



Meta-modeling based efficient global sensitivity analysis for wastewater treatment plants – An application to the BSM2 model

Al, Resul; Behera, Chitta Ranjan; Zubov, Alexandr; Gernaey, Krist V.; Sin, Gürkan

Published in:
Computers & Chemical Engineering

Link to article, DOI:
[10.1016/j.compchemeng.2019.05.015](https://doi.org/10.1016/j.compchemeng.2019.05.015)

Publication date:
2019

Document Version
Peer reviewed version

[Link back to DTU Orbit](#)

Citation (APA):
Al, R., Behera, C. R., Zubov, A., Gernaey, K. V., & Sin, G. (2019). Meta-modeling based efficient global sensitivity analysis for wastewater treatment plants – An application to the BSM2 model. *Computers & Chemical Engineering*, 127, 233-246. <https://doi.org/10.1016/j.compchemeng.2019.05.015>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Accepted Manuscript

Meta-modeling based efficient global sensitivity analysis for wastewater treatment plants – An application to the BSM2 model

Resul AI , Chitta Ranjan Behera , Alexandr Zubov ,
Krist V. Gernaey , Gürkan Sin

PII: S0098-1354(19)30146-2
DOI: <https://doi.org/10.1016/j.compchemeng.2019.05.015>
Reference: CACE 6449



To appear in: *Computers and Chemical Engineering*

Received date: 5 February 2019
Revised date: 10 April 2019
Accepted date: 7 May 2019

Please cite this article as: Resul AI , Chitta Ranjan Behera , Alexandr Zubov , Krist V. Gernaey , Gürkan Sin , Meta-modeling based efficient global sensitivity analysis for wastewater treatment plants – An application to the BSM2 model, *Computers and Chemical Engineering* (2019), doi: <https://doi.org/10.1016/j.compchemeng.2019.05.015>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Highlights

- Introduces a new framework for building advanced metamodels for efficient GSA.
- Illustrates why Sobol indices method should be used instead of the SRC method.
- Discusses how the accuracy of indices can be improved with cross-validation.
- Provides software to implement the framework in the easyGSA toolbox.

ACCEPTED MANUSCRIPT

Meta-modeling based efficient global sensitivity analysis for wastewater treatment plants - An application to the BSM2 model

Resul Al^a, Chitta Ranjan Behera^a, Alexandr Zubov^a, Krist V. Gernaey^a,
Gürkan Sin^{a,*}

^aProcess and Systems Engineering Center (PROSYS), Department of Chemical and Biochemical Engineering, Technical University of Denmark, Building 229, 2800 Kgs. Lyngby, Denmark

***Corresponding Author:** Gürkan Sin, Associate Professor, Department of Chemical and Biochemical Engineering, Technical University of Denmark, Building 229, 2800 Kgs. Lyngby, Denmark, Email: gsi@kt.dtu.dk, Tel: +45 45252980

Abstract

Global sensitivity analysis (GSA) is a powerful tool for quantifying the effects of model parameters on the performance outputs of engineering systems, such as wastewater treatment plants (WWTP). Due to the ever-growing sophistication of such systems and their models, significantly longer processing times are required to perform a system-wide simulation, which makes the use of traditional Monte Carlo (MC) based approaches for calculation of GSA measures, such as Sobol indices, impractical. In this work, we present a systematic framework to construct and validate highly accurate meta-models to perform an efficient GSA of complex WWTP models such as the Benchmark Simulation Model No. 2 (BSM2). The robustness and the efficacy of three meta-modeling approaches, namely polynomial chaos expansion (PCE), Gaussian process regression (GPR), and artificial neural networks (ANN), are tested on four engineering scenarios. The results reveal significant

computational gains of the proposed framework over the MC-based approach without compromising accuracy.

Keywords: global sensitivity analysis; Sobol method; wastewater treatment plant modeling; polynomial chaos expansions; Gaussian process regression; artificial neural networks.

ACCEPTED MANUSCRIPT

1. Introduction

Over the past decades, computational models describing technical and natural systems have become increasingly complex, requiring incorporation of many model parameters, which add up to model output uncertainty. Sensitivity analysis is widely acknowledged as a good practice to better understand model behavior, providing valuable insights into how much of the model output variance can be attributed to the uncertainty in a specific model parameter. While different methods exist for sensitivity analysis, one class of methods is called global sensitivity analysis (GSA) since these methods look at the model behavior in a global fashion, allowing more than one factor (i.e. an investigated model parameter) to vary at the same time as opposed to one-factor-at-a-time (OAT) methods. Therefore, unlike OAT methods, GSA methods can account for and quantify the effects of factor interactions on the investigated model outputs. Although use of OAT methods in sensitivity analysis is fiercely criticized (Saltelli et al., 2017), OAT is still the most widely used technique in sensitivity analysis (Ferretti et al., 2016). Among two different classes of GSA methods are the variance decomposition based methods such as the Sobol indices method (Sobol, 2005), and regression based methods such as standardized regression coefficients (SRC) (Saltelli et al., 2008). Sobol sensitivity indices are considered as the reference measures among practitioners, setting the benchmark for all other sensitivity analysis methods. The Monte Carlo

simulations approach has been traditionally employed to evaluate Sobol sensitivity indices. Although this approach provides robust results, one major drawback is its high computational cost, requiring a large number of model evaluations, typically in the order of 2,000 to 10,000 to ensure convergence of the indices with a satisfactory precision level (Burnaev et al, 2017). Therefore, to reduce the computational burden associated with the Sobol method, alternative ways of calculating Sobol indices, such as metamodeling-based approaches, have gained significant interest in recent years. Among the most widely used algorithms of surrogate models for global sensitivity analysis are polynomial chaos expansions (Sudret, 2008), Gaussian processes regression (Marrel et al., 2009), and artificial neural networks (Li et al., 2016).

In the field of wastewater treatment modeling, the importance of performing global sensitivity analysis has been widely recognized with applications on both process models and plant-wide frameworks. In some of the earliest applications of the Monte Carlo method, Flores-Alsina et al. (2008) investigated the effect of activated sludge input uncertainties on the control strategies of the BSM2 plant, whereas Sin et al. (2009) identified and analyzed different sources of uncertainties in the BSM1 plant. Using the SRC method, Sin et al. (2011) also performed a global sensitivity analysis on the BSM1 plant to identify the most significant parameters influencing the plant performance criteria. Flores-Alsina et al. (2012)

applied the SRC method to identify and rank the most important design variables of activated sludge process plants. Rojas and Zhelev (2012) used GSA to identify the operating conditions with the strongest impact on the energy requirements of a thermophilic aerobic digestion system, on which they perform energy efficiency optimization using the optimization variables identified by the GSA. Cosenza et al. (2013) comparatively applied SRC, Morris screening and Extended-FAST methods to a membrane bioreactor (MBR) model and reported that Morris screening provides inconsistent results in comparison to the other two methods. The same authors also studied variance decomposition based methods in WWTP modeling and concluded that for model outputs like ammonia, nitrate and phosphorus, interactions among input factors provided significantly different sensitivity results compared to regression-based methods, such as SRC, which do not take interactions into consideration (Cosenza et al., 2014). Ramin et al. (2014) studied first order and second order secondary settling tank models within the BSM2 plant and conducted a global sensitivity analysis using the SRC method and Morris screening. Dragan et al. (2017) used the Morris screening method to identify significant design degrees of freedom of the BSM2 plant, which are used to perform design optimization. Mannina et al. (2018) used the SRC method, albeit with a poor linearization of the MC simulation results ($R^2 < 0.7$), to identify calibration parameters of an integrated activated sludge and membrane bioreactor

model. Ochoa et al. (2016) studied process synthesis of wastewater treatment network design and applied GSA using the Sobol method on the kinetic dynamic model parameters of the optimal network configuration to identify opportunities for further design optimization. More recently, Fortela et al. (2019) utilized a Morris screening-based GSA method along with principal component analysis to identify sensitive biochemical mechanism parameters of anaerobic digestion kinetic models.

Although the SRC method is widely used by researchers, benchmarking of the reported sensitivity measures with more reliable variance decomposition based methods like the Sobol indices method is necessary to strengthen the conclusions drawn from the global sensitivity studies, especially in the context of large WWTP modeling case studies. However, to the best of our knowledge, the Sobol indices method has thus far not seen an application in plant-wide models of WWTPs, such as the BSM2, mainly because of the high computational cost associated with calculation of these indices. As the Sobol indices method provides researchers with more reliable sensitivity measures compared to the popular SRC method, there is a need for a systematic framework that will extend the knowledge on construction of efficient surrogate models to be used for global sensitivity analysis purposes. The main objective of this paper is, therefore, to present a systematic methodology to perform efficient Sobol global sensitivity analysis of complex biological systems, such as WWTPs, with

the help of advanced meta-modeling techniques, such as polynomial chaos expansions, Gaussian processes regression, and artificial neural networks. For this purpose, four different scenarios including various sources of epistemic uncertainties are framed: (1) uncertainty in the influent fractionation, (2) uncertainty in stoichiometric and kinetic model parameters, (3) uncertainty about the hydraulics and design-related parameters of the plant, and (4) simultaneous realization of all the uncertainties in scenarios (1), (2), and (3). The efficiency and the accuracy of the different approaches for global sensitivity analysis are compared and discussed.

This paper is organized as follows. First, we present traditional global sensitivity analysis methods as well as more recent meta-modeling based methods. Second, we introduce four practical scenarios in more detail, in which a global sensitivity analysis is needed in order to address the questions which design engineers encounter. Then we explain the plant model and simulation strategy used in this study. Next, we introduce the proposed framework for construction of efficient surrogate models with the different steps being detailed and the implemented techniques briefly described. Then, we discuss the model development statistics of the PCE, GPR and ANN models, and we demonstrate the advantages of using the Sobol indices method over the SRC method. Finally, the effectiveness of the meta-modeling-based approach for GSA is illustrated with the results of

scenario analysis: the most influential parameters in all scenarios are identified, and are then discussed in more detail.

2. Methods

2.1. Global sensitivity analysis methodologies

Global sensitivity analysis is established as a powerful tool for quantifying how much of the variance of an output of interest is due to the uncertainty in the input parameters of a mathematical model. There exist a wide array of techniques and sensitivity measures available in the literature, e.g. the Morris screening method (Morris, 1991), linear regression-based methods, variance-based methods (Saltelli et al., 2008), to name a few. For a recent review of these methods, readers are referred to the studies of (Iooss and Lemaître, 2015; Pianosi et al., 2016). Among the most commonly used methods in GSA are standardized regression coefficients (SRCs) and Sobol sensitivity indices. SRCs provide a measure of sensitivity using a linear regression approximation of the model response, whereas the Sobol sensitivity indices approach decomposes the total model output variance into each single input variance and the combinations thereof. This method is especially useful for the case of highly nonlinear computational models. Both methods employ Monte Carlo simulations, a technique which is discussed below together with the two GSA methods.

2.1.1. Monte Carlo simulations approach

Monte Carlo simulation (MCS) is a universal mathematical technique relying on repeated random sampling to obtain numerical results. It has found wide use in fields as disparate as finance, energy, uncertainty, and sensitivity analyses. Calculation of Sobol sensitivity indices using MCS is a relatively well-established procedure and can be divided into 4 steps: (1) specification of uncertainty ranges for the input parameters; (2) sampling of the parameters within their ranges using a sampling algorithm such as Sobol sampling, Latin hypercube sampling, etc.; (3) propagation of uncertainty by repeated runs of the simulator for each combination of sampled parameters in the input space; (4) post-processing of the acquired dataset of quantities of interests. Regardless of the complexity of a single simulator run, the MCS procedure remains the same and the repeated runs are independent from one another, which makes it inherently suitable for parallelization. However, the technique has a serious efficiency drawback, as the size of the sampling matrices needed to converge to accurate results for sensitivity indices is no less than 10^3 , typically in the order of 10^4 (Burnaev et al., 2017). This results in an unaffordable number of model runs in practice. Furthermore, slightly varying estimators are suggested for the numerical computation of Sobol sensitivity indices with Monte Carlo simulations. A comparison of these can be found in (Saltelli

et al., 2010), who have found the Jansen's estimator more accurate, which is therefore selected as the estimator of the Sobol sensitivity indices calculated from the MCS method in this study.

2.1.2. Standardized regression coefficients (SRC)

The SRC method involves obtaining model outputs by performing Monte Carlo simulations and fitting a linear regression model to the Monte Carlo simulation output using the input variables $X_{1,\dots,k}$ as in the following functional form.

$$y = b_0 + \sum_{k=1}^n b_k X_k \quad (1)$$

where the coefficients b_0, \dots, b_k are determined by least-squares minimization of the differences between the y -values produced by the regressed linear model and the actual model output produced by the Monte Carlo simulation. By scaling the raw coefficients b_0, \dots, b_k with the standard deviations of the input variables and the output, one can obtain the SRCs as follows.

$$SRC_k = b_k \frac{\sigma_{X_k}}{\sigma_y} \quad (2)$$

The absolute value of SRC_k provides a direct interpretation of the individual effect of input variable X_k on the output variance; the higher the value, the stronger the influence. However, one limitation of this metric is that it does not quantify interaction effects, i.e. the amount of variance caused on the output due to the interactions of input variables X_k . For a linear model, the sum of squares of the SRCs should add up to 1 ($\sum_k SRC^2 = 1$), whereas for non-linear models, this sum amounts to the coefficient of determination R^2 . According to Saltelli et al. (2008), the SRC is a valid measure of sensitivity if the coefficient of determination (R^2) of the linear model is higher than 0.7, which is the fraction of the output variance of the MCS data explained by the regressed model. For cases where R^2 is below 0.7, the use of SRC as a sensitivity measure comes at the risk of being ignorant of the $100 \times (1 - R^2)$ % of the original output variance. Calculation of these SRCs is relatively straightforward using any regression software, and one example is to use the `fitlm` regressor in the MATLAB computing environment for generating the SRCs.

2.1.3. Sobol sensitivity indices (S_i and S_{Ti})

Sobol indices provide valuable information about both individual effects of each input variable on the output variance, as well as interaction effects among input variables. Among the most commonly used indices are first order Sobol indices S_i and total order Sobol indices S_{Ti} . The first order

sensitivity indices can be interpreted as the expected amount of variance that can be subtracted from the total output variance, if we decide to fix the value of an input variable within its uncertainty range. It can then be formulated as conditional variance over the unconditional variance as follows.

$$S_i = \frac{V[E(Y|X_i)]}{V(Y)} \quad (3)$$

The first order indices show the individual impact of each input variable on the uncertainty of the model output. They are indicative of the individual contribution of each input, excluding any contribution that may arise from interactions with other inputs. They are also called main effects. The sum of all first order indices should be equal to 1 for perfectly additive (linear) models and less than 1 for non-additive (nonlinear) models. The difference $(1 - \sum_i S_i)$ is an indicator of the presence of interactions between input variables (Saltelli et al., 2008).

Total sensitivity indices, on the other hand, quantify the total effect of an input variable on the model output (accounting also for its interactions with other input variables), which can be interpreted as the expected amount of output variance that would remain unexplained if only that input

variable was allowed to vary over its uncertain range. The expression for the total sensitivity indices reads,

$$S_{Ti} = 1 - \frac{V[E(Y|X_{-i})]}{V(Y)} \quad (4)$$

The sum of all total indices should be equal to 1 for perfectly additive (linear) models, and always higher than 1 for non-additive models (Saltelli et al., 2008). Another interesting property of Sobol indices is that the difference $(S_{Ti} - S_i)$ can be regarded as an indicator of the strength of interactions between input variables. For definitions of the first and the total indices when there are dependent inputs, readers are referred to the study of Mara et al., (2015). Traditionally these indices are calculated using nested Monte Carlo simulations (e.g. Janssen, Sobol or Saltelli approximation) for individual variance contributions. However, more recent approaches employ metamodels, such as polynomial chaos expansion and Gaussian process regression and artificial neural network models. We discuss these three approaches below.

2.1.4. Metamodeling-based approaches

To overcome the aforementioned computational cost of the MCS approach, metamodeling-based approaches are developed. Metamodeling, also referred to as surrogate modeling, has received increased attention in recent years as computational models became much more sophisticated

with the advancements in computing power. Metamodels are computationally undemanding alternative models of usually complex simulation models and they come in many types with numerous engineering applications. Extensive discussions on how to construct these models can be found elsewhere (Forrester et al., 2008). We discuss three classes of metamodels that are commonly used for global sensitivity analysis purposes, namely, polynomial chaos expansions (PCE), Gaussian processes regression (GPR), and artificial neural networks (ANN). This selection of metamodels is mainly because these models were found to be particularly effective for GSA application in an earlier study (Al et al., 2018), which also analyzed radial basis function (RBF) interpolation, multivariate adaptive regression splines (MARS) comparatively. Readers are referred to the original work for more discussions of the performance comparison of the other models. Besides these metamodels, relatively newer studies also proposed the use of orthogonal augment radial basis functions (Wu et al., 2019) and support vector regression (Cheng et al., 2019) for performing efficient global sensitivity analysis.

2.1.4.1 Polynomial Chaos Expansions

Polynomial chaos expansions are polynomial approximations of model responses generated by complex simulators. To emulate a random model response $y = M(x)$ from N independent random variables $x = \{x_1, x_2, \dots, x_N\}$, PCE assumes the following functional shape,

$$y^{PC} = M^{PC}(x) = \sum_{\alpha \in A} S_{\alpha} \Psi_{\alpha}(x) \quad (5)$$

using the following chaos representation,

$$y^{PC} = s_0 \psi_0 + \sum_{i=1}^N s_i \psi_1(x_i) + \sum_{j \geq i}^N s_{ij} \psi_2(x_i, x_j) + \sum_{k \geq j}^N s_{ijk} \psi_3(x_i, x_j, x_k) + \dots \quad (6)$$

where ψ_{α} is an α -th order multivariate orthogonal polynomial and S_{α} are the polynomial coefficients. By definition, the orthogonality property of the polynomials translates to an ANOVA decomposition (Fajraoui et al., 2011). Each multivariate polynomial ψ_{α} in this representation is a tensor product of univariate polynomials. These univariate polynomials are selected based on the distribution of the input variables, e.g. Legendre polynomials for uniform distributions and Hermite polynomials for standard Gaussian distributions. More details about these polynomials can be found elsewhere (Marelli and Sudret, 2016). For practical computations, the chaos representation is truncated to a simpler form in such a way that one only retains those basis polynomials with degrees not exceeding a user-defined degree p . Thus the number of unknown coefficients to be calculated, denoted by N_p , leads to

$$N_p = \binom{N+p}{p} = \frac{(N+p)!}{N!p!} \quad (7)$$

Least squares minimization, among other techniques, is often used to estimate PC coefficients by minimizing the mean square residual. However, in order to avoid an ill-conditioned regression problem, one needs to have an initial experimental design of a size greater than N_p , which increases factorially with respect to N and p . To address this, a degree adaptive algorithm, which tries to choose the best polynomial degree given the experimental design, based on least angle regression (LAR), is proposed to create a sparse PC expansion consisting of the significant coefficients of the full PC expansion (Blatman and Sudret, 2011). Once this sparse PC approximation of the model response is constructed, Sobol sensitivity indices may then be computed analytically from the PC coefficients S_α . The generalization error of the resulting PC model can then be estimated by the leave-one-out (LOO) cross-validation error ϵ_{LOO} as follows

$$\epsilon_{LOO} = \frac{\sum_{i=1}^N (M(x^i) - M^{PC \setminus i}(x^i))^2}{\sum_{i=1}^N (M(x^i) - \mu_y)^2} \quad (8)$$

where $M^{PC \setminus i}(x^i)$ is the prediction of the model constructed from an experimental design excluding x^i and $M(x^i)$ is the actual model response at the design point x^i . The upper part of this expression is also called predictive error sum of squares (PRESS) whereas the denominator is a

total sum of squares. The corresponding coefficient of determination, with analogy to R^2 , is often called the leave-one-out coefficient Q^2 and defined as follows.

$$Q^2 = 1 - \epsilon_{LOO} \quad (9)$$

Unlike R^2 , Q^2 does not tend to 1 as the number of model parameters grows, thereby providing a better measure for model selection. The UQLab framework written in the MATLAB environment provides implementations of PCE models, and was used in this study to develop these models (Marelli and Sudret, 2016).

2.1.4.2 Gaussian Process Regression

Gaussian process regression (GPR) is a Bayesian probabilistic nonlinear regression algorithm, which uses kernels to explain a given model response as a realization of a random function of the following shape,

$$y^{GP} = M^{GP}(x) = \beta f(x) + \sigma^2 Z(x, \omega) \quad (10)$$

where the first term $\beta f(x)$ is the mean value of the Gaussian process, σ^2 is the variance and $Z(x, \omega)$ is a zero mean unit variance stochastic

process. β denotes regressed coefficients and $f(x)$ a set of basis functions (i.e. constants, polynomials, etc.). The error variance σ^2 and the coefficients β are estimated from the experimental design, i.e. dataset of simulation inputs and outputs used to construct the GPR model. The crucial information in a GPR model is contained in its stochastic part, $Z(x, \omega)$, which introduces latent variables to correlate observations x and new points x' in the input space with correlation functions (also referred to as kernel or covariance functions in the literature). The most commonly used correlation function types include squared exponential kernel, exponential kernel, matern 3/2, matern 5/2 and rational quadratic kernel. For more details about Gaussian process regression and kernel functions, the reader is referred to the book of Rasmussen and Williams (2006). With a wide array of available correlation function types that can be used to describe the similarity between two samples in a space, GPR provides modelers with a powerful surrogate model capable of approximating highly nonlinear random responses. The Statistics and Machine Learning toolbox of the MATLAB software (The MathWorks), which has been used for this study, provides users with an implementation of different types of GPR models. Once a GPR model is built from an experimental design, one can then use this model instead of the computationally costly original model to calculate Sobol sensitivity indices following the Monte Carlo simulations approach as described above.

2.1.4.3. *Artificial Neural Networks*

Artificial neural networks, initially inspired by biological neurons, have attracted widespread attention in recent years especially due to their successful applications in highly complex machine learning tasks, such as image classification, speech recognition, computer vision, etc (Géron, 2017). Feedforward neural networks (FNN), where connections between the nodes do not form a cycle such as in recurrent neural networks (RNN), are among the earliest and the simplest types of neural networks. One example of such networks is the multilayer perceptron (MLP), where there are at least 3 layers (input, hidden, and output layers) with interconnecting transfer functions (also referred to as activation functions) such as sigmoids, logistic function, hyperbolic tangent, etc. Figure 1 demonstrates a single hidden layer architecture of a feedforward MLP. If more than one hidden layers are added in the network architecture, it is then called a deep neural network (DNN). Due to their highly versatile and scalable architectures, they are very effective at capturing vastly nonlinear relationships in datasets, making them an ideal candidate also for generation of surrogate models. One of the most widely used training algorithms for neural networks is the backpropagation algorithm (Rumelhart et al., 1986), in which a prediction is made for each training observation in a forward pass, and the error contribution from each connection in the network is calculated in a reverse pass to adjust the corresponding

weights so that the error in the final prediction is reduced. The Deep Learning Toolbox of the MATLAB software provides implementations of the backpropagation algorithm variants, which are used to develop these models for this study as explained in section 3.1.

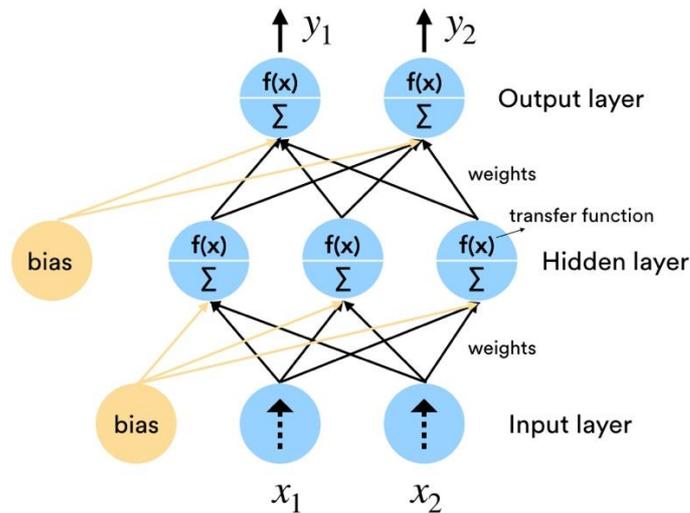


Figure 1: A network architecture of a feed-forward MLP containing a single hidden layer and an output layer is shown. Input variables (x_1 , and x_2) are passed through the input layer, and the hidden and the output layers (also being affected by the bias and the weight terms) to produce the network outputs (y_1 , and y_2).

2.2. Scenarios for global sensitivity analysis

To elucidate the application of the metamodeling-based global sensitivity analysis of WWTPs, four realistic scenarios are postulated as shown in Table 1. In scenario 1, the goal is to quantify the sensitivity of the plant key performance indicators (KPIs); such as methane gas production, aeration energy demand, etc., to the uncertainty in data regarding the incoming influent wastewater characterization, which is represented with a

set of influent fractionation parameters and their uncertainty ranges as shown in Table S1. In scenario 2, the emphasis is put on the investigation of the parametric sources of uncertainty, namely stoichiometric and biokinetic parameters of the ASM1 model, and their effects on the plant KPIs. Table S2 shows the list of those parameters and their uncertain bounds. In scenario 3, the objective is to examine the design and hydraulics related parameters (listed in Table S3) of the plant and their effects on the KPIs. In scenario 4, all different sources of uncertainty considered in the previous scenarios are combined in order to identify the most influential parameters and quantify their effects on the plant KPIs. More detailed descriptions of all scenarios are also included in the supplementary material.

Table 1: Different sources of uncertainties considered in the 4 scenarios to be studied

Scenarios	Sources of uncertainty		
	Influent data	Kinetics and stoichiometry	Hydraulics and design related parameters
1	Uncertain	Certain	Certain
2	Certain	Uncertain	Certain
3	Certain	Certain	Uncertain
4	Uncertain	Uncertain	Uncertain

2.3. The model: BSM2 plant layout, simulation strategy and plant performance evaluation

The widely-used Benchmark Simulation Model No. 2 framework, originally developed by Jeppsson et al. (2006), was taken as the base WWTP layout and is further modified as shown in Figure 2. The final modified plant layout consists of a primary clarifier (PC), an activated sludge (AS) process, and a secondary clarifier (SC) as the mainstream technologies, whereas a thickener, an anaerobic digester (AD), and a dewatering unit are used to treat the sludge coming out of the AS and PC units. The AS tanks are modelled using the Activated Sludge Model No. 1 (ASM1) (Henze et al., 1987). The first two tanks are anoxic (aiming at denitrification), whereas the remaining three tanks are aerobic (aiming at nitrification and COD removal). The secondary clarifier (to separate active biomass from effluent) is modelled as a non-reactive system using an exponential settling velocity function proposed by Takács et al., (1991), while the AD unit is modelled using Anaerobic Digestion Model No. 1 (ADM1) (Batstone et al., 2002). In the modified layout, the waste activated sludge is discharged from the last aerobic tank (adopted from other studies (Behera et al., 2018; Boiocchi et al., 2017; Sin et al., 2009)) instead of the SC unit as in the original BSM2 configuration. Further details about the employed modeling and simulation strategy can be found in the supplementary material.

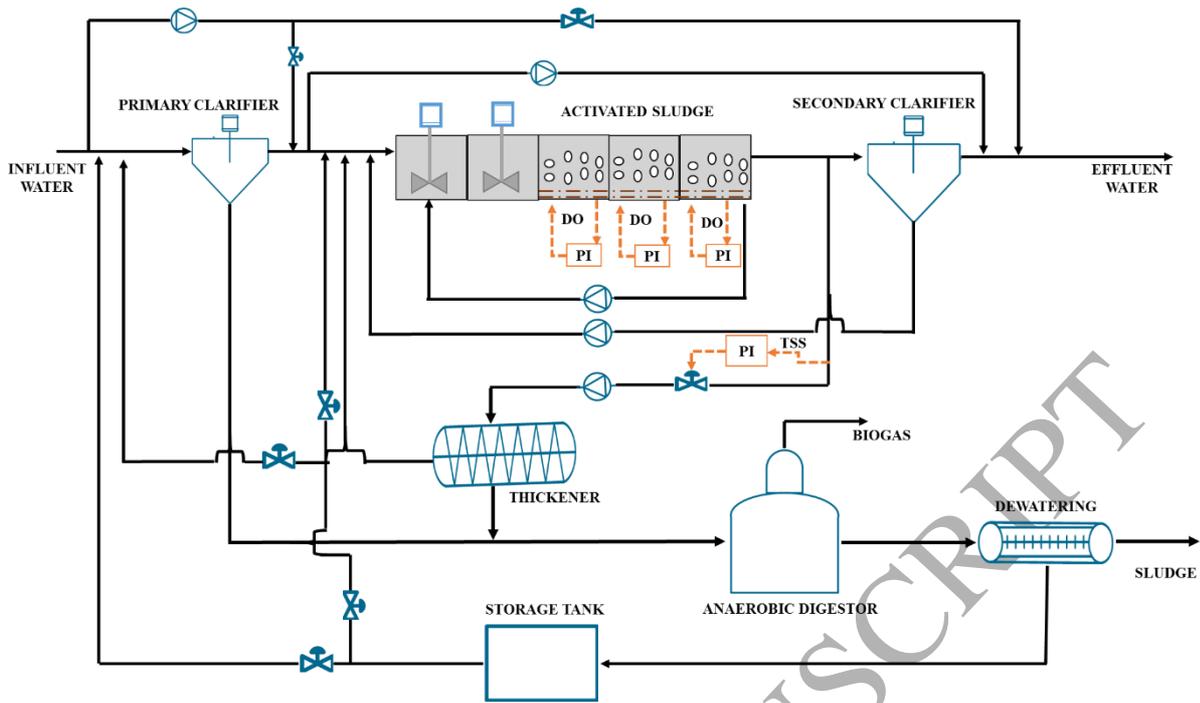


Figure 2: Schematic overview of the modified Benchmark Simulation Model No. 2 (BSM 2) framework used for global sensitivity analysis.

The sensitivity analysis results of each scenario are evaluated based on 6 different plant performance indicators; namely effluent quality index (EQI), methane gas production in the AD tank, aeration energy demand by the AS tank, net sludge discharge from the plant, effluent ammonium, and effluent nitrate concentrations. The EQI definition is adopted from the BSM2 publications (Germaey et al., 2014; Jeppsson et al., 2006), and it represents a sum of pollutant loads of the effluent. It is important to note that the EQI is calculated for one day (average load used for steady state simulation), and not for a longer period (i.e., for the last 364 days) as used for dynamic simulation of the BSM2 (Germaey et al., 2014; Jeppsson et al., 2006). The aeration energy demand in the AS tank is calculated using the correlation suggested in (Germaey et al., 2014). The details of

the model equations used for calculating the above-mentioned performance indicators (such as EQI, AE demand etc.) are provided in the supplementary material.

2.4. Generic framework for meta-model development

A generic computational framework for building various types of surrogate models to be used for global sensitivity analysis is presented in Figure 3. First, a set of parameters that are subject to sensitivity analysis are chosen to create an input space. All the available information regarding bounds and distributions of these parameters can be introduced at this step. Second, an experimental design is created using a sampling algorithm, such as space-filling designs like Latin Hypercube sampling (LHS) (McKay et al., 1979), as well as quasi-random sequences with uniformity properties like Sobol and Halton sequences. A recent comparative study found the best results when using samples generated with Sobol compared to LHS and Halton designs (Davis et al., 2018). More efficient designs can also be sought with the use of adaptive algorithms, such as those defined in (Burnaev et al., 2017; Steiner et al., 2019). Monte Carlo simulations are employed to generate the input and output pairings of the experimental design. This experimental design is then fed into the surrogate model development algorithms to train surrogate models. One often needs to estimate the generalization capability of these models before their use. Among the techniques that can be used

for this purpose are the holdout validation method, the k-fold cross validation method, the leave-one-out cross validation method, and the Monte Carlo cross validation method. For a detailed discussion of these methods, readers are referred to the work of (Molinaro et al., 2005). The training sample size directly determines the total computational cost of the surrogate model development. An appropriate size of this sample is decided by the framework, which monitors the evolution of a user-defined target accuracy on an error metric as the size of the experimental design is incrementally increased.

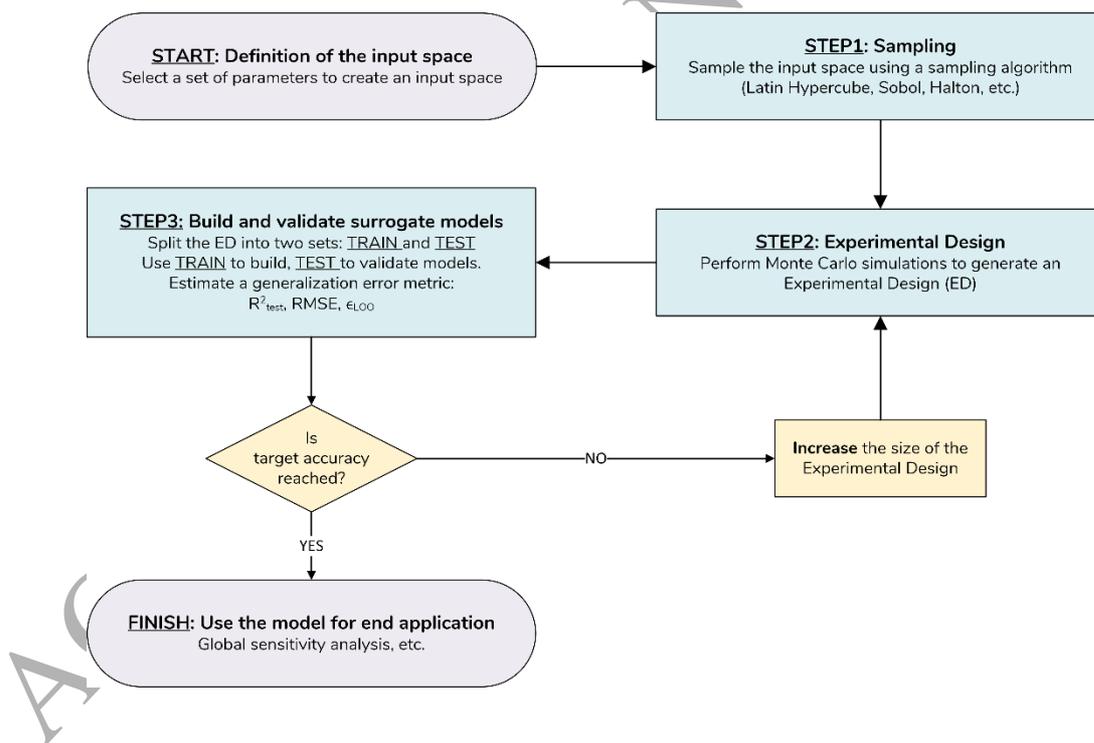


Figure 3: The flowchart of the generic framework for building surrogate models. An iterative loop is added to ensure a user-set target model accuracy by incrementally increasing the size of the experimental design.

It should also be noted here that the generic feature of the framework allows for the construction of all different metamodels from the same dataset without a further need for tuning of the dataset. On the other hand, each metamodel, which is coded as functions returning a model given a dataset, has its own corresponding model development algorithm to identify an optimal metamodel structure and its parameters. For instance, hyperparameters of a GPR model, such as kernel function specific parameters, are optimized using a Bayesian solver to improve its performance of cross validation statistic (Rasmussen and Williams, 2006).

3. Results and discussions

3.1. Performance statistics of PCE, GPR, and ANN models

PCE models were developed following the proposed framework, which sequentially increases the experimental design size and the maximum allowable polynomial order until a target accuracy ($0.95 < Q^2$) is reached.

Table 2 shows the performance statistics of the models developed for scenario 1, in which 7 different parameters were considered in the input space. The orders of the final selected PCE models are relatively moderate, although the maximum order limit imposed was 20. The size of the experimental design was initially set to 100, with an incremental size increase of 50, to a maximum of 250 samples, which was hit by the framework for developing models of effluent ammonium and effluent quality index for the scenario 1 (see

Table 2) as the target accuracy was not reached. The leave-one-out error metric ϵ_{LOO} (as defined in Eq. 8) is used to select the best PCE models in the adaptive sparse PC expansion algorithm since it gives a better estimate of the model's accuracy compared to R^2 , which systematically tends to 1 as the number of model parameters increases (Blatman and Sudret, 2011). The adaptive algorithm allows creation of highly accurate models with a few hundred samples due to the sparse expansion, which calculates only a low number of coefficients (see the number of nonzero coefficients in

Table 2). Statistics of the models developed for scenarios 2, 3, and 4 are also included in the supplementary material.

Table 2: Performance statistics of PCE models developed for the scenario 1.

PCE model output	Size of experimental design	PCE order p	Coefficient of determination R^2	Leave-one-out coefficient Q^2	Number of nonzero coefficients
Effluent nitrate	150	6	0.988	0.950	37
Effluent ammonium	250	5	0.955	0.900	26
Effluent quality index	250	5	0.976	0.910	48
Sludge disposal	100	2	0.999	0.999	21
Aeration energy	100	4	0.993	0.968	28
Methane production	100	5	0.999	0.999	24

GPR models are developed following the same framework. The algorithm incrementally increases the experimental design and for each different size of the experimental design, the algorithm seeks the best kernel function among a family of available alternatives and performs a hyperparameter optimization. The target accuracy was set to obtain models with a cross-validation score ($0.95 < Q^2$). Similar to the PCE model development, the chosen increment was 50. An increase in the size of the experimental design is only made if the desired accuracy was not reached. Table 3 shows the statistics of the final selected models for scenario 1 outputs. For more details of the different kernel functions, readers are referred to the book of Rasmussen and Williams (2006). The supplementary material also contains the statistics of the models developed for the other scenarios.

Table 3: Performance statistics of GPR models developed for the scenario 1.

GPR model output	Size of experimental design	Coefficient of determination R^2	Leave-one-out coefficient Q^2	Kernel function
Effluent nitrate	100	1	0.977	ARD Matern 3/2
Effluent ammonium	150	0.999	0.976	Squared Exponential
Effluent quality index	150	1	0.984	ARD Matern 5/2
Sludge disposal	100	0.999	0.999	ARD Matern 3/2
Aeration energy	100	1	0.994	ARD Matern 3/2
Methane production	100	0.999	0.999	ARD Matern 3/2

ANN models are built following a grid search approach which tries to find the best combination of user-configurable network parameters such as the number of neurons in each layer, training algorithms, and transfer functions, etc. The Deep Learning Toolbox of the MATLAB software is used to develop the ANNs with backpropagation training algorithm variants, i.e. Levenberg-Marquardt (LM), Bayesian regularization (BR), scaled conjugate gradient (SCG), and conjugate gradient with Powell-Beale restarts (CGB) were all tested for the each output. 5-fold cross-validation is used to decide on the right size of the training sample, i.e. further increase in the sample size is avoided if the target accuracy (cross-validation R^2 above 0.95) is reached. One fifth of the training sample is used as a validation dataset while the rest is used for training. Bayesian regularization is most often selected as the training algorithm by the framework whereas the size of the hidden layer (i.e. number of hidden nodes) is allowed to vary from 10 to 15 in order to keep the number of network connections at a reasonable level. Table 4 shows performance scores and configuration details (number of nodes and the training algorithm) of the final selected models for the scenario 1. The supplementary material also contains the statistics of the models developed for the other scenarios.

Table 4: Performance statistics of ANN models developed for the scenario 1.

ANN model output	Size of experimental design	Training R^2	5-fold cross-validation R^2	Configuration selected
Effluent nitrate	150	0.999	0.978	7×12×1 with BR
Effluent ammonium	150	1	0.992	7×10×1 with BR
Effluent quality index	150	1	0.994	7×10×1 with BR
Sludge disposal	100	0.999	0.998	7×15×1 with BR
Aeration energy	150	0.999	0.993	7×11×1 with BR
Methane production	100	0.999	0.998	7×14×1 with BR

It is significant to emphasize that the accuracy of the above-shown metamodelling are valid in the entire domain of their input parameter space, as they are developed for global sensitivity analysis purposes in this contribution. This means that for different application purposes, the framework needs to be aligned accordingly. For example if one is interested in understanding the use of metamodelling in a local sensitivity analysis context (i.e. how much a small variation of a single input parameter affects the outputs), then the scope of design of experiments in the sampling stage of the framework needs to be aligned with this purpose. In that case, one should define a narrow range of input space around the nominal point where local sensitivity analysis is to be performed and construct a metamodel fit for this purpose.

3.2. Comparison of SRCs and first order Sobol indices

According to Saltelli et al. (2004), standardized regression coefficients can be used as a valid measure of sensitivity if the coefficient of determination (R^2) of the linear regression of the Monte Carlo simulation results has a value higher than 0.7. Similar to first order Sobol indices, for a perfectly linear model, summation of the squared SRCs should add up to 1. Therefore, this sensitivity measure is comparable to Sobol first order indices. In order to show a comparison of these two sensitivity measures, Monte Carlo simulation results of scenario 1 are used to calculate SRCs of the input parameters. Figure 4 shows a comparison of squared SRCs with first order Sobol indices obtained from the same Monte Carlo simulations. Also shown are the R^2 values obtained from the regressed linear models. For low values of R^2 , such as effluent ammonium, the two sensitivity measures significantly diverge, whereas for high values of R^2 , such as sludge production, they converge to the same values and rankings. However, for effluent ammonium, the SRC method results in a different ranking of the most significant parameters compared to the Sobol method. As the Sobol method relies on variance decomposition, the measures obtained from this method are more reliable compared to regression-based methods whose accuracy depends on the percent of the output variance explained by the fitted linear model, i.e. R^2 . These results clearly show why the Sobol method should be preferred over SRC as a more reliable method for global sensitivity analysis.

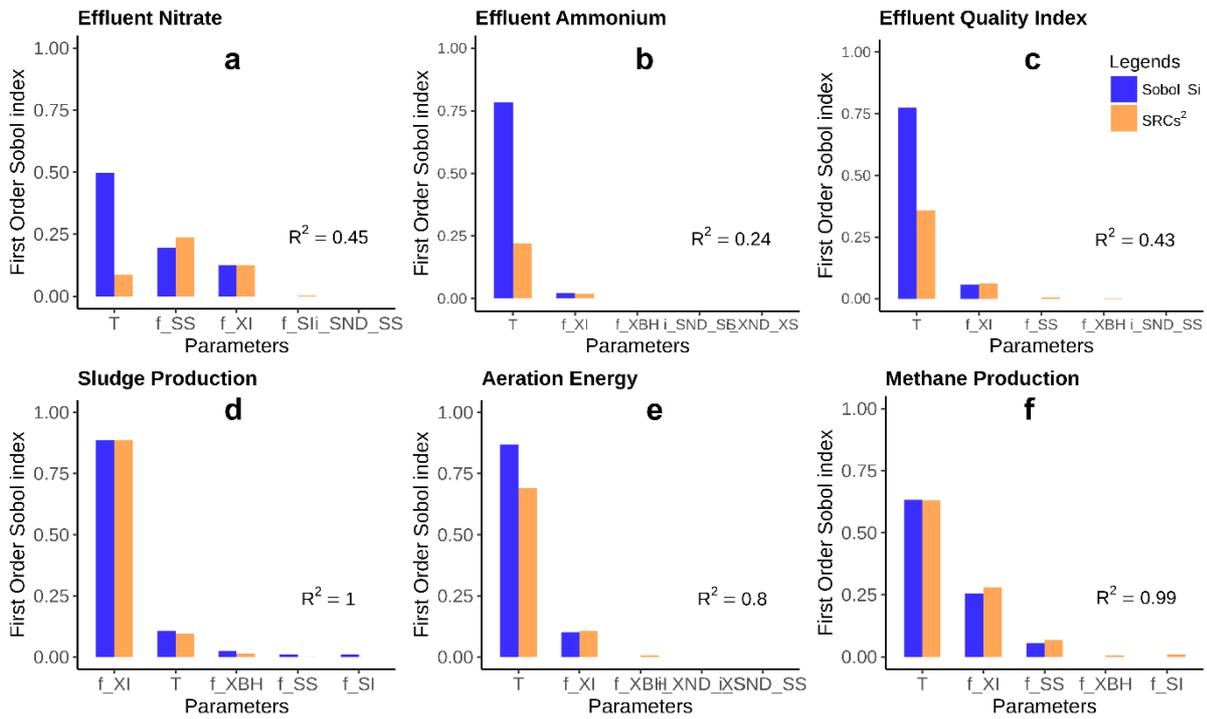


Figure 4: Comparison of standardized regression coefficients (squared) with first order Sobol indices. As the coefficient of determination of the linear models decreases, SRC provides increasingly diverging indices compared to the corresponding Sobol indices.

In addition, the varying performance results of the linear models also imply that certain KPIs of the system, such as effluent ammonium and nitrate concentrations, exhibit a nonlinear behavior with the chosen set of input parameters. For instance, the input-output relationships between the influent temperature and effluent ammonium as well as effluent nitrate are visualized in Figure S1 and Figure S2, respectively. As there is a clear nonlinear type correlation, the predictions from the linear model will diverge from the true simulation results. Figure 5 shows a comparison of the linear model predictions of effluent nitrate to those obtained from PCE and GPR models. This result implies a limitation to the amount of relational

information that can be extracted from a dataset due to the choice of functional form of the metamodel. Overall, this emphasizes the importance of systematically exploring different functional forms (metamodel structure) to describe input-output data relations as presented in this framework.

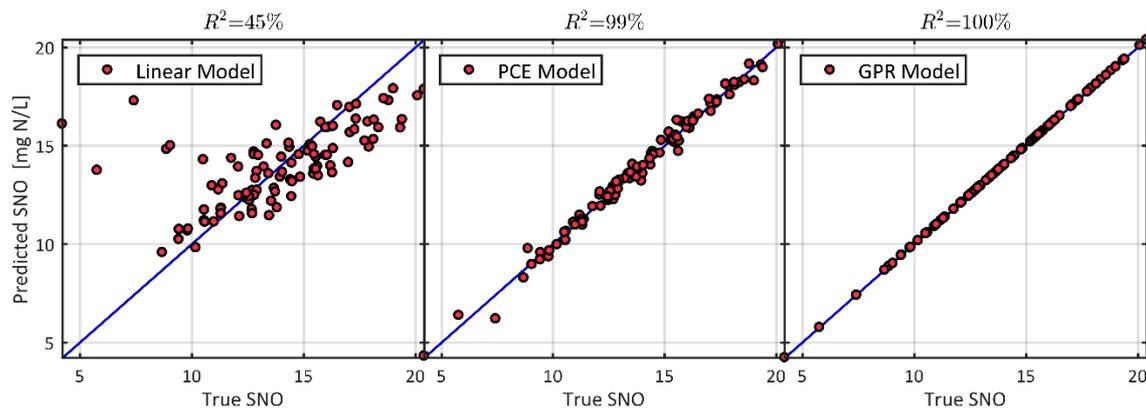


Figure 5: Predictive performance of the linear model used by the SRC method is compared to more advanced metamodels of type PCE and GPR for the scenario 1 results of effluent nitrate concentration.

3.3. Comparison of Monte Carlo and meta-model-based approaches for calculation of Sobol indices

3.3.1. Sensitivity analysis of the uncertain influent fractionations – Scenario 1

From the sensitivity analysis results shown in Figure 6 for scenario 1, it is clear that the most significant parameters are temperature (T) and the influent fractions, f_{SS} and f_{XI} respectively. For nitrification, temperature is the most influential input, capturing more than 95% of the output variation, and this is in line with process understanding i.e., the effect of T on the nitrification rate is described by an Arrhenius equation where the rate increases rapidly with temperature. Likewise, the effluent nitrate

concentration depends on the denitrification rate which inherently relies on the availability of biodegradable COD (both soluble and particulates), and the heterotrophic growth rate. The sensitivity analysis also supports the development of improved process understanding by ranking T (influences growth of heterotrophs and autotrophs), f_{SS} (represents soluble biodegradable COD) and f_{XI} (inversely related to biodegradable particulate COD) as influential parameters. For EQI and aeration energy demand, T and f_{XI} are the most influential parameters as the growth of heterotrophs and autotrophs depends on temperature. Higher growth rate and availability of biodegradable particulate COD (which correlates with lower fractions of inerts, f_{XI}) translates to a higher demand for dissolved oxygen (i.e., higher aeration energy demand). The EQI estimation depends on the effluent COD, BOD₅, TSS, TKN and nitrate concentration (see the supplementary material for detailed equations) and the temperature influences the effluent concentration by altering the growth rate. The methane production in the AD depends on the COD load, which comes from the sludge produced in the AS and PC underflow. From activated sludge process understanding it is known that the availability of biodegradable COD influences the sludge production in the AS among other factors that include the decay rate and the SRT. The decay rate itself is also a function of temperature. Likewise, the influent COD fractions also directly influence the composition of the underflow solids from the PC, which are sent to the AD where sludge is

digested anaerobically to produce methane gas. Therefore, the sensitivity analysis identifies T , f_{XI} and f_{SS} as influential parameters for methane gas production as well. The sludge production from the plant is mainly influenced by the influent f_{XI} (which directly affects the amount of sludge production in the plant). Temperature has a lower impact on the sludge production, mainly through its effect on the decay rate of biomass in the system. These observations are all in agreement with general process engineering understanding of activated sludge systems. The results of scenario 1, where the purpose is to study the importance of influent quality on plant design performance, also imply that uncertainty in the influent wastewater fractionations (i.e. COD) and temperature (i.e. T), which could be affected by climate change, are crucial parameters affecting all the key design performance indicators. Therefore in plant design studies, appropriate engineering measures need to be proposed to ensure robustness of plant performance against future changes in influent quality and temperature (e.g. by increasing design safety margins).

Comparing the sensitivity analysis methods, one observes that the sensitivity indices obtained from the Monte Carlo procedure were in very good agreement with the indices obtained from PCE, GPR, and ANN metamodels, with GPR and ANN models giving slightly better results compared to PCE models. This is mainly because GPR and ANN models have a better cross-validation performance score compared to PCE models (see

Table 2, Table 3, Table 4), especially for models of effluent nitrate, effluent ammonium, and the effluent quality index. However, the ANN

model performs even better than the GPR model for the effluent quality index. For methane and sludge production, all the models give very similar results, as they all have cross-validation scores close to 1.

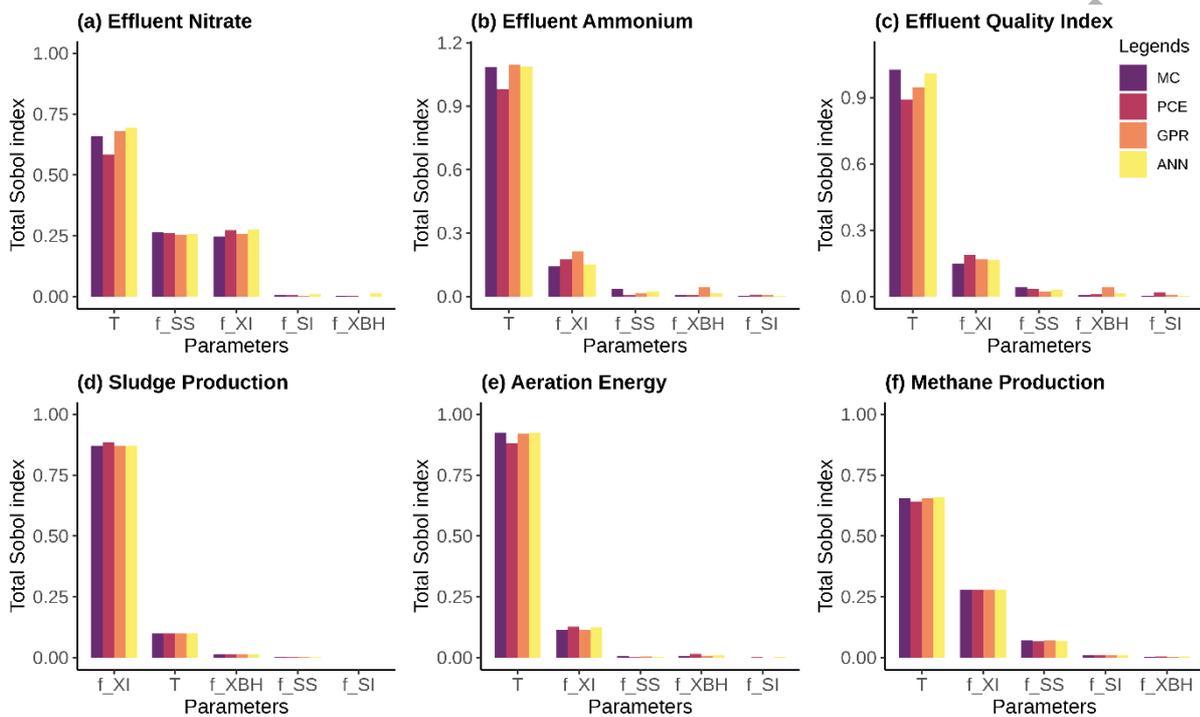


Figure 6: Comparison of Sobol indices obtained from the original plant model (BSM2) using the MC procedure and metamodels (PCE, GPR, and ANN) for the scenario 1. The most influential influent fractionation parameters are shown for the key plant performance indicators: (a) Effluent nitrate (b) Effluent ammonium (c) Effluent quality index (d) Sludge production (e) Aeration Energy (f) Methane production.

3.3.2. Sensitivity analysis of stoichiometric and biokinetic parameters - Scenario 2

The objective of the scenario 2 was to assess the importance of process model parameters on the plant design performance metrics. In

BSM2, ASM1 is used to describe microbial conversion dynamics in activated sludge tanks, hence (stoichiometric and kinetic) parameters of the ASM1 model are analyzed in this scenario as input sources of uncertainty. The sensitivity analysis results, as shown in Figure 7, reveals that for effluent nitrate, the influential parameters (in descending order) are i_{XB} , K_{OH} , η_{yh} , Y_H , and η_g . When the nitrogen fraction in biomass (i.e. i_{XB}) increases, the ammonium concentration inside the reactor increases as this nitrogen fraction is hydrolyzed after biomass decay. Therefore, the effluent nitrate concentration increases due to increased availability of ammonium for nitrification/oxidation. The effluent nitrate concentration can also depend on the anoxic growth rate of heterotrophs (K_{OH} , η_g), which convert nitrate to N_2 . As expected from process engineering understanding, the heterotrophic yield, a stoichiometric parameter, is one of the most influential parameters for this output. It is noteworthy to mark that the EQI is also influenced by the same set of parameters (i_{XB} , K_{OH} , η_{yh} , Y_H , and η_g), because the effluent nitrate concentration has a considerable weight in the EQI calculation (see supplementary material for EQI equation). Likewise, for the effluent ammonium concentration, the sensitivity analysis reveals that the nitrification parameters (K_{NH} , K_{OA} , b_A , u_A) are significant. Moreover, other parameters, namely X_{ITSS} (i.e., TSS to COD ratio), are found to be important. The latter parameter directly affects the sludge production, and, through SRT control, this parameter affects the sludge

wastage rate which determines the amount of nitrifying biomass in the system. It indicates that the sludge production in the AS system increases (recycling more active biomass from SC to AS) with an increase in X_{ITSS} value, and the SRT decreases (as constant TSS is maintained by the controller) by manipulating the flowrate Q_w (waste sludge goes to the AD via the thickener) and reduces the nitrification rate. This understanding agrees well with engineering knowledge and findings reported by Sin et al. (2009). The aeration energy consumption is influenced by the parameters such as i_{XB} , f_p , Y_H , K_{OH} and X_{ITSS} which can be related to the biomass availability in the AS tank. This can be because of a higher active biomass recycle to the AS tank or a higher yield or a lower inert fraction in the biomass or a higher nitrogen fraction in the biomass leading to more ammonium substrate resulting in more AOB growth.

As for the performance comparison of metamodelling, there is no clear difference between the models except that PCE models give slightly diverging results in comparison to the others, which can be attributed to their slightly lower cross-validation scores. Moreover, GSA results from metamodelling also show a good agreement with those obtained from the MCS approach in terms of correctly ranking the most important parameters.

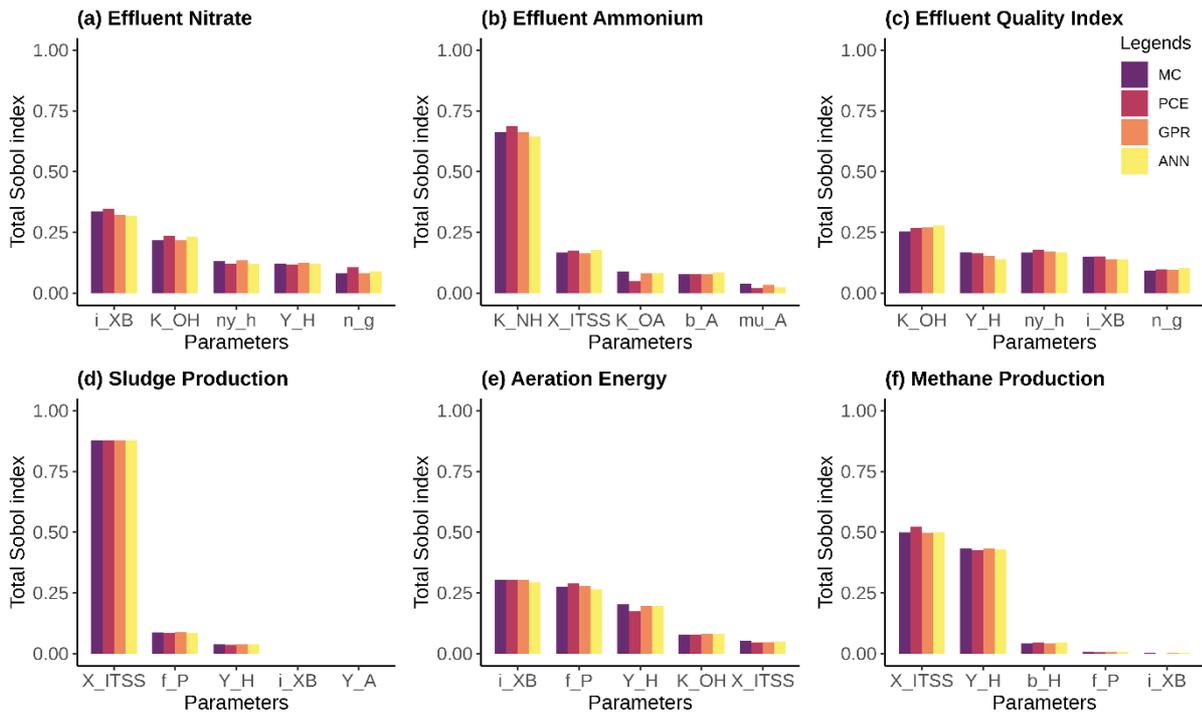


Figure 7: Comparison of Sobol indices obtained from the original plant model (BSM2) using the MC procedure and metamodels (PCE, GPR, and ANN) for the scenario 2. The most influential stoichiometric and kinetic parameters are shown for the key plant performance indicators: (a) Effluent nitrate (b) Effluent ammonium (c) Effluent quality index (d) Sludge production (e) Aeration energy (f) Methane production.

3.3.3. Sensitivity analysis of hydraulics and design related parameters – scenario 3

Regarding the influence of hydraulic and design parameters on the effluent nitrate concentration, which was the scope of scenario 3, Figure 8 shows that slow settling in the secondary clarifier (r_P), the anoxic tank size ($VOL1$, $VOL2$) and the internal recycle flowrate (Q_{intr}) are important parameters. The settling parameter (r_P) will affect the amount of biomass lost in the effluent, which in turn will impact the active biomass concentration in the activated sludge tanks. Likewise, $VOL1$ and $VOL2$

determine the HRT of the anoxic tank and the internal recycle helps to transport nitrate from the aerobic tank to the anoxic tank (Baeza et al., 2004). Altering these parameters can lead to inefficient denitrification that will influence the effluent nitrate concentration. The effluent ammonia is influenced by aerobic tank size (i.e., $VOL3$, $VOL4$, and $VOL5$) as nitrification largely depends on the aerated fraction of the HRT in the activated sludge systems (when the SRT is controlled/fixed). The longer aerated HRT will allow more nitrification to take place in the system. On the contrary, the reduction of HRT (as aerobic tank size reduces) leads to washout of biomass and further accumulation of ammonium (Li et al., 2013) inside the aerobic tank. However, the EQI solely depends on settling parameters (r_p) as poor settling in the SC can drastically increase total suspended solids and nitrate concentrations of the effluent. This also means poor settling leads to higher discharge of biomass in the effluent rather than ending up in the anaerobic digester and later in the sludge. The aeration energy is mainly influenced by the aeration tank size and again by the settling parameter. A similar reasoning can be used here, i.e., poor settling in the SC leads to a lower recycle of active biomass to the AS tank which can lead to a lower aeration energy demand. Therefore, based on scenario 3 results, it is evident that hydraulics (i.e. internal recirculation flow rates in aeration tanks) and effective volumes of reactors (which may change due to construction or mixing issues for

aeration as well as settling tanks) have the potential to affect the design performance metrics. Therefore good engineering practices and design standards (such as Metcalf & Eddy Inc. et al., 2014) should be followed in plant design and commissioning.

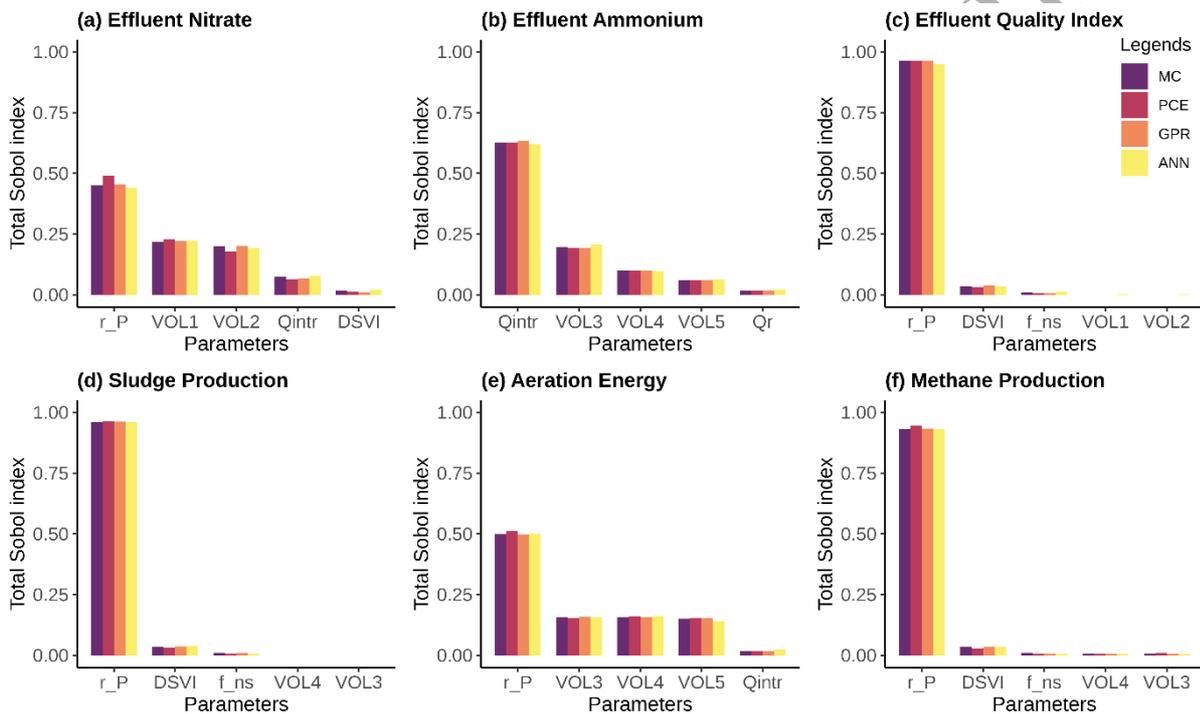


Figure 8: Comparison of Sobol indices obtained from the original plant model (BSM2) using the MC procedure and metamodels (PCE, GPR, and ANN) for the scenario 3. The most influential hydraulics and design parameters are shown for the key plant performance indicators: (a) Effluent nitrate (b) Effluent ammonium (c) Effluent quality index (d) Sludge production (e) Aeration energy (f) Methane production.

3.3.4. Sensitivity analysis of uncertain influent fractionations, kinetic and biokinetic parameters, and hydraulics and design related parameters – Scenario 4

The objective of the scenario 4 is twofold: to analyze the comparative effects of different sets of parameters, and secondly, to test the robustness of the meta-model based GSA techniques in high dimensional settings. For that reason, all the previously studied scenario parameters are gathered to obtain a larger parameter set. Based on the parameter sensitivity rankings obtained in this scenario, the following conclusions are drawn. As was also found in scenario 1, all the plant KPIs show a high sensitivity to the variation in the incoming influent temperature (T), except total sludge production, which is more dictated by the particulate organic matter fraction (f_{XI}) of the influent, as expected. For effluent ammonium, effluent quality index, and aeration energy demand, the total impact of influent temperature is close to one whereas, for methane production, the influent fractions of f_{XI} and f_{SS} were also found as significant parameters. The most important parameter results of this scenario differ from those of scenario 1 only with the inclusion of some scenario 2 parameters (such as i_{XB} , f_P , X_{ITSS}) among the top listed parameters. Moreover, this comparison of different sets of uncertain parameters on the plant KPIs implies that influent uncertainty is the most critical source of uncertainty in WTP design, as it causes the largest deviations from the means of the performance metrics (see Table S5). On the other hand, hydraulics and design related parameters of scenario 3 play a negligible role when

compared with the other sets. Regarding the parameter ranking results obtained from the metamodels, a good agreement with the costly MCS approach is still preserved (see Figure 9), even when the dimension of the parameter space was relatively high ($d=37$). PCE models give slightly less accurate results compared to ANN and GPR models (for example effluent nitrate in Figure 9), though they all correctly rank the most sensitive parameters. This further strengthens the reliability of the use of the metamodeling-based approach for global sensitivity analysis of complex plant-wide models.

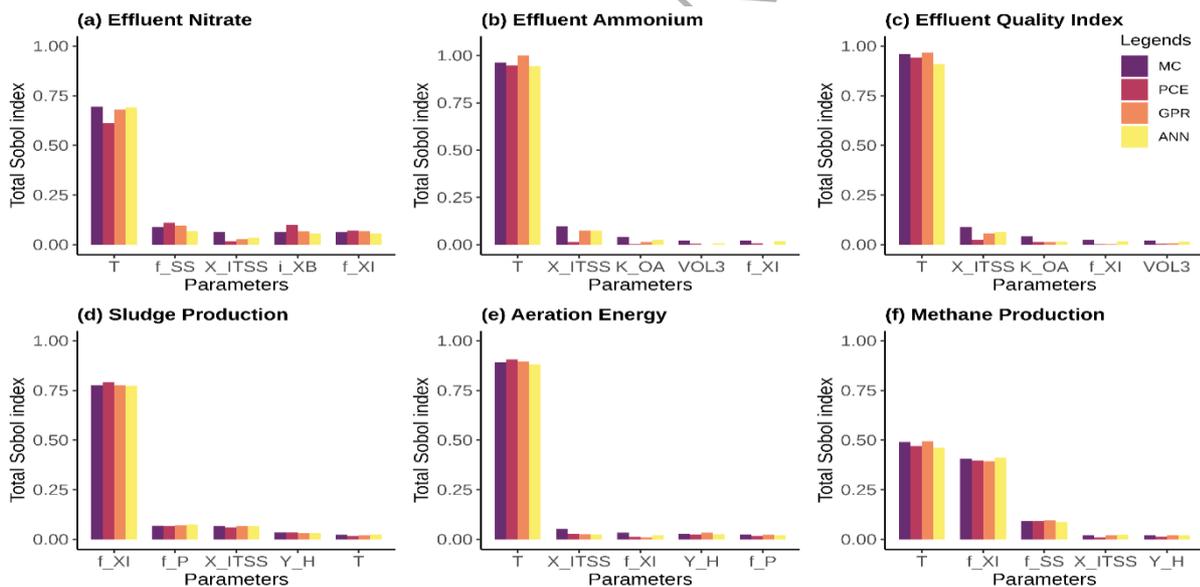


Figure 9: Comparison of Sobol indices obtained from the original plant model (BSM2) using the MC procedure and metamodels (PCE, GPR, and ANN) for the scenario 4. The most influential parameters considered in all scenarios are shown for the key plant performance indicators: (a) Effluent nitrate (b) Effluent ammonium (c) Effluent quality index (d) Sludge production (e) Aeration energy (f) Methane production.

3.4. Comparison of computational cost of different GSA approaches

In terms of their computational costs, the approaches discussed above differ considerably. As the actual elapsed clock time for performing the required simulations will vary depending on the simulation and the solver settings, the type of processing units, the computing architecture (clusters, clouds, etc.), and the computing strategy (level of parallelization), a comparison is given in terms of the number of costly model evaluations needed, which is listed in Table 5 for each approach. To obtain Sobol indices from the Monte Carlo procedure, sampling matrices of size 2000 were used, which resulted in 78000 plant model simulations for the scenario 4 in which 37 parameters were sampled in total. For investigating all the scenarios, a total of 164000 plant simulations were performed in parallel using a high performance cluster connected to 100 processing units. The large size of the matrices is necessary to ensure accurate convergence of each sensitivity index of all the investigated parameters. Expectedly, the metamodeling based approaches used a significantly lower number of original model simulations, which is the size of the experimental design they were built from. The incremental algorithm of the framework used as few as 100 simulations to build GPR and ANN models for most of the outputs of scenarios 1, 2, and 3; however, a larger dataset is needed to attain the target cross-validation scores in scenario 4. A total of 600 original model simulations were used to produce GPR based indices,

plus the time spent on performing the Monte Carlo simulations with metamodels (around 10 minutes in the cluster), which is negligible considering a single plant simulation time of 6 minutes (t_{BSM2}). On the other hand, PCE models needed a total of 750 simulations without the need for Monte Carlo simulations as explained above. The standardized regression coefficients are generated using the Monte Carlo simulation results (using the BSM2), for which a sample size of 1000 was used for each scenario, resulting in 4000 simulations in total. Unlike the Sobol method, the MC procedure for SRC does not depend on the number of parameters sampled. In general, using metamodels for conducting the scenario analysis with BSM2 results in an order of magnitude computational gain compared to SRC, and a two orders of magnitude gain compared to the Sobol method.

Table 5: Comparison of computational costs of different approaches for global sensitivity analysis

Approach	# of plant-wide simulations used				Total computational cost
	Scenario 1	Scenario 2	Scenario 3	Scenario 4	

	$(d = 7)$	$(d = 20)$	$(d = 10)$	$(d = 37)$	
SRC with MCS	1000	1000	1000	1000	$4000 \times t_{BSM2}$
Sobol indices with MCS using BSM2	18000	44000	24000	78000	$164000 \times t_{BSM2}$
Sobol indices with MCS using GPR	150	100	100	250	$600 \times t_{BSM2}$
Sobol indices with MCS using ANN	150	100	100	450	$800 \times t_{BSM2}$
Sobol indices with PCE	250	150	100	250	$750 \times t_{BSM2}$

4. Conclusions

In this work, a new methodology for performing efficient global sensitivity analysis using advanced surrogate models is presented and demonstrated with four different scenarios covering various kinds of epistemic system uncertainties in wastewater treatment plants. The key findings of the presented work can be summarized as follows:

- The proposed framework facilitates gaining valuable insights into performance of complex process systems like WWTPs by combining advanced GSA methods (such as variance decomposition-based Sobol sensitivity analysis) with powerful machine learning algorithms.
- The Sobol sensitivity method provides more reliable sensitivity measures compared to the widely-used SRC method, especially when highly nonlinear relations exist between the model outputs of interest and the input parameters. When the degree of linearization

is low, the SRC method fails to provide correct parameter rankings, and therefore should be avoided in plant-wide model applications that feature strong nonlinearities between inputs and outputs.

- In general, compared to MC-based techniques, using surrogate models to carry out system-wide GSA of WWTP models yields computational gains as large as two orders of magnitude.
- The easyGSA toolbox provided with the framework expedites the process of constructing highly accurate surrogate models, and enables performing computationally heavy model-based analysis such as global sensitivity analysis and beyond.
- Polynomial chaos expansion is a particularly efficient technique that can be used to calculate Sobol sensitivity indices of complex biological systems at a very low computational cost. To further increase the accuracy of the PCE-based indices and to cross-validate them, GPR and ANN type metamodels can also be constructed to replace the computationally demanding models in Monte Carlo simulations.
- Influent fractionations, stoichiometric and biokinetic model parameters as well as hydraulics and design related parameters are all found to be influential in estimating performance metrics of WWTP systems when analyzed separately. However, the comparative analysis of all

the parameters confirms that the influent parameters are more important for most of the plant KPIs.

Finally, although it has not seen wide use in plant-wide modeling of WWTPs, Sobol-based global sensitivity analysis is a powerful tool for unveiling the underlying complex parameter dependence of system behaviors. The computational cost can be further brought down by exploring advanced surrogate modeling techniques, for which the sequential framework provides a useful workflow. As the framework is flexible, cross-validated surrogate models can also be generated for other application purposes than global sensitivity analysis, such as process synthesis and model-based design optimization, which is an ongoing work.

Acknowledgments

The authors acknowledge funding from the EU Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant

agreement no.675251 as well as the funding from the Water Joint Programming Initiative under the PIONEER_STP project.

ACCEPTED MANUSCRIPT

References

- Al, R., Behera, C.R., Zubov, A., Sin, G., 2018. Systematic framework development for the construction of surrogate models for wastewater treatment plants, in: *Computer Aided Chemical Engineering*. pp. 1909-1914. <https://doi.org/10.1016/B978-0-444-64241-7.50313-X>
- Baeza, J., Gabriel, D., Lafuente, J., 2004. Effect of internal recycle on the nitrogen removal efficiency of an anaerobic/anoxic/oxic (A2/O) wastewater treatment plant (WWTP). *Process Biochem.* 39, 1615-1624. [https://doi.org/10.1016/S0032-9592\(03\)00300-5](https://doi.org/10.1016/S0032-9592(03)00300-5)
- Batstone, D.J., Keller, J., Angelidaki, I., Kalyuzhnyi, S.V., Pavlostathis, S.G., Rozzi, A., Sanders, W.T., Siegrist, H., Vavilin, V.A., 2002. The IWA Anaerobic Digestion Model No 1 (ADM1). *Water Sci. Technol.* 45, 65-73.
- Behera, C.R., Santoro, D., Gernaey, K.V., Sin, G., 2018. Organic carbon recovery modeling for a rotating belt filter and its impact assessment on a plant-wide scale. *Chem. Eng. J.* 334. <https://doi.org/10.1016/j.cej.2017.11.091>
- Blatman, G., Sudret, B., 2011. Adaptive sparse polynomial chaos expansion based on least angle regression. *J. Comput. Phys.* 230, 2345-2367. <https://doi.org/10.1016/j.jcp.2010.12.021>
- Boiocchi, R., Gernaey, K. V., Sin, G., 2017. Understanding N₂O formation mechanisms through sensitivity analyses using a plant-wide benchmark simulation model. *Chem. Eng. J.* 317, 935-951. <https://doi.org/10.1016/j.cej.2017.02.091>
- Burnaev, E., Panin, I., Sudret, B., 2017. Efficient design of experiments for sensitivity analysis based on polynomial chaos expansions.

- Cheng, K., Lu, Z., Zhang, K., 2019. Multivariate output global sensitivity analysis using multi-output support vector regression. *Struct. Multidiscip. Optim.* <https://doi.org/10.1007/s00158-018-2184-z>
- Cosenza, A., Mannina, G., Vanrolleghem, P.A., Neumann, M.B., 2014. Variance-based sensitivity analysis for wastewater treatment plant modelling. *Sci. Total Environ.* 470-471, 1068-1077. <https://doi.org/10.1016/j.scitotenv.2013.10.069>
- Cosenza, A., Mannina, G., Vanrolleghem, P.A., Neumann, M.B., 2013. Global sensitivity analysis in wastewater applications: A comprehensive comparison of different methods. *Environ. Model. Softw.* 49, 40-52. <https://doi.org/10.1016/j.envsoft.2013.07.009>
- Davis, S.E., Cremaschi, S., Eden, M.R., 2018. Efficient Surrogate Model Development: Impact of Sample Size and Underlying Model Dimensions, in: *Computer Aided Chemical Engineering*. Elsevier Masson SAS, pp. 979-984. <https://doi.org/10.1016/B978-0-444-64241-7.50158-0>
- Dragan, J.M., Zubov, A., Sin, G., 2017. Methodology for Plantwide Design and Optimization of Wastewater Treatment Plants, 27th European Symposium on Computer Aided Process Engineering. Elsevier Masson SAS. <https://doi.org/10.1016/B978-0-444-63965-3.50145-8>
- Fajraoui, N., Ramasomanana, F., Younes, A., Mara, T.A., Ackerer, P., Guadagnini, A., 2011. Use of global sensitivity analysis and polynomial chaos expansion for interpretation of nonreactive transport experiments in laboratory-scale porous media. *Water Resour. Res.* 47, 1-14. <https://doi.org/10.1029/2010WR009639>

- Ferretti, F., Saltelli, A., Tarantola, S., 2016. Trends in sensitivity analysis practice in the last decade. *Sci. Total Environ.* 568, 666-670.
<https://doi.org/10.1016/j.scitotenv.2016.02.133>
- Flores-Alsina, X., Corominas, L., Neumann, M.B., Vanrolleghem, P.A., 2012. Assessing the use of activated sludge process design guidelines in wastewater treatment plant projects: A methodology based on global sensitivity analysis. *Environ. Model. Softw.* 38, 50-58.
<https://doi.org/10.1016/j.envsoft.2012.04.005>
- Flores-Alsina, X., Rodríguez-Roda, I., Sin, G., Gernaey, K. V., 2008. Multi-criteria evaluation of wastewater treatment plant control strategies under uncertainty. *Water Res.* 42, 4485-4497.
<https://doi.org/10.1016/j.watres.2008.05.029>
- Forrester, A.I.J., Sbester, A., Keane, A.J., 2008. *Engineering Design via Surrogate Modelling*. John Wiley & Sons, Ltd, Chichester, UK.
<https://doi.org/10.1002/9780470770801>
- Fortela, D.L.B., Sharp, W.W., Revellame, E.D., Hernandez, R., Gang, D., Zappi, M.E., 2019. Computational evaluation for effects of feedstock variations on the sensitivities of biochemical mechanism parameters in anaerobic digestion kinetic models. *Biochem. Eng. J.* 143, 212-223.
<https://doi.org/10.1016/j.bej.2019.01.001>
- Géron, A., 2017. *Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems*. "O'Reilly Media, Inc."
- Henze, M., Grady, C.P.L., Gujer, W., Maris, G.V.R., Matsuo, T., 1987. *Activated sludge process model No. 1*.

- Iooss, B., Lemaître, P., 2015. A Review on Global Sensitivity Analysis Methods BT - Uncertainty Management in Simulation-Optimization of Complex Systems: Algorithms and Applications, in: Dellino, G., Meloni, C. (Eds.), . Springer US, Boston, MA, pp. 101-122.
https://doi.org/10.1007/978-1-4899-7547-8_5
- Jeppsson, U., Rosen, C., Alex, J., Copp, J., Gernaey, K. V., Pons, M.N., Vanrolleghem, P.A., 2006. Towards a benchmark simulation model for plant-wide control strategy performance evaluation of WWTPs. *Water Sci. Technol.* 53, 287-295. <https://doi.org/10.2166/wst.2006.031>
- Li, H., Zhang, Y., Yang, M., Kamagata, Y., 2013. Effects of hydraulic retention time on nitrification activities and population dynamics of a conventional activated sludge system. *Front. Environ. Sci. Eng. China* 7, 43-48. <https://doi.org/10.1007/s11783-012-0397-8>
- Li, S., Yang, B., Qi, F., 2016. Accelerate global sensitivity analysis using artificial neural network algorithm: Case studies for combustion kinetic model. *Combust. Flame* 168, 53-64.
<https://doi.org/10.1016/j.combustflame.2016.03.028>
- Mannina, G., Cosenza, A., Viviani, G., Ekama, G.A., 2018. Sensitivity and uncertainty analysis of an integrated ASM2d MBR model for wastewater treatment. *Chem. Eng. J.* 351, 579-588.
<https://doi.org/10.1016/j.cej.2018.06.126>
- Mara, T.A., Tarantola, S., Annoni, P., 2015. Non-parametric methods for global sensitivity analysis of model output with dependent inputs. *Environ. Model. Softw.* 72, 173-183.
<https://doi.org/10.1016/j.envsoft.2015.07.010>

- Marelli, S., Sudret, B., 2016. UQLab The Framework for Uncertainty Quantification 2016.
- Marrel, A., Iooss, B., Laurent, B., Roustant, O., 2009. Calculations of Sobol indices for the Gaussian process metamodel. *Reliab. Eng. Syst. Saf.* 94, 742-751. <https://doi.org/10.1016/j.ress.2008.07.008>
- McKay, M.D., Beckman, R.J., Conover, W.J., 1979. Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics* 21, 239-245. <https://doi.org/10.1080/00401706.1979.10489755>
- Metcalf & Eddy Inc., Tchobanoglous, G., Burton, F.L., Tsuchihashi, R., Stensel, H.D., 2014. *Wastewater Engineering: Treatment and Resource Recovery*, 5th ed. McGraw-Hill Education.
- Molinaro, A.M., Simon, R., Pfeiffer, R.M., 2005. Prediction error estimation: A comparison of resampling methods. *Bioinformatics* 21, 3301-3307. <https://doi.org/10.1093/bioinformatics/bti499>
- Morris, M.D., 1991. Factorial Sampling Plans for Preliminary Computational Experiments. *Technometrics* 33, 161-174. <https://doi.org/10.2307/1269043>
- Ochoa, M.P., Estrada, V., Hoch, P.M., 2016. Wastewater stabilisation ponds system: Global sensitivity analysis on network design. *Chem. Eng. Trans.* 50, 187-192. <https://doi.org/10.3303/CET1650032>
- Pianosi, F., Beven, K., Freer, J., Hall, J.W., Rougier, J., Stephenson, D.B., Wagener, T., 2016. Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environ. Model. Softw.* 79, 214-232. <https://doi.org/10.1016/j.envsoft.2016.02.008>
- Ramin, E., Flores-Alsina, X., Sin, G., Gernaey, K. V., Jeppsson, U., Mikkelsen, P.S., Plósz, B.G., 2014. Influence of selecting secondary

settling tank sub-models on the calibration of WWTP models - A global sensitivity analysis using BSM2. *Chem. Eng. J.* 241, 28-34.

<https://doi.org/10.1016/j.cej.2013.12.015>

Rasmussen, C.E., Williams, C.K.I., 2006. *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, Massachusetts.

Rojas, J., Zhelev, T., 2012. Energy efficiency optimisation of wastewater treatment: Study of ATAD. *Comput. Chem. Eng.* 38, 52-63.

<https://doi.org/10.1016/j.compchemeng.2011.11.016>

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. *Nature* 323, 533-536.

<https://doi.org/10.1038/323533a0>

Saltelli, A., Aleksankina, K., Becker, W., Fennell, P., Ferretti, F., Holst, N., Li, S., Wu, Q., 2017. Why So Many Published Sensitivity Analyses Are False. *A Systematic Review of Sensitivity Analysis Practices* 1-36.

Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., Tarantola, S., 2010. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Comput. Phys. Commun.* 181, 259-270. <https://doi.org/10.1016/j.cpc.2009.09.018>

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S., 2008. *Global Sensitivity Analysis. The Primer*, in: *Global Sensitivity Analysis. The Primer*. John Wiley & Sons, Ltd, Chichester, UK, pp. 237-275.

Sin, G., Gernaey, K. V., Neumann, M.B., van Loosdrecht, M.C.M., Gujer, W., 2009. Uncertainty analysis in WWTP model applications: A critical discussion using an example from design. *Water Res.* 43, 2894-2906.

<https://doi.org/10.1016/j.watres.2009.03.048>

- Sin, G., Gernaey, K. V, Neumann, M.B., van Loosdrecht, M.C.M., Gujer, W., 2011. Global sensitivity analysis in wastewater treatment plant model applications: Prioritizing sources of uncertainty. *Water Res.* 45, 639-651. <https://doi.org/10.1016/j.watres.2010.08.025>
- Sobol, I.M., Kucherenko, S.S., 2005. Global sensitivity indices for nonlinear mathematical models. *Review. Wilmott* 2005, 56-61. <https://doi.org/10.1002/wilm.42820050114>
- Steiner, M., Bourinet, J.M., Lahmer, T., 2019. An adaptive sampling method for global sensitivity analysis based on least-squares support vector regression. *Reliab. Eng. Syst. Saf.* 183, 323-340. <https://doi.org/10.1016/j.ress.2018.11.015>
- Sudret, B., 2008. Global sensitivity analysis using polynomial chaos expansions. *Reliab. Eng. Syst. Saf.* 93, 964-979. <https://doi.org/10.1016/j.ress.2007.04.002>
- Takács, I., Patry, G.G., Nolasco, D., 1991. A dynamic model of the clarification-thickening process. *Water Res.* 25, 1263-1271. [https://doi.org/10.1016/0043-1354\(91\)90066-Y](https://doi.org/10.1016/0043-1354(91)90066-Y)
- Wu, Z., Wang, W., Wang, D., Zhao, K., Zhang, W., 2019. Global sensitivity analysis using orthogonal augmented radial basis function. *Reliab. Eng. Syst. Saf.* 185, 291-302. <https://doi.org/10.1016/j.ress.2018.12.028>