rain,i}$ and $Q_{evap,i}$ stand for volumetric flow (m^3/d). $C_{in,i,j,k}$ and $C_{out,i,j,k}$, are the concentrations of component j related to pond i at layer m; $V_{i,m}$ is the pond volume i of the layer m (m^3) and $r_{i,j,m}$ is the reaction term obtained by the sum of the stoichiometric coefficients with $\rho_{k,i,m}$ as the process reaction rates that take place within the pond i and layer m. Note that only facultative type pond has two layers.

Configuration of the system was a result of the optimisation, where pond type was classified by the processes that occur within the lagoons into anaerobic, facultative and aerobic. Binary variables, $z_{type,i}$, were associated to the different types of processes, through big M constraints: Eq(4), and to different bounds of pond dimensions in order to define the pond type. Eq(5) ensured that there exists a pond of any type.

$$\rho_{k,i} - z_{type,i} \cdot M \leq 0 \quad k \in type, \; type = anaer, fac, aer \tag{4}$$

$$\sum_{type} z_{type,i} = 1 \tag{5}$$

Where anaer, fac and aer represent the sets of anaerobic, facultative and aerobic processes, respectively.

The optimisation of pond stabilisation systems consists of minimising the total cost subject to environmental discharge regulations. The total cost is the sum of capital and operating costs, considering cost of land, construction, lining and maintenance of the system. Medri et al. (2007) proposed linear expressions for each type of cost: land, construction, revetment and maintenance. The objective function for minimising the total cost system is then described by Eq(6).

$$\min C_t = \sum_i^n C_{l,i} + C_{c,i} + C_{r,i} + C_{m,i} \tag{6}$$

Where C_t is the total cost of the system (U$S) and n is the number of ponds.

The model is formulated as a MINLP problem implemented in GAMS (Rosenthal, 2016) and solved by DICOPT, a solver based on the extensions of the outer-approximation algorithm for the equality relaxation strategy. More details of the model and solution strategy can be found in Ochoa et al. (2016a).

2.2 Global sensitivity analysis technique

Sobol's method is used to compute sensitivity indices. This method is based on variance decomposition, using Monte-Carlo simulation methods (Sobol' I., 2001; Saltelli and Tarantola, 2002).

Given a function y=f(x,t), where y is a differential or algebraic state variable (e.g. heterotrophic bacteria concentration), x is a vector of k input parameters and t is the independent variable in differential equations, e.g., time; when all uncertain parameters x_i vary under its probability density function, the uncertainty in y(x,t) can be quantified by its unconditional variance V(y,t). It can be decomposed as the sum of the variance of a conditional expected value and the expected value of a conditional variance, conditioning with respect to both x_i and x_{-i}.

$$V(y, t) = V\big(E(y|x_i, t)\big) + E\big(V(y|x_i, t)\big) \tag{7}$$

$$V(y, t) = V\big(E(y|x_{-i}, t)\big) + E\big(V(y|x_{-i}, t)\big) \tag{8}$$

V and E correspond to variance and expected value operators, respectively. In Eq(7), $V(y,t)_i = V(E(y|x_i,t))$ computes the variance (over all possible realizations of parameter x_i) of the conditional expected value of the state variable y under all parameters variation, except x_i. It represents the expected reduction in the state variable variance that could be obtained if x_i could be known or fixed. $V(y,t)_i$ is the first-order effect associated to parameter x_i. The residual effect, $E(y,t)_i = E(V(y|x_i,t))$, is the expected value (over all realizations of parameter x_i.) of the conditional variance of the state variable y when all parameters except x_i change. It represents the average state variable variance if x_i could be known or fixed.

The same can be stated for Eq(8), by replacing x_i for "all parameters except x_i" (x_i). Thus, the term $V(y,t)_i^{TOT} = E(V(y|x_{-i},t))$ computes the average state variable variance if all parameters except x_i could be known or fixed. If equations (7) and (8) are normalised, the first-order sensitivity index, $S(y,t)_i$ and the total sensitivity index $S(y,t)_i^{TOT}$ are defined as:

$$S(y, t)_i = \frac{V(E(y|x_i,t))}{V(y,t)} = \frac{V(y,t)_i}{V(y,t)} \tag{9}$$

$$S(y, t)_i^{TOT} = \frac{E(V(y|x_{-i},t))}{V(y,t)} = \frac{V(y,t)_i^{TOT}}{V(y,t)} \tag{10}$$

To calculate sensitivity indices it is necessary to compute the unconditional and conditional variances of each state variable, involving the calculation of multiple integrals. However, Sobol' proposes a methodology to compute the variances considering only evaluations of functions (y=f(x,t)), as defined by Eqs (11), (12) and (13) (Sobol' and Shukhman, 2007).

$$\frac{1}{2N}\sum_{j=1}^{N}\big(y_A^j - y_B^j\big)^2 \rightarrow V(y) \tag{11}$$

$$\frac{1}{N}\sum_{j=1}^{N} y_A^j \cdot \big(y_{C_i}^j - y_B^j\big) \rightarrow V(y, t)_i \qquad i = 1..k \tag{12}$$

$$\frac{1}{2N}\sum_{j=1}^{N}\big(y_A^j - y_{D_i}^j\big)^2 \rightarrow V(y, t)_i^{TOT} \qquad i = 1..k \tag{13}$$

A, B, C_i and D_i, are matrices of dimension (N x k), N the sample size used for the Monte Carlo estimate and k the number of uncertain model parameters. Each column of both A and B matrices is a sample from the distribution function of the relative parameter. Each row is an input sample, for which a model output y can be evaluated. A is considered the 'sample' matrix and B the 're-sample' matrix. C_i is the matrix where all parameters except x_i are re-sampled, whereas D_i is the matrix where only x_i is re-sampled. y_A, y_B, y_{Ci} and y_{Di} are vectors of N model outputs values obtained when model variables are evaluated in matrices A, B, C_i and D_i, respectively. Thus, the total number of model evaluation is N(k+2).

First order ($S(y,t)_i$) and total effect ($S(y,t)_i^{TOT}$) indices measure the effect of the variation of the parameters on the model state variables. First sensitivity indices provide the reduction on the unconditional variance of the state variable that can be obtained if x_i is fixed at its true value. On the other hand, total sensitivity indices take into account the interactions among parameters, providing information on the non-additive part of the model.

The interactional sensitivity index is defined by the difference between the total sensitivity index and the first order sensitivity index.

$$S_i^{int} = S_i^{TOT} - S_i \tag{14}$$

A more extended explanation of the method can be found in Ochoa et al. (2016b).

Table 1: Uncertain parameters of the model

Uncertain Parameter	Description	Unit	Nominal Value	Standard Deviation
θ	Temperature coefficient		1.070	0.1338
k_m	Mass transfer coefficient between layers in facultative pond		0.022	0.0027
K_{NHa}	Saturation/Inhibition coefficient of ammonium and ammonia nitrogen, C_{NH}, for autotrophic bacteria, C_a	g_N/m^3	0.200	0.0250
i_{nxs}	Fraction of nitrogen in slowly biodegradable particulate COD, C_s	g_N/g_{COD}	0.040	0.0050
K_{NHh}	Saturation/Inhibition coefficient of ammonium and ammonia nitrogen, C_{NH}, for heterotrophic bacteria, C_h	g_N/m^3	0.050	0.0063
i_{nsf}	Fraction of nitrogen in fermentable, readily biodegradable soluble COD, C_f	g_N/g_{COD}	0.030	0.0038
b_a	Decay rate of autotrophic bacteria, C_a	d^{-1}	0.015	0.0019
b_{fb}	Decay rate of fermenting bacteria, C_{fb}	d^{-1}	0.020	0.0025
b_{asrb}	Decay rate of acetotrophic sulphate reducing bacteria, C_{asrb}	d^{-1}	0.012	0.0015
b_{amb}	Decay rate of acetotrophic methanogenic bacteria, C_{amb}	d^{-1}	0.008	0.0010

3. Numerical results

With the model presented in the previous section, a GSA analysis is performed. To generate the sample matrices N=1000 scenarios were considered. Time horizon was set to the operating time (12 days).

As can be seen in Figure 1 to 4, for ammonium and ammonia nitrogen (C_{NH}) and nitrate and nitrite nitrogen (C_{NO}) total sensitivity indices profiles, it is evident that the temperature coefficient θ is the most influential parameter within the three ponds for almost the entire operating time. This behaviour is repeated for the 19 differential variables studied, not only due to first order effect, but also through interactional effects between the temperature coefficients with the other parameters. Therefore, focusing efforts on fixing this parameter to its true value would lead to the greatest reduction in the unconditional variance of the output.

Another relevant parameter is the mass transfer coefficient between layers, k_m. It only contributes to the variance output in the upper and lower layer of the facultative pond. For the nitrate and nitrite nitrogen concentration, $C_{NO,3,2}$, it explains up to 93% of the total variance, affecting almost the entire operating time (Figure 4b). In the case of $C_{NH,3,2}$, k_m accounts up to the 20% of the total variance with a greater influence at the end of the time horizon, as it can be seen in Figure 2b. It also contributes to the total variance of dissolved oxygen, $C_{O,3,2}$, and sulphate sulphur, $C_{SO4,3,2}$, concentration of the lower layer of the facultative pond. And to a lesser extent, it influences the fermentation products, $C_{SA,3,2}$, and inert soluble COD concentration, $C_{I,3,2}$.
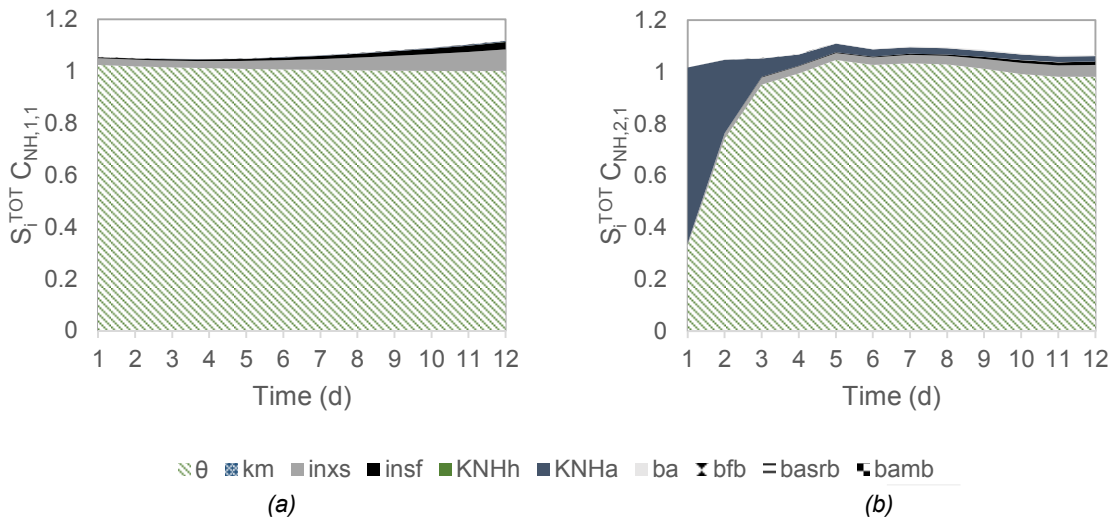

(a) *(b)*

Figure 1: Total sensitivity indices profiles for ammonium and ammonia nitrogen concentration in (a) the first aerobic pond ($C_{NH,1,1}$) and (b) the second aerobic pond ($C_{NH,2,1}$).
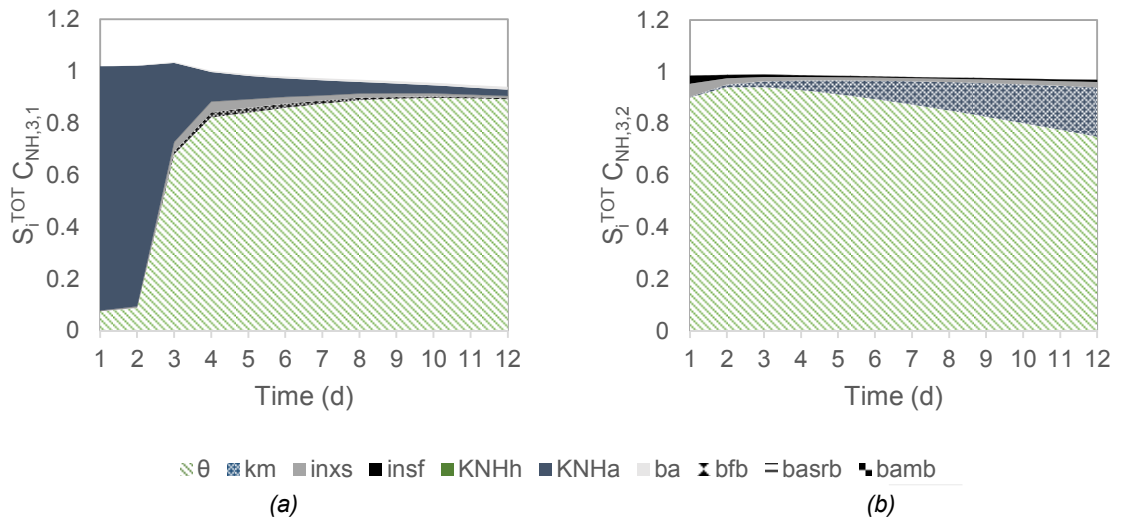
(a)   *(b)*

Figure 2: Total sensitivity indices profile for ammonium and ammonia nitrogen concentration in the facultative pond: (a) the upper layer ($C_{NH,3,1}$) and (b) lower layer($C_{NH,3,2}$).
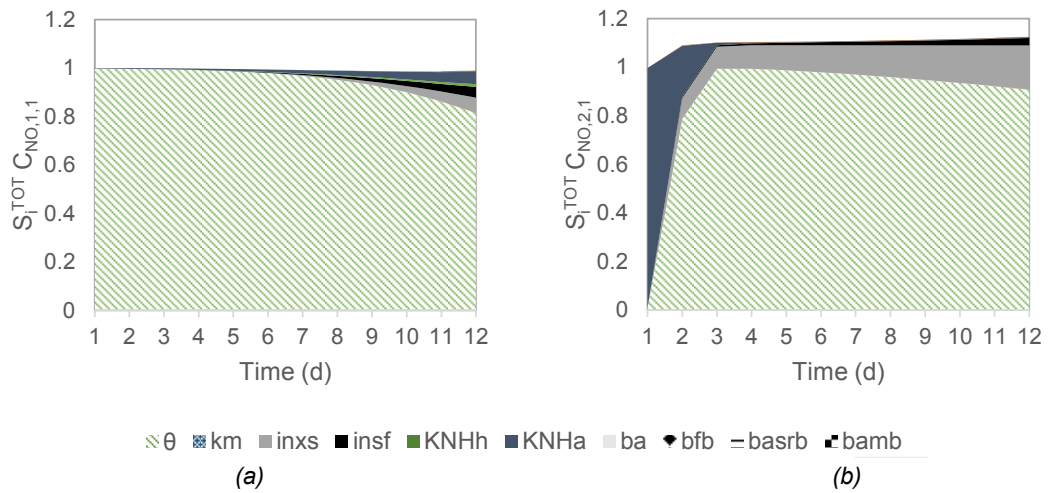

(a)   *(b)*

Figure 3: Total sensitivity indices profiles for nitrate and nitrite nitrogen concentration in (a) the first aerobic pond ($C_{NO,1,1}$) and (b) the second aerobic pond ($C_{NO,2,1}$).
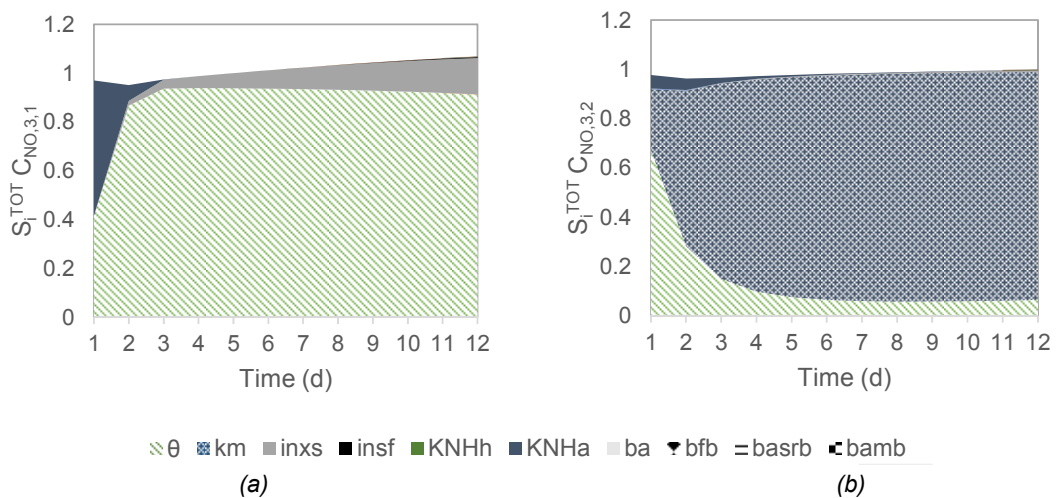

(a)   *(b)*

Figure 4: Total sensitivity indices profile for nitrate and nitrite nitrogen concentration in the facultative pond: (a) the upper layer ($C_{NO,3,1}$) and (b) lower layer($C_{NO,3,2}$).

Saturation/Inhibition coefficient of C_{NH} for C_a, K_{NHha}, accounts for 65% and 95% of the total variance of C_{NH} at the beginning of the process in the second aerobic pond and in the upper layer of the facultative pond, respectively (Figure 1b and Figure 2a). This fact is in concordance with the high growth rate of the autotrophic bacteria at that point of the time horizon ($C_{a,2,1}$ and $C_{a,3,1}$). This parameter also contributes to the total variance of nitrate and nitrite nitrogen of the second aerobic pond, $C_{NO,2,1}$, (Figure 3b) through its first order and interactional effect, and to the total variance of nitrate and dissolved oxygen in the second aerobic pond and in the upper layer of the last pond ($C_{O,2,1}$ and $C_{O,3,1}$) only by its interactional effects. The influence is again appreciated at the beginning of the operating time.

Finally, the variance output is comprised of two more parameter: the Saturation/Inhibition coefficient of C_{NH} for C_h, K_{NHh}, and the fraction of nitrogen in C_s, i_{nxs}, and. The contribution of both parameters is low; the former affects the fermentation products ($C_{SA,2,1}$ and $C_{SA,3,1}$), fermentable, readily biodegradable soluble COD ($C_{F,2,1}$ and $C_{F,3,1}$) and dissolved oxygen ($C_{O,1,1}$) concentration profiles. Whereas, the latter only influences the nitrate and nitrite nitrogen variance output less than 18% along all operating days, as it can be appreciated in Figure 3b and Figure 4a.

4. Conclusions

A proper knowledge of the influence of the model parameters on the state variables has allowed their classification and provides useful information for parameter estimation. GSA allows determining which parameters are the most influential through their effects due to interactions with other parameters and due to first order effects along the entire time horizon. The temperature coefficient θ is the parameter that has the strongest first order influence during the time horizon in the wastewater stabilisation pond system model, which indicates the need of its proper determination. It is important to estimate the most accurate values of other parameters like k_m and K_{NHha} in order to avoid their first order effects that explain 100% of the variance during some intervals within the time horizon of some variables output. Such information is crucial for identifying the set of these three parameters that need to be determined more precisely, as stated in the results section, and allows the identification of the set of parameters which has no contribution to the total output variance.

Acknowledgments

References

Medri, W., Costa, R.H.R., Medri, V., Belli Filho, P., 2007, Stabilization Pond Systems: Cost Estimation for the Treatment of Piggery Waste, Transactions of the American Society of Agricultural and Biological Engineers, 50(4):1409-1414.

Ochoa, M.P., Estrada, V., Hoch, P.M., 2016a, MINLP Wastewater Stabilisation Ponds Synthesis using Rigorous Models under Different Scenarios, Computer Aided Chemical Engineering.

Ochoa, M.P., Estrada, V., Di Maggio, J., Hoch, P.M., 2016b, Dynamic global sensitivity analysis in bioreactor networks for bioethanol production, Bioresource Technology, 200: 666-679.

PSEnterprise Ltd., 2014, gPROMS Model Developer Guide. London: Process Systems Enterprise Ltd.

Rosenthal, R.E., 2016, GAMS- A User's Guide, Washington, DC, USA.

Sah, L., Rousseau, D. P. L., Hooijmans, C.M., Lens, P.N.L., 2011, 3D model for a secondary facultative pond. Ecological Modelling, 222(9), 1592-1603.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S., 2008, Global Sensitivity Analysis. The Primer. John Wiley & Sons, Ltd.

Saltelli, A., Tarantola, S. (2002). On the relative importance of input factors in mathematical models: safety assessment for nuclear waste disposal. Journal of the American Statistical Association, 97, 702–709.

Sobol', I. M. (1993). Sensitivity estimates for nonlinear mathematical models. Mathematical Modelling & Computational Experiment, 1, 407-414.

Sobol', I. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Mathematics & Computers in Simulation, 55, 271-280.

Sobol', I., Shukhman, B. (2007). On global sensitivity indices: Monte Carlo estimates affected by random errors. Monte Carlo Methods and Appl., 13(1), 89-97.