



Global Sensitivity Analysis of Metabolic Models for Phosphorus Accumulating Organisms in Enhanced Biological Phosphorus Removal

Minh Nguyen Quang¹, Tim Rogers², Jan Hofman¹ and Ana B. Lanham^{1*}

¹ Department of Chemical Engineering, Water Innovation and Research Centre, University of Bath, Bath, United Kingdom,

² Department of Mathematical Sciences, Centre for Networks and Collective Behaviour, University of Bath, Bath, United Kingdom

OPEN ACCESS

Edited by:

Manfred Zinn,
HES-SO Valais-Wallis, Switzerland

Reviewed by:

Zhen Chen,
Tsinghua University, China
Antonio Espuña,
Universitat Politècnica de Catalunya,
Spain

*Correspondence:

Ana B. Lanham
a.lanham@bath.ac.uk

Specialty section:

This article was submitted to
Bioprocess Engineering,
a section of the journal
Frontiers in Bioengineering and
Biotechnology

Received: 20 December 2018

Accepted: 09 September 2019

Published: 04 October 2019

Citation:

Nguyen Quang M, Rogers T,
Hofman J and Lanham AB (2019)
Global Sensitivity Analysis of
Metabolic Models for Phosphorus
Accumulating Organisms in Enhanced
Biological Phosphorus Removal.
Front. Bioeng. Biotechnol. 7:234.
doi: 10.3389/fbioe.2019.00234

The aim of this study was to identify, quantify and prioritize for the first time the sources of uncertainty in a mechanistic model describing the anaerobic-aerobic metabolism of phosphorus accumulating organisms (PAO) in enhanced biological phosphorus removal (EBPR) systems. These wastewater treatment systems play an important role in preventing eutrophication and metabolic models provide an advanced tool for improving their stability via system design, monitoring and prediction. To this end, a global sensitivity analysis was conducted using standard regression coefficients and Sobol sensitivity indices, taking into account the effect of 39 input parameters on 10 output variables. Input uncertainty was characterized with data in the literature and propagated to the output using the Monte Carlo method. The low degree of linearity between input parameters and model outputs showed that model simplification by linearization can be pursued only in very well defined circumstances. Differences between first and total-order sensitivity indices showed that variance in model predictions was due to interactions between combinations of inputs, as opposed to the direct effect of individual inputs. The major sources of uncertainty affecting the prediction of liquid phase concentrations, as well as intra-cellular glycogen and poly-phosphate was due to 64% of the input parameters. In contrast, the contribution to variance in intra-cellular PHA constituents was uniformly distributed among all inputs. In addition to the intra-cellular biomass constituents, notably PHB, PH₂MV and glycogen, uncertainty with respect to input parameters directly related to anaerobic propionate uptake, aerobic poly-phosphate formation, glycogen formation and temperature contributed most to the variance of all model outputs. Based on the distribution of total-order sensitivities, characterization of the influent stream and intra-cellular fractions of PHA can be expected to significantly improve model reliability. The variance of EBPR metabolic model predictions was quantified. The means to account for this variance, with respect to each quantity of interest, given knowledge of the

corresponding input uncertainties, was prescribed. On this basis, possible avenues and pre-requisite requirements to simplify EBPR metabolic models for PAO, both structurally via linearization, as well as by reduction of the number of non-influential variables were outlined.

Keywords: global sensitivity analysis, Monte Carlo, enhanced biological phosphorus removal, phosphorus accumulating organism, metabolic model, EBPR, standard regression coefficients, sobol sensitivity analysis

1. INTRODUCTION

Enhanced biological phosphorus removal (EBPR), a variation of conventional Activated Sludge (AS), is a widely employed technology to remove phosphorus from wastewater. This prevents eutrophication in waterways due to the excess of nutrients (Metcalf and Eddy, 2003; Oehmen et al., 2007). By engineering alternating anaerobic and aerobic conditions, the resultant community of phosphorus accumulating organisms (PAO) assimilates phosphorus from the bulk liquid phase by intra-cellular accumulation as poly-phosphate (poly-P). Under stable operation, EBPR installations have been demonstrated to achieve very high phosphorus removal efficiencies, with effluent phosphate (PO_4) concentrations as low as 0.5 mg/l (Lopez-Vazquez et al., 2008a).

However, EBPR is known to suffer from operational instabilities from causes which are not completely understood. The deterioration of phosphorus removal performance is often attributed to the proliferation of glycogen accumulating organisms (GAO), which do not directly contribute to phosphorus removal yet compete with PAO for carbon sources, e.g., volatile fatty acids (VFA). In an effort to suppress their growth, different studies have investigated factors that influence the competition between these organisms, e.g., the carbon source (Pijuan et al., 2004; Oehmen et al., 2005b), pH (Filipe et al., 2001a,b,c), temperature (Whang and Park, 2006; Lopez-Vazquez et al., 2008b), the P/C ratio in the influent (Liu et al., 1997) and the aeration levels (Carvalho et al., 2014a). Metabolic models, generally a more detailed approach than conventional activated sludge models (ASM), can be used to effectively describe their metabolisms and thereby consider the effect of these factors on the competition between PAO and GAO.

Metabolic models offer a way to integrate information from various sources within a common mathematical framework. They describe and predict the metabolism of key organisms relevant to EBPR on a mechanistic basis. The EBPR metabolic model stemmed mainly from the work of Smolders et al. (1994), describing the anaerobic metabolic pathways of PAO. It has since grown to include the aerobic (Smolders et al., 1995) and anoxic (Kuba et al., 1996) metabolisms of *Accumulibacter* PAO, as well as that of *Competibacter* and *Defluviicoccus*-related GAO (Filipe et al., 2001b; Zeng et al., 2003b; Oehmen et al., 2005b). These models have also expanded to include the effects of carbon source, temperature and pH (Oehmen et al., 2005a, 2006; Lopez-Vazquez et al., 2009). Interfaced with ASM, they have been successfully used to describe the performance of full-scale EBPR installations (Van Veldhuizen et al., 1999; Meijer et al., 2001). Furthermore, they offer a means with which to

test hypotheses concerning microbial ecology and explore their population dynamics under multi-parametric settings, e.g., as conducted by Lopez-Vazquez et al. (2009) between PAO and GAO, as well as within their respective sub-groups, e.g., in Oehmen et al. (2010b) between *Accumulibacter* PAO clades I and II.

In recent years, the application of advanced molecular techniques has led to a more comprehensive understanding of both the metabolic activity and phylogenetic diversity (Oehmen et al., 2007), e.g., differing capacities of PAO clades for denitrification (Zeng et al., 2003a; Flowers et al., 2009; Oehmen et al., 2010b), utilization of the TCA cycle to supply reducing power for PHA formation (Lanham et al., 2014) or shift to glycolysis-driven VFA uptake (Acevedo et al., 2014). Consequently, the complexity of metabolic models, in terms of stoichiometric and kinetic descriptions of the underlying processes, as well as the number of different organisms to account for, has been on the rise. This trend is set to continue with the incorporation of newly-recognized putative PAO and GAO, e.g., *Tetrasphaera* (Maszenan et al., 2000), more detailed characterization of GAO metabolisms, namely glycolysis pathways and VFA-uptake mechanisms, as well as correct differentiation of alternative metabolic pathways from the presence of different bacterial strains (Oehmen et al., 2010a).

This ever-increasing complexity and sprawling-nature of metabolic models must however be justified, given the level of detail and number of organisms already described. Variations in model predictions could be exacerbated by uncertainty associated with the input parameters (Sin et al., 2005). Therefore, this work seeks to address for the first time, in a comprehensive way, the subjective uncertainty in metabolic models for EBPR, i.e., that arising from incomplete knowledge about the true value of the model's input parameters, by performing a global sensitivity analysis (GSA) to determine the key parameters that influence their predictions.

Although previous studies reported local sensitivity analyses, such procedures consider the variation of different input parameters only one at a time and typically at the same initial conditions, e.g., in Lopez-Vazquez et al. (2009), Oehmen et al. (2010b), and Lanham et al. (2014). As such, they do not account for interactions between different sources of uncertainty, apart from the case of simple linear models (Saltelli et al., 2006). In contrast, GSA seeks to ascertain the relative importance of key processes and/or parameters driving the output dynamics of the overall system by perturbing all inputs simultaneously. It has been used effectively in areas of risk assessment (Mokhtari and Frey, 2005), experimental design (Kent et al., 2013) and model development (Sin et al., 2009) applied to biochemical systems.

GSA provides a basis for the justification (or rejection) of previous assumptions regarding the estimation of input parameters during model calibration, thereby generating feedback with which to improve existing model formulations. Most importantly, GSA enables meaningful comparisons of different model formulations by decoupling uncertainty in the required input parameters from the heterogeneity in the calibration procedure. In this study, a mechanistic model was developed to integrate available knowledge concerning the metabolism of *Accumulibacter* PAO in anaerobic-aerobic EBPR systems. Given the model's complexity, underlying assumptions and uncertainty of the true value of its various input parameters, a GSA was conducted by simultaneous perturbation of all input parameters.

First, standard regression coefficients (SRC) (Saltelli et al., 2008) were determined from the ordinary least squares (OLS) method using linear regression on the output of Monte Carlo simulations. Alternative regression methods, e.g., step-wise, ridge or least absolute shrinkage and selection operator (LASSO) were considered, owing to the additional step of selecting predictor variables that aid in the interpretation of the fitted model (Tibshirani, 1996). Nevertheless, regression by OLS was chosen as the most fit-for-purpose, given that (1) sampling of the input parameter space was randomized (low collinearity), (2) the overall number of sample points was large relative to the number of predictor variables, and (3) the consequences of over-fitting were negligible, thereby negating the advantages of more sophisticated methods (Melkumova and Shatskikh, 2017).

SRC results were complemented with sensitivity indices obtained via the Sobol method for variance decomposition to investigate non-linear interactions (Sobol, 2001). Furthermore, Sobol sensitivity indices were also computed to determine the combined influence of multiple-inputs to reveal higher-order interactions (Saltelli et al., 2008). Although alternative variance-based methods are available, e.g., the Fourier amplitude sensitivity test (FAST), which can be computationally more efficient, the Sobol method was selected on account of being more robust, particularly for the determination of total-order sensitivities (Saltelli and Bolado, 1998).

The contribution of each input parameter's uncertainty on the variance of model outputs was then calculated. In this way, input parameters that most significantly affect the model output were identified. Opportunities for factor fixing and model simplification were investigated in view of reducing the complexity of existing models, thereby facilitating parameter estimation their integration within the ASM framework. This would ensure that the limitations and uncertainty intervals of current metabolic models would be better understood. In addition, it can flag key parameters that carry the greatest uncertainty, which would need careful measurement or further experimental investigation. Models would become more applicable to the general case rather than being specific to any particular experimental design, which could significantly improve their homogeneity and applicability. This would ultimately lead to improved efficiency and reliability of EBPR systems, an important pollution control and resource recovery technology.

2. MATERIALS AND METHODS

2.1. Model Construction, Evaluation, and Analysis Environment

Data processing and analysis were done in the Python programming language, ver 3.7. This included: (1) characterization of input uncertainty, (2) sampling from the space of input factors, (3) propagation of input uncertainty to the output and (4) quantifying the contribution of input uncertainty on the variance of model outputs. Where possible, parts of the implementation were adapted from SALib ver 1.1.2 (Herman and Usher, 2017). Python script files can be made available upon request.

2.2. Model Description

The model described the anaerobic-aerobic metabolism of *Accumulibacter* PAO. Stoichiometric and kinetic dependencies on the carbon source as acetate (HAc) and propionate (HPr) were implemented according to Smolders et al. (1995) and Oehmen et al. (2005b), respectively. Modifications to the maximum substrate uptake rates were implemented in accordance with Carvalheira et al. (2014b). The dependence of intra-cellular poly-hydroxyalkanoate (PHA) formation on carbon source followed that of Zeng et al. (2003b). Temperature dependencies were accounted for as described in Brdjanovic et al. (1997). The effect of pH on stoichiometric yields and kinetics was implemented in line with Filipe et al. (2001a) and Filipe et al. (2001c), respectively. Sequential maintenance on PHA, glycogen and poly-P followed Lanham et al.'s (2014) formulation.

Overall, the model tracked one 5 h operational cycle, split between an anaerobic and aerobic phase of equal duration, supplied with a mixture of HAc and HPr. These included 4 dissolved components in the bulk liquid phase: O₂, HAc, HPr and PO₄. It also accounted for the concentration of PAO biomass, alongside its intra-cellular storage compounds: poly-β-hydroxybutyrate (PHB), poly-β-hydroxyvalerate (PHV), poly-β-hydroxy-2-methylvalerate (PH₂MV), glycogen and poly-P. The process of biomass growth was described as the difference between total PHA degradation and that used for the replenishment of poly-P and glycogen reserves (Murnleitner et al., 1997; Lopez-Vazquez et al., 2009). In aggregate, the model consisted of 10 components and 10 kinetic processes (see **Appendices 1, 5** in Supplementary Materials).

2.3. Characterization of Input Uncertainty

A total of 39 input parameters were considered. These were allocated into four categories for analysis: metabolic parameters, kinetic parameters, Arrhenius temperature coefficients and initial conditions. Metabolic parameters included: the ATP requirement for biomass synthesis from Acetyl-CoA (K_1) and Propionyl-CoA (K_2), the yield of ATP per unit of NADH₂ oxidized (δ), the phosphate transport coefficient (ϵ) and the half-saturation constants (K_{HAc} , K_{HPr} , K_{PO_4} , K_{PHA} , K_{Gly} , K_{fPHA} , K_{pp}). Kinetic parameters included the maximum specific rates of anaerobic HAc (q_{HAc}) and HPr (q_{HPr}) uptake, the anaerobic maintenance coefficient ($m_{\text{ATP,an}}$), as well as the rates of aerobic

PHA degradation (q_{PHA}), glycogen production (q_{Gly}), poly-P formation (q_{PP}) and the aerobic maintenance coefficient ($m_{\text{ATP,ox}}$). The maximum intra-cellular fractions of glycogen ($f_{\text{Gly,max}}$) and poly-P ($f_{\text{PP,max}}$) were also regarded as kinetic parameters. Arrhenius temperature coefficients included those corresponding to the effect of temperature on: anaerobic uptake of HAC and HPr ($\theta_{q_{\text{VFA}}}$), anaerobic maintenance ($\theta_{m_{\text{ATP,an}}}$), as well as aerobic PHA degradation ($\theta_{q_{\text{PHA}}}$), glycogen production ($\theta_{q_{\text{Gly}}}$), poly-P formation ($\theta_{q_{\text{PP}}}$) and aerobic maintenance ($\theta_{m_{\text{ATP,ox}}}$). Finally, initial conditions, i.e., the total concentration of carbon sources $S_{\text{VFA},i}$ and the corresponding fraction of HAC ($r_{\text{HAC/HPr},i}$) in the influent, the ratio of PO_4 to carbon ($r_{\text{P/C},i}$), the initial biomass concentration ($X_{\text{PAO},i}$), as well as the intra-cellular fractions of PHA ($X_{\text{PHB},i}$, $X_{\text{PHV},i}$, $X_{\text{PH}_2\text{MV},i}$), glycogen ($X_{\text{Gly},i}$) and poly-P ($X_{\text{PP},i}$) were also taken into consideration. As the distribution from which the input parameters were sampled was not uniform, readers are directed to the **Supplementary Materials** for a detailed account.

Input uncertainty was characterized by systematic review of quantities whose values had been measured or otherwise estimated from experimental data in the literature. Each parameter was considered a random variable that followed a uniform distribution between a minimum and maximum value of a and b , respectively. These ranges were selected based on the degree of variation, var , around the average, x_{mean} . Variability was scaled to reflect the number of data reported in the literature, as 50, 25, or 5% for 1, less or equal to 10, or more measurements, respectively. To ensure that the range extended at least to the limits found in the literature, the minimum value in the collected data was taken as the lower bound a if it was smaller than that determined using var . The upper bound was selected in a similar manner, as defined by Equations (1) and (2).

$$a = \min((1 - var) x_{\text{mean}}, x_{\text{min}}) \quad (1)$$

$$b = \max((1 + var) x_{\text{mean}}, x_{\text{max}}) \quad (2)$$

To conform to bounds within which the model has been experimentally validated, temperature was sampled between the limits of 10 to 30°C, and pH was sampled between 6 and 7.5. To reflect the nature of knowledge completeness regarding the value of parameters where sufficient experimental data could be obtained (i.e., > 50 points), the data was fitted to the Erlang distribution.

Model inputs were generated via Monte Carlo sampling, where each X_i constituted a set of input parameter values with which to evaluate the model:

$$X_i = [x_{1,i}, x_{2,i}, \dots, x_{K,i}] \quad \text{for } i = 1, 2, \dots, N \quad (3)$$

where N is the total number of samples. Initial concentrations in the bulk liquid phase, biomass and intra-cellular components were also included as part of the input uncertainty, given that characterization of the influent to a sufficient level of detail is a known obstacle. In this sense, the initial concentrations of HAC,

HPr and PO_4 were derived as ratios from the total concentration of carbon substrate. Intra-cellular fractions of PHB, PHV and PH_2MV were sampled according to the Dirichlet distribution, in order to ensure that their sum would be equal to the total PHA fraction, which itself followed the uniform distribution. The implementation was based on Saltelli et al.'s (2008, 2010) extension of Sobol sequences to generate quasi-random points.

The $N \times M$ dimensional matrix of inputs, where M is the total number of parameters was propagated via the metabolic model, yielding a three-dimensional matrix of outputs $Y_{T \times K \times N}$, where T is the number of time steps and K is the number of output variables. The mean of the concentration profile over one cycle was calculated, as the methods for quantifying the sensitivity indices require scalar values.

2.4. Sensitivity Analysis Measures

2.4.1. Standard Regression Coefficients

Standard regression coefficients were obtained by performing a series of linear regressions on data from the matrix of inputs and each column of the output matrix:

$$y_{i,k} = b_m + \sum_{m=1}^M b_{m,k} x_{i,m} + \epsilon_{i,k} \quad \text{for } i = 1, 2, \dots, N \text{ and } k = 1, 2, \dots, K \quad (4)$$

where $y_{i,k}$ is the vector of values corresponding to the k th output variable, $b_{m,k}$ is the coefficient of the m th input parameter, $x_{i,m}$ is the value of the m th parameter and $\epsilon_{i,k}$ is the residual error. As shown in Equation (5), the standard regression coefficient $\beta_{m,k}$, i.e., the effect of parameter m on output k , relative to all other parameters, was determined by scaling the input and output values by their respective means and standard deviations:

$$\frac{y_{i,k} - \mu_k}{\sigma_k} = \sum_{m=1}^M \beta_{m,k} \frac{x_{i,m} - \mu_m}{\sigma_m} + \epsilon_{i,k} \quad (5)$$

2.4.2. Sobol Method for Variance Decomposition

The variance observed in the output can be expressed as the sum of variances of individual parameters and their combinations, as shown in Equations 6 to 8 (Saltelli et al., 2008):

$$\text{Var}(Y) = \sum_{i=1}^M V_i + \sum_{i<j} V_{ij} + \dots + V_{1,2,\dots,m} \quad (6)$$

$$V_i = V_{X_i}(E_{X_i}(Y|X_i)) \quad (7)$$

$$V_{ij} = V_{X_{ij}}(E_{X_{\sim ij}}(Y|X_i, X_j)) - V_i - V_j \quad (8)$$

where $\text{Var}(Y)$ is the overall variance of the output, V_i is the variance of the input parameter i and V_{ij} is the partial variance of parameters i and j . X_i denotes the set values taken by parameter i , and $X_{\sim i}$ denotes the set of values taken by all parameters except i .

$$S_i = \frac{V_i}{\text{Var}(Y)} \quad (9)$$

$$S_{Ti} = \frac{E_{X_{\sim i}}(\text{Var}(Y|X_{\sim i}))}{\text{Var}(Y)} = 1 - \frac{\text{Var}_{X_{\sim i}}(E_{X_i}(Y|X_{\sim i}))}{\text{Var}(Y)} \quad (10)$$

From this, the first-order sensitivity index, also referred to as the “effect” of a given parameter i , can be calculated as shown in Equation (9). The total effect of parameter i , including the direct and interactions with other parameters can be expressed according to Equation (10). Numerical estimation of the first and total-order Sobol indices was implemented according to Saltelli et al. (2010).

3. RESULTS

3.1. Quantification of Output Uncertainty Using Monte Carlo Simulations

A comprehensive literature review was conducted on relevant articles from 1993 to 2017 to characterize the variance of experimentally measured (or indirectly estimated from experimental data) values of the input parameters necessary to initialize the metabolic model. As shown in **Figure 1**, both the abundance and the minimum and maximum range of the data varied considerably from one parameter to another. Note that **Figure 1** only shows the variation of input parameters that have been reported in the literature, and does not include (single) point values, e.g., the ATP yield per unit of biomass on Propionyl-CoA (K_2) and the half-saturation constant for propionate uptake (K_{HPr}), nor the characteristic variation of initial conditions. The underlying data was normalized by subtracting the mean and dividing by the corresponding standard deviation to illustrate the variance, rather than the difference in magnitude. In general, more complete data was obtained for kinetic parameters, namely the maximum specific rates of anaerobic HAc uptake (q_{HAc}) and aerobic Poly-P formation (q_{PP}), for which 33 and 90 measurements were reported, respectively.

Only 2 measurements were available for the Arrhenius temperature coefficients, as only two studies have quantified the short and long-term temperature dependencies (Brdjanovic et al., 1997, 1998). The sparsity of data for the half-saturation coefficients results from the fact that the same values have been reused in nearly all previous studies concerning EBPR metabolic models. This is likely due to the assumption that, as in ASM models, they do not deviate to any appreciable extent. In the case of metabolic parameters, namely the yield of ATP per unit of NADH_2 (δ) and the phosphate transport coefficient (σ), the small number of measurements likely arises from the difficulty in measuring these quantities experimentally. While another factor may be that their deviation from the mean has had no appreciable effect on model predictions, as demonstrated in Zeng et al. (2003b), this conclusion was drawn from a local sensitivity analysis. Whether this conclusion holds for initial conditions other than those applied in the aforementioned study has yet to be determined.

Once characterized, the input uncertainty was propagated through the model to obtain a set of concentration profiles for each permutation of input parameters and initial conditions. **Figure 2** shows a subset obtained by varying the input parameters, while holding the initial conditions constant to

illustrate the variance in model predictions due to input uncertainty. Specifically, the total concentration of VFA ($S_{VFA,i}$) in the influent was 1.5 C-mmol/l, the ratio of HAc to HPr ($r_{HAc/HPr,i}$) was 0.67, and the ratio of PO_4 to VFA ($r_{P/C,i}$) was 1.5. The initial concentration of biomass was 4.43 C-mmol/l, and the initial fractions of PHA ($X_{PHA,i}$), glycogen ($X_{Gly,i}$), and poly-P ($X_{PP,i}$) were 0.14, 0.36, and 0.3, respectively, in the first part of this study. When HAc and HPr were not fully consumed in the anaerobic phase, the model disregarded aerobic consumption of carbon substrates, so as not to influence the model outputs with other aerobic processes due to aerobic consumption of VFA.

The uncertainty of model predictions varied from one output variable to another. In addition, the output variance was also dependent on time, in line with changing redox conditions. Most notably, the concentration of biomass was insensitive to any input uncertainty in the anaerobic phase, since according to the model no growth is expected in these conditions, but diverges significantly in the aerobic phase. A similar trend was observed for concentration profiles of intra-cellular glycogen (X_{Gly}).

The intra-cellular concentrations of PHV and PH_2MV were less sensitive in the anaerobic compared to the aerobic phase, whereas the concentration of PHB showed greater variance in both redox environments. However, it must be noted that PHB was also the most abundant fraction of intra-cellular PHA, and one that results from the metabolism of both HAc and HPr, whereas PHV results mainly from the metabolism with HPr. Consequently, the low variance in the final value of PHV and PH_2MV at the end of the anaerobic phase was likely due to these factors aggravated by the lower $S_{HPr,i}$. Likewise, the lower variance at the end of the aerobic phase results from the near complete depletion of all PHA components. As shown in **Figure 2**, the final concentration of PHV and PH_2MV in the majority of simulations was below 0.1 C-mmol/l.

The mean (over time) of the PAO output concentration profiles for each complete anaerobic-aerobic cycle are summarized in the form of histograms in **Figure 3**. These were obtained by taking into account the uncertainty associated with the input parameters alongside the initial conditions, e.g., concentration of carbon sources in the influent, the prevailing temperature, pH and the starting content of intra-cellular polymers in the biomass cells, for a total of 164,000 input samples. While there is considerable variance for all model outputs, the distribution of the concentrations of carbon compounds (extra-cellular VFA and intra-cellular PHA) were skewed toward the lower end while the rest were centered around a particular mean. The mean concentration profiles of phosphate in the bulk liquid phase and biomass were subject to the highest degree of variance, whereas the concentrations of intra-cellular PHV, poly-P and PH_2MV deviated the least under different permutations of input parameters. The mean intra-cellular poly-P concentration (X_{PP}) was the only variable with a bi-modal distribution, with a value near zero in over half of the outcomes. Based on the skewness of the final outcomes, it can be expected that the coefficient of determination, as a proxy of the validity of SRC, would also suffer.

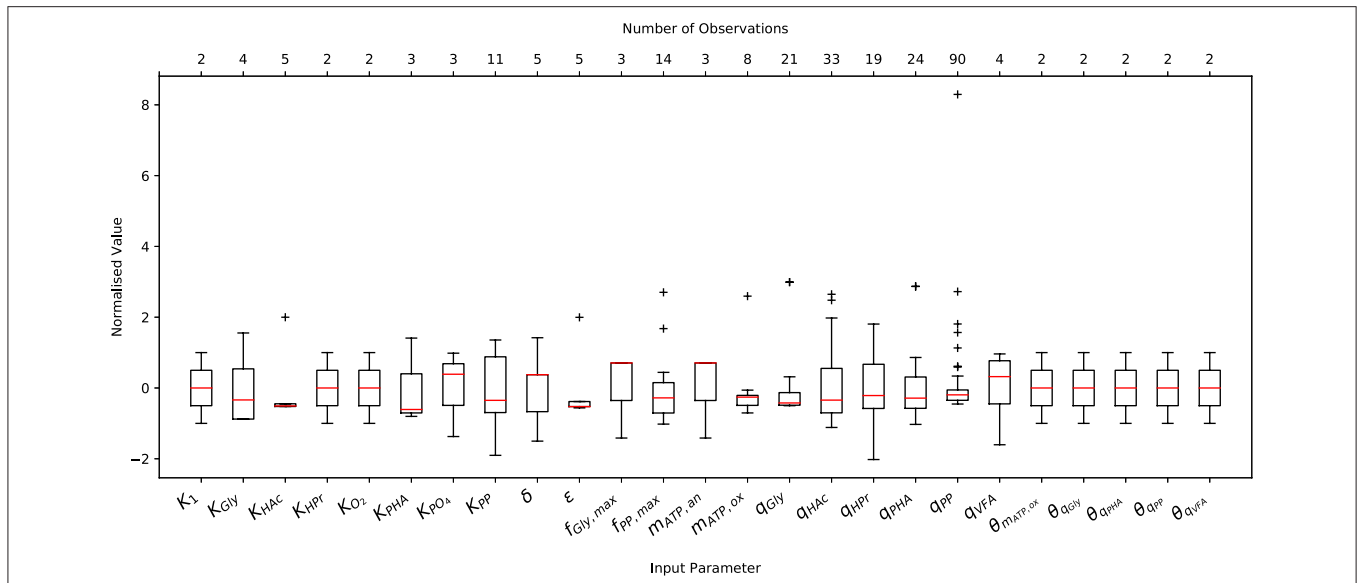


FIGURE 1 | Variation in the measured or predicted values of the input parameters normalized by their respective mean and standard deviation. The boxplot includes the normalized median (red horizontal line within the box), interquartile range (IQR), i.e., from 25th to 75th percentile (box), 1.5 times the lowest and highest percentiles (whiskers) and the number of observations (upper x-axis).

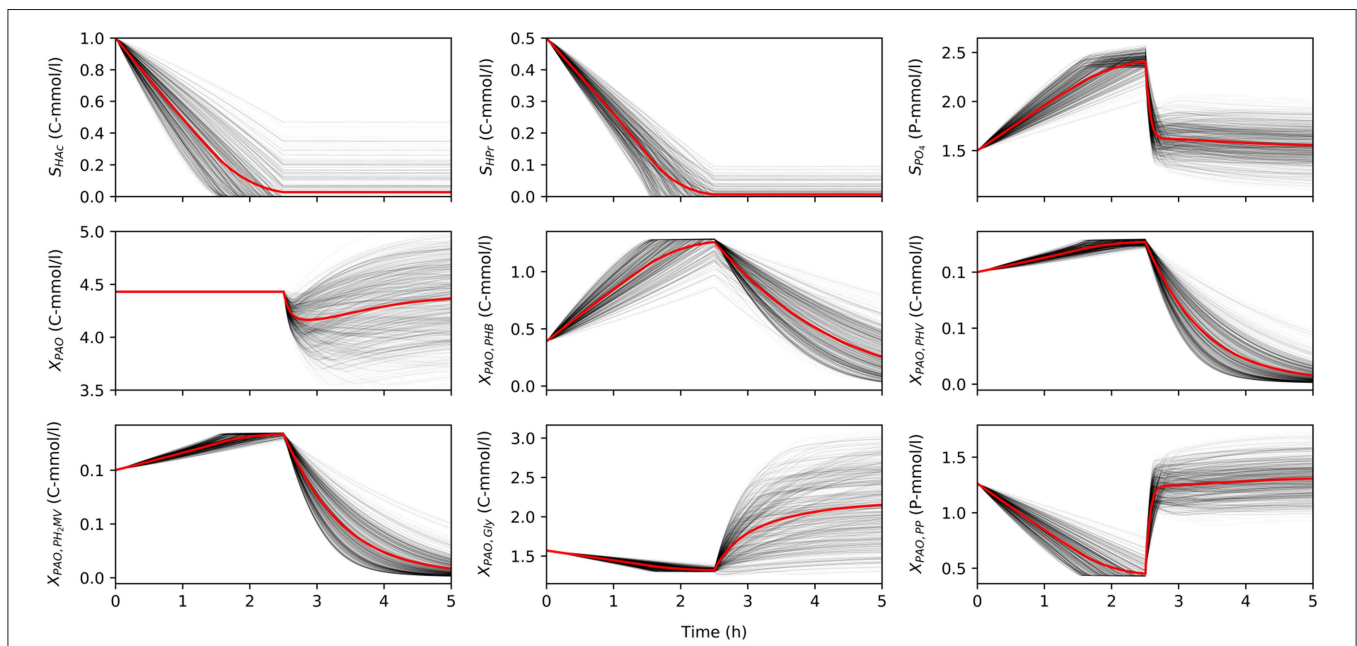


FIGURE 2 | The variance of model predictions given the uncertainty of the input parameters for one set of initial conditions in a 5 h anaerobic-aerobic cycle (2.5 h each). Mean of the concentration profile is highlighted in red.

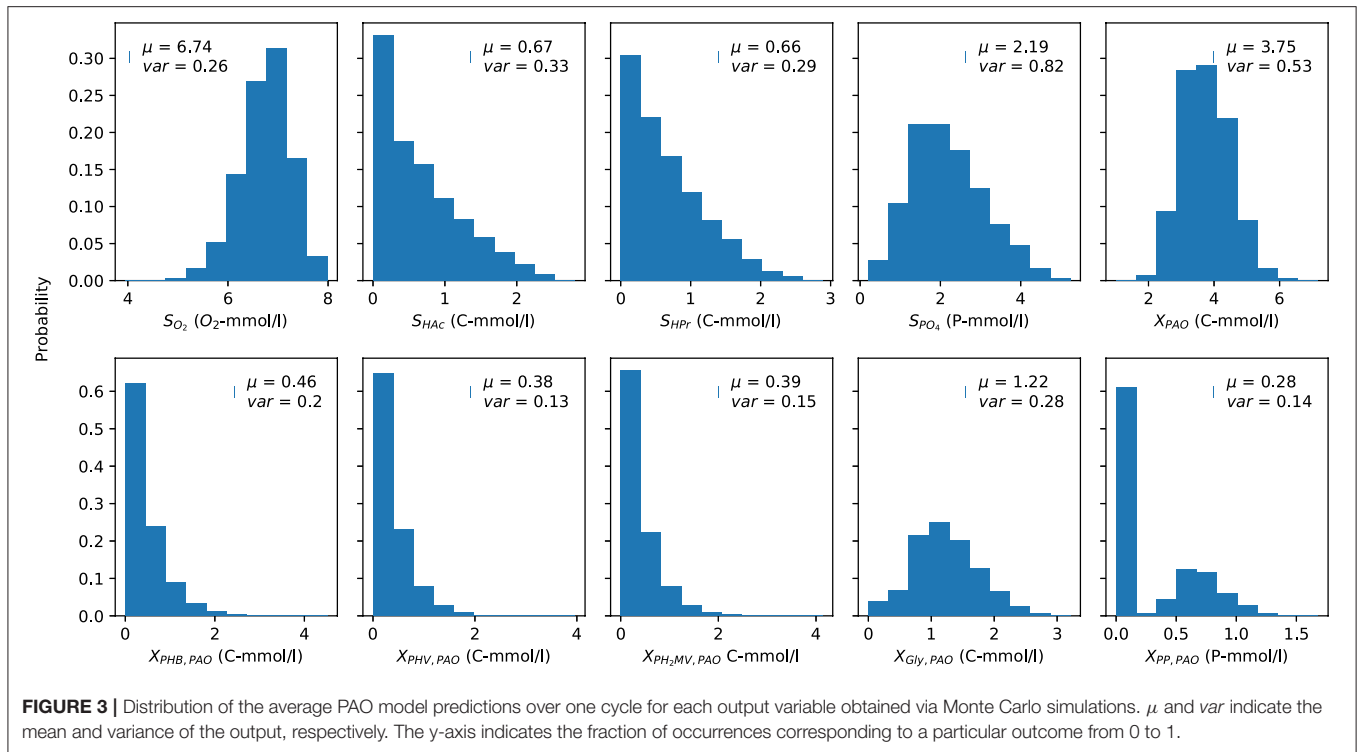
3.2. Sensitivity Analysis

3.2.1. First-Order Effects

Two methods were used to assess the first-order effects of each input parameter on the output parameters: estimation of SRC and the Sobol method for variance decomposition. Unlike local sensitivity analyses, these methods assess the effect of a given input parameter

on an output variable under all possible permutations of different inputs.

For a linear model, the square of the SRC, i.e., β^2 , would be equal to the corresponding first-order index obtained via the Sobol method (see section 2.3.2). The value of β lies between -1 and +1, indicating both the direction as well as the magnitude of the effect of a given input parameter on the output variable.



However, the robustness of this sensitivity measure is dependent on the degree of linearity between the input with respect to the output. As such, the validity of SRC as an estimate of the first-order effect depends on the coefficient of determination (R^2). As a rule of thumb, this value should be ≥ 0.7 for the variance of the output variable to be sufficiently correlated to the variance of the input parameter (Campolongo and Saltelli, 1997). Although the sum of all β^2 with respect to a given output for a linear model would be equal to 1, owing to the non-linear form of the underlying equations in this work, it was expected to be less than one (Saltelli et al., 2008).

Figure 4 compares the β^2 coefficients and first-order Sobol sensitivity indices as estimates of the direct effect of each input on a given output. The higher the value of the index, the higher the effect of that input parameter on the corresponding output parameter relative to the other inputs. Specifically, the Sobol index indicates the relative degree of output that arises from the uncertainty the input in question. The input parameters were listed in order of greatest to lowest influence on the model output uncertainty, reading from left to right. Likewise, the output variables were ordered from most to least uncertain, reading from top to bottom.

The first-order Sobol index was found to be greater than that of the SRC in every instance. This was in line with expectations, as the calculation of the first-order Sobol index does not rely on the assumption of linearity, and is therefore able to capture non-linear effects as well. Furthermore, the R^2 corresponding to each output ranged from only 0.006 to 0.018, indicating that SRC is not an appropriate sensitivity measure for this metabolic model. This suggests that model simplification by linearization of

any part of the system of differential equations by which model predictions are defined is not a feasible option as long as accurate predictions are desired in the operating range specified by the input uncertainty.

As shown by the first-order Sobol indices, the most influential parameters on the prediction of concentrations in the bulk liquid phase were the initial concentration of PAO biomass ($X_{PAO,i}$) and the intra-cellular concentrations of PHA constituents (defined in this study as fractions of the total initial PHA concentration, $f_{PHB,i}$, $f_{PHV,i}$ and $f_{PH2MV,i}$), followed by that of intra-cellular glycogen ($f_{Gly,i}$). That the variance of components in the liquid phase depends on the concentration of PAO is a given, as the rate at which substrates are utilized is directly proportional to the number of organisms. The importance of intra-cellular PHA and glycogen, particularly on the prediction of dissolved oxygen (S_{O_2}) and VFA concentrations is explained by their role in the transport and storage of carbon sources to survive in alternating feast-famine conditions, serving as sources of energy and reducing agents for substrate uptake, growth and cell maintenance. **Table 1** presents a more specific ranking of the influential input parameters for each output.

Intra-cellular glycogen ranked highly for all liquid phase outputs apart from the PO_4 concentration. The lower influence of $f_{Gly,i}$ on the prediction of S_{PO_4} compared to that of intra-cellular PHA can be explained by the degree of relatedness of the processes in question. While PO_4 uptake for poly-P formation under aerobic conditions is directly dependent on intra-cellular PHA reserves, it is only indirectly dependent on the intra-cellular glycogen concentration, where the availability of PHA is affected

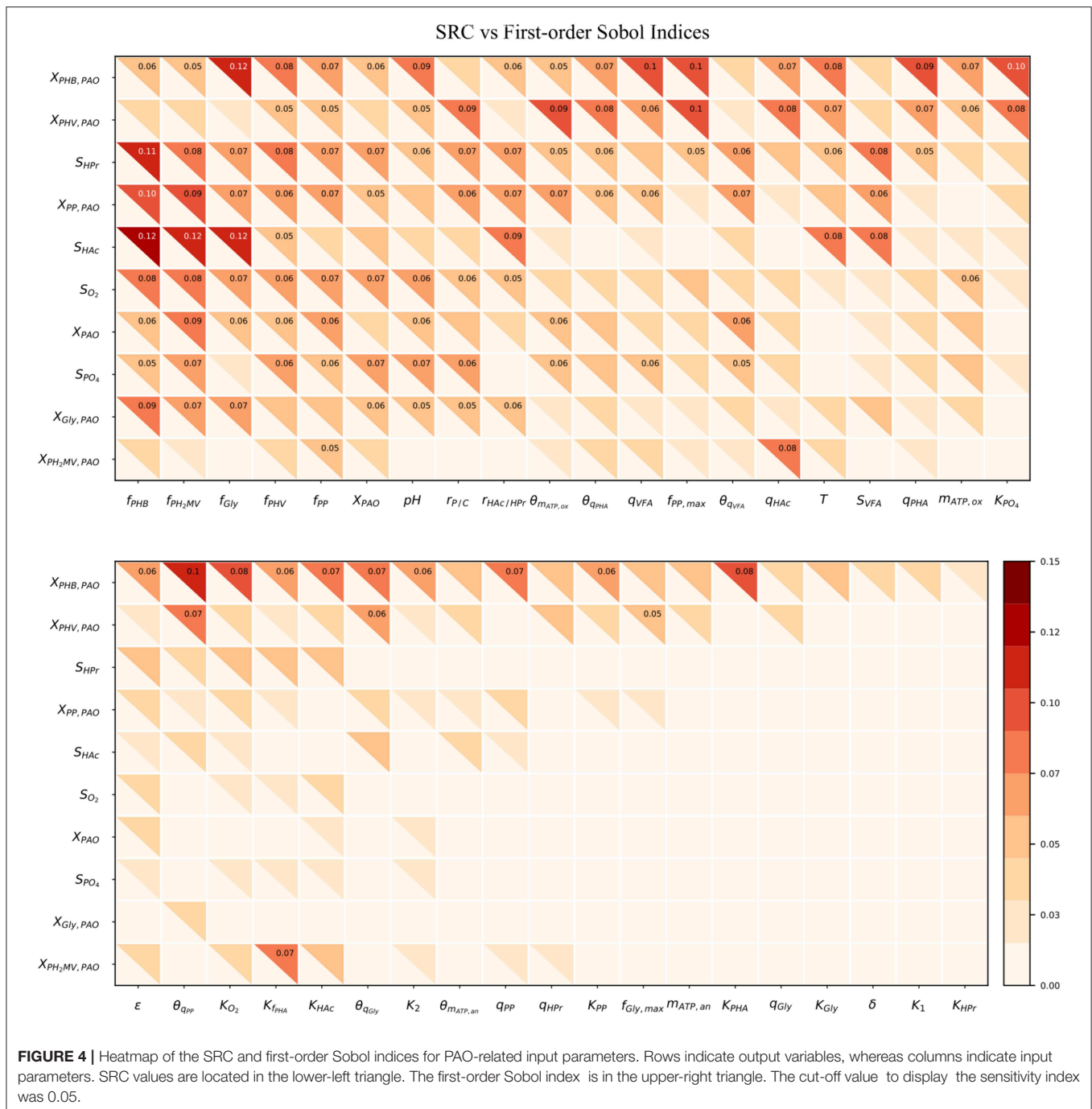


FIGURE 4 | Heatmap of the SRC and first-order Sobol indices for PAO-related input parameters. Rows indicate output variables, whereas columns indicate input parameters. SRC values are located in the lower-left triangle. The first-order Sobol index is in the upper-right triangle. The cut-off value to display the sensitivity index was 0.05.

by glycogen content in the biomass for PHA formation in the preceding anaerobic phase.

The pH was highly influential with respect to the prediction of all liquid phase concentrations, most notably that of S_{PO_4} . The one exception was S_{HAC} . While seemingly surprising, as pH is linked directly to the amount of energy (ATP) required to transport VFA across the cell membrane, this discrepancy could be explained by its involvement in the stoichiometric yield of PHV and PH_2MV formation (as opposed to the more

straightforward conversion to PHB), in addition to glycogen and poly-P degradation during anaerobic PO_4 release in the presence of HPr. In contrast, the stoichiometry with respect to anaerobic glycogen degradation is independent of pH when HAC is the primary VFA.

Although temperature ranked among the most influential input parameters with respect to the intra-cellular concentration of PHB (X_{PHB}), S_{HAC} and S_{HPr} , it did not do so consistently for the remaining outputs. As shown in **Table 1**, it is uncertainty

TABLE 1 | Input parameters ranked according to the most to the least influential with respect to each of the model output variables, as determined by the first-order Sobol sensitivity index.

S_{O_2}	S_{HAc}	S_{HPr}	S_{PO_4}	X_{PAO}	X_{PHB}	X_{PHV}	X_{PH_2MV}	X_{Gly}	X_{PP}
$f_{PH_2MV,i}$	$f_{PHB,i}$	$f_{PHB,i}$	pH	$f_{PH_2MV,i}$	$f_{Gly,i}$	$f_{PP,max}$	q_{HAc}	$f_{PHB,i}$	$f_{PHB,i}$
$f_{PHB,i}$	$f_{Gly,i}$	$S_{VFA,i}$	$X_{PAO,i}$	$f_{PP,i}$	K_{PO_4}	$\theta_{m_{ATP,ox}}$	$K_{f_{PHA}}$	$f_{PH_2MV,i}$	$f_{PH_2MV,i}$
$f_{PP,i}$	$f_{PH_2MV,i}$	$f_{PHV,i}$	$f_{PH_2MV,i}$	$\theta_{q_{VFA}}$	$\theta_{q_{PP}}$	$r_{P/C,i}$	$f_{PP,i}$	$f_{Gly,i}$	$\theta_{m_{ATP,ox}}$
$X_{PAO,i}$	$r_{HAc/HPr,i}$	$f_{PH_2MV,i}$	$f_{PHV,i}$	$f_{PHV,i}$	$\theta_{q_{VFA}}$	$\theta_{q_{PHA}}$	$r_{HAc/HPr,i}$	$f_{HAc/HPr,i}$	$f_{Gly,i}$
$f_{Gly,i}$	T	$X_{PAO,i}$	$r_{P/C,i}$	$f_{Gly,i}$	$f_{PP,max}$	q_{HAc}	$X_{PAO,i}$	$\theta_{m_{ATP,ox}}$	$\theta_{m_{ATP,ox}}$
$f_{PHV,i}$	$S_{VFA,i}$	$r_{P/C,i}$	q_{VFA}	pH	q_{PHA}	K_{PO_4}	pH	$r_{HAc/HPr,i}$	$r_{HAc/HPr,i}$
pH	$f_{PHV,i}$	$f_{Gly,i}$	$\theta_{m_{ATP,ox}}$	$\theta_{m_{ATP,ox}}$	pH	q_{PHA}	$r_{P/C,i}$	$f_{PP,i}$	$f_{PP,i}$
$r_{P/C,i}$	$X_{PAO,i}$	$r_{HAc/HPr,i}$	$f_{PP,i}$	$f_{PHB,i}$	K_{PHA}	T			$r_{P/C,i}$
$m_{ATP,ox}$		$f_{PP,i}$	$f_{PHB,i}$		K_{O_2}	$\theta_{q_{PP}}$			$f_{PHV,i}$
$r_{HAc/HPr,i}$		$\theta_{q_{VFA}}$	$\theta_{q_{VFA}}$		$f_{PHV,i}$	q_{VFA}			$S_{VFA,i}$
		$\theta_{q_{PHA}}$			T	$\theta_{q_{Gly}}$			q_{VFA}
		T			K_{HAc}	$m_{ATP,ox}$			$\theta_{q_{VFA}}$
		pH			$f_{PP,i}$	$f_{PHV,i}$			$X_{PAO,i}$
		$\theta_{m_{ATP,ox}}$			$\theta_{q_{Gly}}$	$f_{PP,i}$			
		q_{PHA}			q_{HAc}	$f_{Gly,max}$			
		$f_{PP,max}$			$m_{ATP,ox}$	pH			
		q_{HAc}			q_{PP}				
					$\theta_{q_{PHA}}$				
					$K_{f_{PHA}}$				
					K_{PP}				
					ϵ				
					$X_{PAO,i}$				
					$r_{HAc/HPr,i}$				
					K_2				
					$f_{PHB,i}$				
					$f_{PH_2MV,i}$				
					$\theta_{m_{ATP,ox}}$				

Input parameters whose sensitivity was lower than the average are not shown.

regarding the Arrhenius coefficients (θ_i) that drives the variance in model outputs in general. When both temperature and Arrhenius coefficient are influential, with the exception of S_{HAc} , the Arrhenius coefficient is more important. Although seemingly surprising, given comparable variances of the corresponding input uncertainties (50% around the mean), this discrepancy may be due to the exponential form of the temperature dependence of the kinetic processes. In general, the Arrhenius coefficient for VFA uptake ($\theta_{q_{VFA}}$), PHA formation ($\theta_{q_{PHA}}$) and poly-P formation ($\theta_{q_{PP}}$) are more influential for all outputs. The Arrhenius coefficient corresponding to aerobic maintenance ($\theta_{m_{ATP,ox}}$) ranks notably high with respect to the prediction of X_{PAO} , the intra-cellular concentration of PHV (X_{PHV}), S_{PO_4} and the concentration of intra-cellular poly-P (X_{PP}).

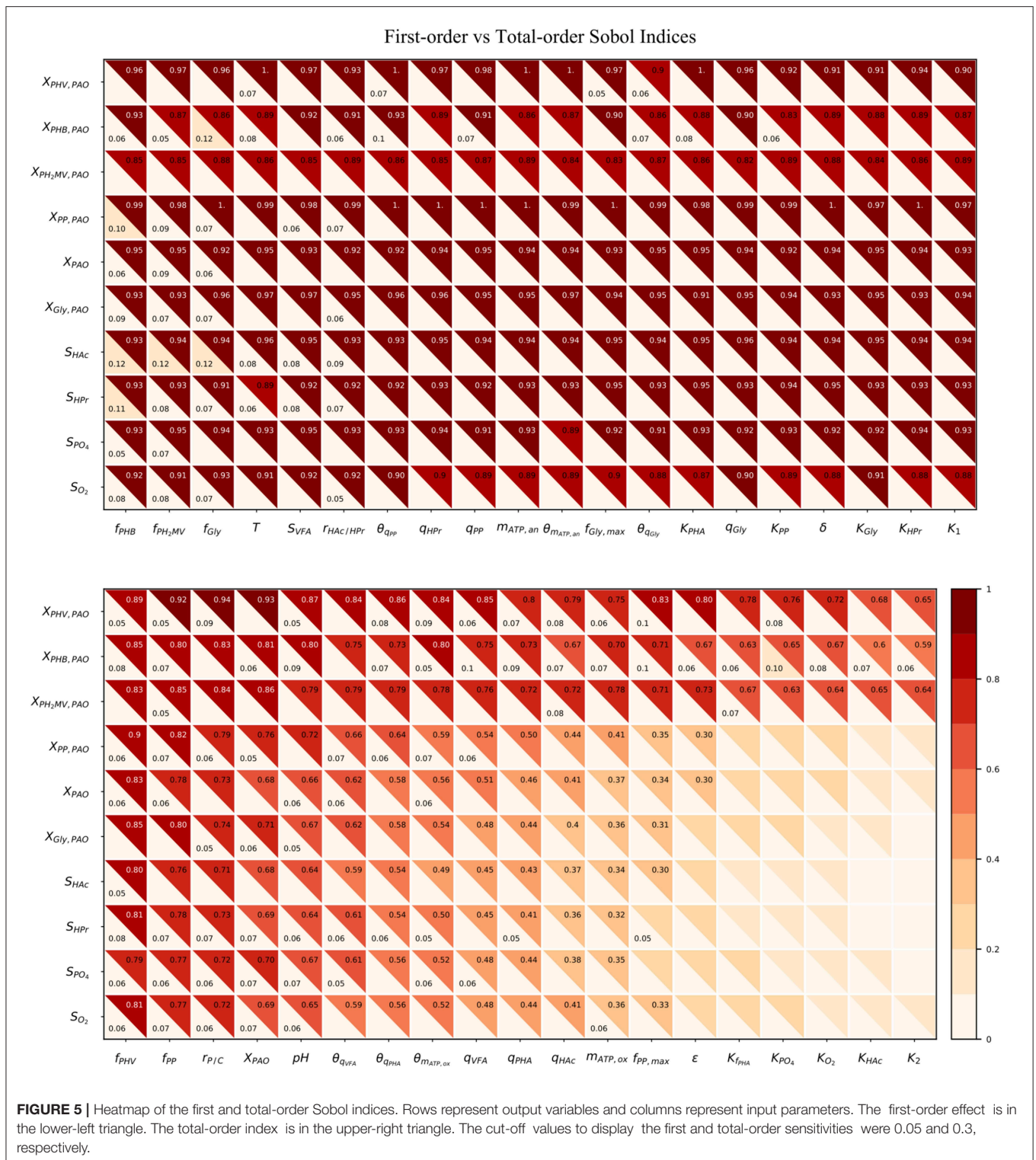
3.2.2. Comparing First-Order and Total-Order Effects

The Sobol method decomposes the variance of the output variables (or sets of variables) into fractions attributable to either individual input parameters or sets of inputs. Unlike SRC, the resultant sensitivity indices take into account non-linear responses, as well as interactions between different input parameters (Saltelli and Annoni, 2010). The total-order sensitivity index quantifies the aggregate influence of a given

input parameter, inclusive of its interaction with all other inputs on a given output variable.

As shown in **Figure 5**, the total-order sensitivity was found to be greater than that of the first-order in all instances. This, coupled with the fact that the confidence interval was on average approximately 8% of the sensitivity index, confirms that the total-order indices were an accurate apportionment of the input uncertainties to the output variance (Saltelli et al., 2010). Although the relative influence of the input parameters on the output variables was identical for both first and total-order sensitivities, the magnitude of the effect was more pronounced in the latter. Whereas the first-order indices ranged from 0 to 0.12, the maximum value of the total-order indices was 1. Differences between the first and sensitivities show that the major cause of variance in the model outputs was not the uncertainty of individual parameters. Rather, it was the interaction among the inputs, revealed by simultaneous perturbation of input parameters using the Monte Carlo method.

As seen from the lower-right quadrant of **Figure 5**, the total effect of the half-saturation constants for f_{PHA} , PO_4 , O_2 , HAc , as well as the yield of ATP per unit of biomass on propionyl-CoA (K_2) and the aerobic phosphate transport coefficient (ϵ) on the output variables were relatively small. The exception to



this trend were the intra-cellular PHA concentrations : X_{PHB} , X_{PHV} and X_{PH_2MV} . However, as the distribution of the total effect was relatively uniform beyond the limits of this quadrant, it was difficult to assess opportunities for model simplification based solely on the raw values of the Sobol indices. In contrast to the

first-order sensitivities, it would no longer be feasible to assign fixed values to the aforementioned input parameters, as their influence on the concentration of intra-cellular PHA components are not negligible when higher order interactions are taken into account.

3.2.3. Relative Contribution of Total Effects

Although the total-order indices were large in many cases, their significance must be taken in the context of the other parameters. It is known that for a linear model the sum of the first-order Sobol indices must be equal to 1. Further, the first-order sensitivities would be equal to the total-order ones in the absence of interactions. Based on this notion, the total-order indices for each output variable were normalized such that their sum would be equal to 1. The data was then separated according to four principal groups of input parameters: (1) metabolic parameters, (2) kinetic parameters, (3) Arrhenius temperature coefficients and (4) initial process conditions. The corresponding heat-maps of the transformed total-order sensitivity indices for each group are shown in **Figures 6A–D**. **Table 2**, discussed in section 4, presents the most influential inputs with respect to each output variable.

From **Figure 6A**, it can be seen that variance in the prediction of liquid phase concentrations, i.e., S_{O_2} , S_{HAc} , S_{HPr} , S_{PO_4} , and X_{PAO} was due primarily to the uncertainty of a subset of metabolic parameters. This is in contrast to variance in the prediction of intra-cellular PHA concentrations, which were due to the uncertainty associated with all metabolic parameters more evenly. Similarly, the prediction of liquid phase concentrations was more sensitive to a smaller subset of kinetic parameters than the prediction of intra-cellular components, specifically the PHA constituents (**Figure 6B**). The largest difference between the influential and non-influential parameters was found with respect to S_{HAc} and S_{HPr} . This was in line with expectations, as the kinetics of anaerobic VFA uptake tend to be the fastest in the metabolic model of interest. It was curious to note that the effect of the maximum intra-cellular fraction of glycogen ($f_{Gly,max}$) was more influential than that of poly-P ($f_{PP,max}$) with respect to all outputs. This may be explained by the difference in input uncertainty, as the reported values of f_{Gly} deviate more from each other among the different studies.

As shown in **Figure 6C** and by the ranking in **Table 2**, the most influential temperature coefficients were those corresponding to anaerobic maintenance ($\theta_{mATP,an}$), aerobic poly-P formation (θ_{qPP}) and glycogen formation (θ_{qGly}), respectively. The greatest difference between the influential and non-influential parameters was with respect to the VFA concentration in the liquid phase and the intra-cellular concentration of glycogen (X_{Gly}).

With regards to the input parameters defining the initial state and the prevailing process conditions, the most influential were the initial intra-cellular fraction of glycogen and PHA, followed by the total VFA concentration in the influent ($S_{VFA,i}$) and the closely related ratio of HAc to HPr ($r_{HAc/HPr,i}$), as shown in **Figure 6D**. Compared with metabolic and kinetic parameters, the contribution of input uncertainties on output variance was spread out more evenly among the Arrhenius temperature coefficients and prevailing process conditions. This indicates that reducing input uncertainty in the model's dependence on temperature and detailed characterization of the influent, inclusive of the intra-cellular components in the biomass, will lead to significant improvements to the accuracy of model predictions with respect to all output variables. In contrast, more

careful consideration in the course of parameter prioritization is required for the metabolic and kinetic parameters, given the relatively larger discrepancy between influential and non-influential ones.

4. DISCUSSION

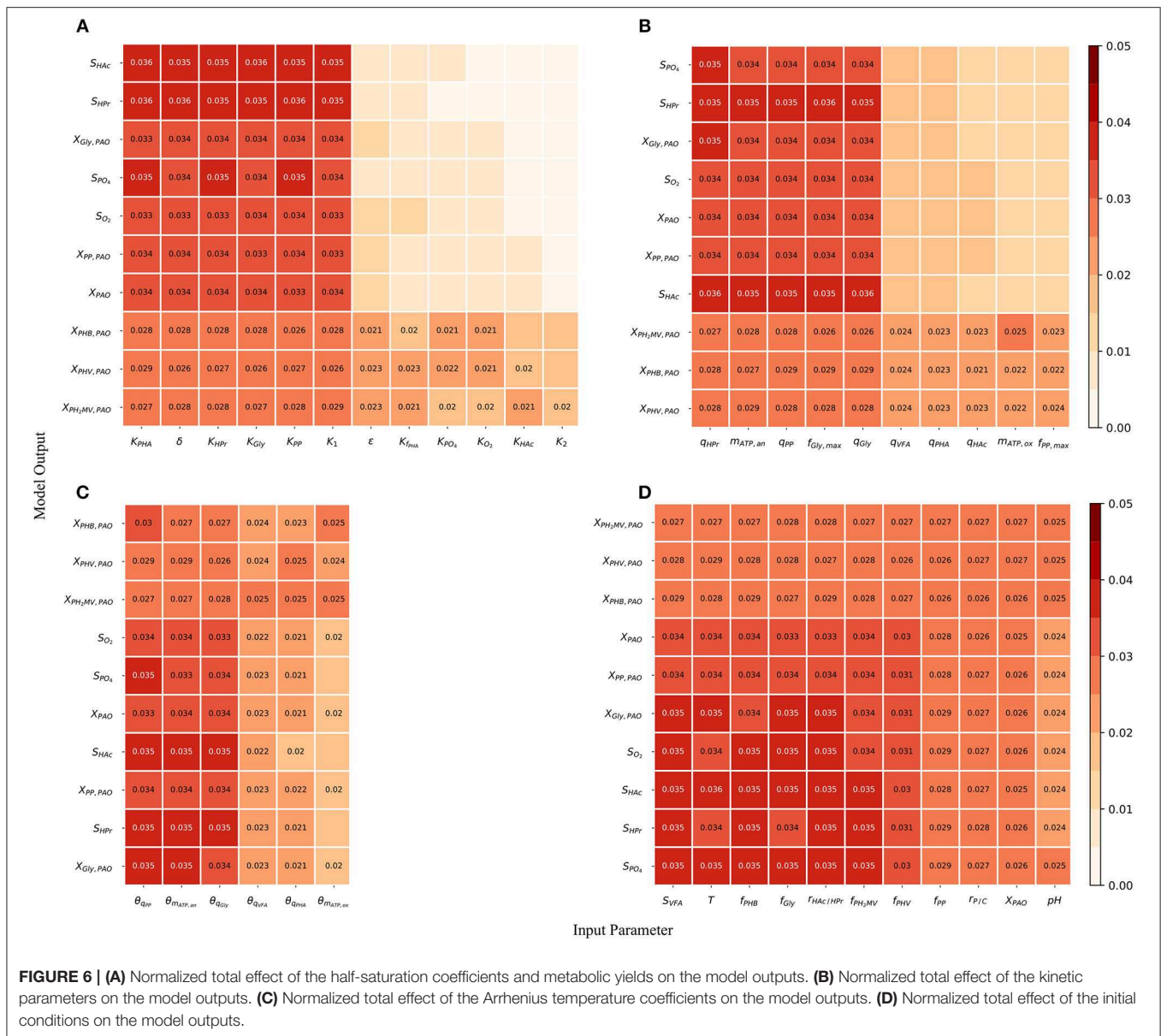
4.1. Opportunities for Model Simplification Based on First-Order Sensitivities

Linearization of at least parts of the model would result in significantly lower computational cost, and would thereby lower the barrier for the effective use of metabolic models in process monitoring and control. It would also result in more efficient use of resources for the study of more detailed scenarios. This would translate to shorter simulation times and knock-on benefits of the accompanying experimental design, where simulations are used to screen for subsequent laboratory or field work. The validity of SRC depends on the associated R^2 values, which indicate the degree to which interactions between input parameters and output variables can be linearized. By extension, given that metabolic models, particularly those that consider the interaction between multiple species of PAO and GAO are over-parametrized (Yagci et al., 2004), coefficients of determination also indicate the feasibility of model simplification.

Given the small SRC values (< 0.1) and the accompanying R^2 (< 0.018) for all output variables, there is little room for model simplification by linearization of its parts. However, this conclusion is drawn from results obtained by simultaneous perturbation of the model's input parameters, inclusive of the initial conditions. Consequently, although linearization is not feasible for the general case, it may be possible where the initial conditions, i.e., information concerning the concentrations in the bulk liquid phase and the intra-cellular state of PAO cells at the beginning of the anaerobic phase, as well as T and pH, are well-defined and not subject to an appreciable extent of fluctuation from one cycle to another. This can be expected to be the case for laboratory-scale experiments. It may also hold true for wastewater treatment plants where the influent is well characterized and not subject to a significant degree of fluctuation, or if the variation itself is well defined. Based on **Figure 2**, linearization may be most readily applied to the prediction of HAc and HPr concentrations. It may also be feasible if the system were to be modeled as discrete stages (with redox conditions defining model boundaries), as the predicted behavior follows distinct patterns in the anaerobic and aerobic stages. However, the applicability of such a model would be limited to systems where the redox conditions can be strictly ensured, e.g., no excess of dissolved oxygen in the anaerobic phase.

4.1.1. Influence of Initial Conditions

As shown in **Figure 5** and highlighted in **Table 1**, the most influential parameters tended to describe process conditions, starting with the initial intra-cellular fractions (PHA, glycogen and poly-P), followed by the biomass concentration ($X_{PAO,i}$), pH, T, and the influent ($r_{P/C,i}$, $r_{HAc/HPr,i}$, $S_{VFA,i}$). It is important to note that the effect of these parameters was typically higher than other metabolic or kinetic parameters. Uncertainty in initial



conditions affects the variance of model outputs to a greater degree than input parameters associated with process kinetics or stoichiometry. This finding poses two challenges.

On one hand, if metabolic models are to be applied in full-scale wastewater treatment works, detailed characterization of the influent, namely of the intra-cellular components in biomass, will significantly reduce the variance of all model outputs, leading to more accurate predictions. While characterization of liquid phase components, i.e., the concentrations of HAc, HPr, and PO₄, as well as T and pH may be readily achieved by installation of simple sensors and basic testing protocols, the quantification of biomass of specific microbial groups and their corresponding intra-cellular fractions remains a challenge due to cost and/or complexity issues.

On the other hand, the uncertainty of experimental methods used to measure these quantities should be acknowledged

and improved where possible. While quantification of the *Accumulibacter* PAO population in both laboratory and full-scale reactors using culture-independent molecular methods has become standard practice (López-Vázquez et al., 2007a; Oehmen et al., 2010b; Lanham et al., 2014), there is some controversy over the accuracy of such measurements. The relative abundances measured by different methods, e.g., quantitative fluorescence *in situ* hybridization (qFISH) via target sequences in the 16S rRNA molecule or real-time quantitative polymerase chain reaction (qPCR) via the poly-P kinase gene (*ppk1*) are not always in complete agreement (Fukushima et al., 2007). This may be attributed to factors such as the sensitivity to low cell counts, differences in cellular metabolic activities, heterogeneities in experimental protocol (e.g., sample preparation), or sample characteristics (Moter and Göbel, 2000; van Loosdrecht et al., 2016; Huber et al., 2018). This is not to mention the added

TABLE 2 | Input parameters ranked according to the most to the least influential with respect to each of the model output variables, as determined by the total-order Sobol sensitivity index.

Group	S _{O₂}	S _{HAc}	S _{HPr}	S _{PO₄}	X _{PAO}	X _{PHB}	X _{PHV}	X _{PH₂MV}	X _{Gly}	X _{PP}
Metabolic	K _{Gly}	K _{Gly}	δ	K _{HPr}	K _{Gly}	δ	K _{PHA}	K ₁	K _{Gly}	δ
	K _{PP}	K _{PHA}	K _{PHA}	K _{PP}	K _{PHA}	K _{HPr}	K _{HPr}	K _{PP}	K _{PP}	K _{HPr}
	K _{HPr}	δ	K _{PP}	K _{PHA}	δ	K _{Gly}	K _{PP}	δ	K ₁	K _{PP}
	K ₁	K ₁	K _{HPr}	K ₁	K _{HPr}	K _{PHA}	K _{Gly}	K _{HPr}	δ	K _{PHA}
	δ	K _{PP}	K _{Gly}	δ	K ₁	K ₁	δ	K _{PHA}	K _{HPr}	K _{Gly}
	K _{PHA}	K _{HPr}	K ₁	K _{Gly}	K _{PP}	K _{PP}	K ₁	K _{Gly}	K _{PHA}	K ₁
Kinetic	q _{Gly}	q _{Gly}	f _{Gly,max}	q _{HPr}	q _{PP}	q _{PP}	m _{ATP,an}	m _{ATP,an}	q _{HPr}	q _{PP}
	f _{Gly,max}	q _{HPr}	m _{ATP,an}	m _{ATP,an}	q _{HPr}	f _{Gly,max}	q _{PP}	q _{PP}	q _{PP}	q _{HPr}
	q _{HPr}	f _{Gly,max}	q _{Gly}	q _{Gly}	m _{ATP,an}	q _{Gly}	q _{HPr}	q _{HPr}	m _{ATP,an}	m _{ATP,an}
	m _{ATP,an}	m _{ATP,an}	q _{HPr}	f _{Gly,max}	q _{Gly}	q _{HPr}	f _{Gly,max}	f _{Gly,max}	q _{Gly}	f _{Gly,max}
	q _{PP}	q _{PP}	q _{PP}	q _{PP}	f _{Gly,max}	m _{ATP,an}	q _{Gly}	q _{Gly}	f _{Gly,max}	q _{Gly}
	Arrhenius coef	θ _{q_{PP}}	θ _{m_{ATP,an}}	θ _{m_{ATP,an}}	θ _{q_{PP}}	θ _{q_{Gly}}	θ _{q_{PP}}	θ _{m_{ATP,an}}	θ _{q_{Gly}}	θ _{m_{ATP,an}}
θ _{m_{ATP,an}}		θ _{q_{Gly}}	θ _{q_{Gly}}	θ _{q_{Gly}}	θ _{m_{ATP,an}}	θ _{m_{ATP,an}}	θ _{q_{PP}}	θ _{q_{PP}}	θ _{q_{PP}}	θ _{m_{ATP,an}}
θ _{q_{Gly}}		θ _{q_{PP}}	θ _{q_{PP}}	θ _{m_{ATP,an}}	θ _{q_{PP}}	θ _{q_{Gly}}	θ _{q_{Gly}}	θ _{m_{ATP,an}}	θ _{q_{Gly}}	θ _{q_{Gly}}
Process condition	f _{Gly,i}	T	f _{PHB,i}	f _{PH₂MV,i}	f _{PH₂MV,i}	f _{PHB,i}	T	r _{HAc/HPr,i}	S _{VFA,i}	f _{Gly,i}
	r _{HAc/HPr,i}	S _{VFA,i}	f _{PH₂MV,i}	S _{VFA,i}	T	S _{VFA,i}	f _{PH₂MV,i}	f _{Gly,i}	T	f _{PHB,i}
	S _{VFA,i}	f _{Gly,i}	S _{VFA,i}	f _{Gly,i}	f _{PHB,i}	r _{HAc/HPr,i}	S _{VFA,i}	X _{PAO,i}	f _{Gly,i}	T
	f _{PHB,i}	f _{PH₂MV,i}	r _{HAc/HPr,i}	f _{PHB,i}	S _{VFA,i}	T	f _{PHB,i}	T	r _{HAc/HPr,i}	r _{HAc/HPr,i}
	f _{PH₂MV,i}	f _{PHB,i}	f _{Gly,i}	r _{HAc/HPr,i}	r _{HAc/HPr,i}	f _{PH₂MV,i}	f _{Gly,i}	f _{PP,i}	f _{PHB,i}	S _{VFA,i}
	T	r _{HAc/HPr,i}	T	T	f _{Gly,i}	f _{Gly,i}	r _{P/C,i}	f _{PHB,i}	f _{PH₂MV,i}	f _{PH₂MV,i}
	f _{PHV,i}	f _{PHV,i}	f _{PHV,i}	f _{PHV,i}	f _{PHV,i}	f _{PHV,i}	X _{PAO,i}	S _{VFA,i}	f _{PHV,i}	f _{PHV,i}
	f _{PP,i}	f _{PP,i}	f _{PP,i}	f _{PP,i}	f _{PP,i}	r _{P/C,i}	r _{HAc/HPr,i}	f _{PH₂MV,i}	f _{PP,i}	f _{PP,i}
	r _{P/C,i}	r _{P/C,i}	r _{P/C,i}	r _{P/C,i}	r _{P/C,i}	X _{PAO,i}	f _{PP,i}	r _{P/C,i}	r _{P/C,i}	r _{P/C,i}
	X _{PAO,i}		X _{PAO,i}	X _{PAO,i}			f _{PHV,i}	f _{PHV,i}		X _{PAO,i}
	pH									

Input parameters whose sensitivity was lower than the average are not shown.

complication of the specificity of the technique to target the desired organism, e.g., the overlap of multiple FISH probes resulting in the over-estimation of *Accumulibacter* (Albertsen et al., 2016).

Similarly, while glycogen is an important source of carbon, energy (ATP) and reducing agents (NADH₂) in both PAO and bacterial metabolisms in other environments, it has proven to be difficult to accurately quantify. Protocols depend on the type of cell and the state of cell-aggregation, i.e., on whether the biomass is floccular or granular (Lanham et al., 2012). Consequently, model predictions have been shown to exceed the measured concentrations of PHA and poly-P (Brdjanovic et al., 2000; Meijer et al., 2002). Lopez-Vazquez et al. (2009) found that current metabolic models consistently overestimated the X_{Gly} concentration by as much as 25%. In view of cumulative errors in long-term multi-cycle simulations, even small errors can lead to significant deviations in prediction accuracy. As such, the problem of accurate, fast and (if possible) simpler quantification of intra-cellular fractions, namely that of glycogen and PHA, requires more attention to reduce prediction error, and thus for the effective deployment of metabolic models in full-scale wastewater treatment plants.

4.1.2. Influence of Temperature Coefficients

As shown in Figure 5 and Table 2, the Arrhenius temperature coefficients for anaerobic maintenance (θ_{m_{ATP,an}}), aerobic glycogen production (θ_{q_{Gly}}) and poly-P formation (θ_{q_{PP}}) rank among the most influential inputs for all output variables. Although relatively abundant data is available for the kinetic uptake rates (Figure 1), relatively limited attention has been devoted to temperature coefficients, apart from the work of Brdjanovic et al. (1997, 1998). One limitation of this work was that the reported temperature dependencies have only been verified to a limited range of operating temperatures (typically between 10 and 20 °C). Given that Arrhenius temperature coefficients are expected to take a constant value in theory, further investigation of the kinetic processes are required to specify the uncertainty in these parameters, and thereby ascertain the accuracy of model predictions.

This is of particular importance in the context of using metabolic models to study the competition between PAO and GAO, as performed by Lopez-Vazquez et al. (2009). It would also enable more accurate study of PAO kinetics at wider temperature ranges, particularly in light of recent, seemingly contradictory findings concerning the population dynamics of PAO and

GAO at higher temperatures (Ong et al., 2014; Shen et al., 2017). Information concerning the temperature coefficients of (*Competibacter*) GAO mirror that of PAO, their values deriving from only two investigations (López-Vázquez et al., 2007b; Lopez-Vazquez et al., 2008a). Recent findings concerning the identity of *Accumulibacter* PAO clades, subdivisions of *Competibacter* GAO, as well as *Defluvicoccus*-related GAO (Oehmen et al., 2010b), call for a more comprehensive re-evaluation of the temperature dependencies to determine their “true” values.

4.1.3. Influence of Kinetic Parameters

After considering initial conditions and temperature parameters, the burden of improving model reliability falls on reducing the input uncertainty associated with kinetic parameters. The priority should be the estimation of the maximum specific rates of anaerobic HPr uptake (q_{HPr}), stand maintenance coefficient ($m_{\text{ATP,an}}$), aerobic poly-P formation (q_{PP}), glycogen production (q_{Gly}) and the aerobic maintenance coefficient ($m_{\text{ATP,ox}}$). Interestingly, these parameters were highly influential on the prediction of all model outputs, despite the fact that data concerning kinetic parameters were the most abundant. This was likely due to the comparatively large variance in the input values, notably in the case of q_{PP} . This suggests that there may be a “natural” level of variation that can be expected in the values of certain kinetic parameters, notably in the maximum rates of change involved in Monod-type expressions, for which additional measurements will not lead to further improvements in model prediction accuracy. This inherent level of variation may be related to the characteristics of the biomass and/or influent of the system in question.

4.1.4. Influence of Metabolic Parameters

As seen from **Table 2**, the most influential metabolic parameters were the half-saturation constants for anaerobic HPr uptake (K_{HPr}), and PHA formation (K_{PHA}), as well as those associated with aerobic glycogen (K_{Gly}), poly-P formation (K_{PP}) and δ . All of these are strongly linked to the transport of HPr across the cell membrane, storage of PHA and subsequent regeneration of glycogen. These parameters may become more influential in determining the composition of the PAO-GAO community in simulations involving both groups of microorganisms. As such, it is important that the uncertainty of these parameters be reduced to improve model prediction reliability, as the preferential uptake of HPr has been shown to be one of the key factors in the competition between the two groups of microorganisms (Oehmen et al., 2005b; Carvalheira et al., 2014b).

4.1.5. Model Simplification by Fixing Non-influential Parameters

The contribution of input uncertainty on the variance of model predictions was found to vary from one output variable to another (**Figure 5**). For the prediction of liquid phase concentrations (S_{O_2} , S_{HAc} , S_{HPr} , S_{PO_4} , X_{PAO}), as well as intra-cellular concentrations of glycogen (X_{Gly}) and poly-phosphate (X_{PP}), the vast majority of uncertainty could be attributed to a smaller subset of input parameters (64%). This suggests

that uncertainty in some input parameters did not affect the prediction accuracy of liquid phase components to an appreciable extent. These include the maximum specific rate of HAc uptake (q_{PAO}) and aerobic PHA degradation (q_{PHA}), the maximum intra-cellular fraction of poly-P ($f_{\text{PP,max}}$), the aerobic maintenance coefficient ($m_{\text{ATP,an}}$), the half-saturation constants for f_{PHA} , PO_4 , O_2 and HAc, as well as the phosphate transport coefficient (ε) and ATP yield per unit of biomass on Propionyl-CoA (K_2). Concerning the prediction of intra-cellular PHA, the total-order sensitivity indices were significant and nearly uniformly distributed among all of these input parameters. Since uncertainty in these input parameters has little effect on the model output outside of the intra-cellular PHA concentrations, they could potentially be fixed to constant values.

Reducing the uncertainty regarding any of the aforementioned input parameters will significantly improve model reliability. In the context of simplifying metabolic models for integration within the wider ASM framework, these results merit further investigation to determine the trade-off between fixing the aforementioned input parameters, especially where the total-order index was lesser than 0.5, and the prediction error with regards to intra-cellular PHA.

4.2. Limitations and Future Perspectives

This work has for the first time assessed the overall effect of input parameter uncertainty on the variance of metabolic model predictions. Most of the propagated uncertainty propagation is not linear from one parameter to one output, but rather due to the interaction of multiple parameters. This brings conclusions in previous studies focused on parameter estimation for metabolic models in EBPR into question. Parameters such as the anaerobic maintenance coefficient ($m_{\text{ATP,an}}$) and the yield of ATP per unit of NADH_2 (δ) have often been justified as constants on the basis of local sensitivity analyses specific to particular conditions, typically in lab-scale reactors, e.g., in Zeng et al. (2003b).

This does not invalidate previous findings, as prediction uncertainty does indeed depend on the particular process conditions. However, in view of applying EBPR metabolic models to the general case, i.e., to characterize systems where the influent cannot be strictly controlled and/or process conditions vary spatially and temporally in the reactor, this study does highlight that local sensitivity analyses are not sufficient to identify regions in which current EBPR metabolic models may be accurately used. More data is needed to determine the true value or at the least the experimental range of a number of parameters. Baring that, if a natural degree of variation exists for certain input parameters, then its distribution should be defined.

Although the input space was sampled to conform to the assumptions of the metabolic model, i.e., within a narrow range of temperatures, pH and an influent mixture of HAc and HPr, the predictions may not be valid in situations where the PAO population is subject to limiting conditions for extended periods of time. For instance, PAO have been found to shift from PHA to glycogen accumulation in anaerobic conditions following a lack of PO_4 in the influent (Acevedo et al., 2014, 2017)—a phenotypic behavior characteristic of their GAO competitors. Similarly, PAO have been shown to employ the tricarboxylic

acid (TCA) cycle to supplement the production of reducing agents to cope with a lack of VFA in the influent, or due to glycogen depletion, possibly as a result of prolonged periods of cellular maintenance (Lanham et al., 2013, 2014). It has also been found that PAO require a longer period of adjustment compared to GAO following extended starvation periods (Vargas et al., 2013; Carvalheira et al., 2014c). As such, the model would require additional considerations of the kinetic and metabolic parameters, or (more likely) structural modifications to the formulation of kinetic rates in order to accurately reflect observations of such behavior in response to or during recovery from long-term duress.

This work used SRC and the Sobol method to conduct a global sensitivity analysis of a metabolic model for PAO. Various groups of input parameters were considered, including metabolic parameters, kinetic parameters, Arrhenius temperature coefficients and the initial conditions required to initialize the model. However, certain factors often held constant in laboratory-scale studies yet crucial to the operation of full-scale wastewater treatment plants were not accounted for. These included the sludge retention time (SRT), hydraulic retention time (HRT) and the lengths of the anaerobic and aerobic phases, which would often be the only means by which to adjust process conditions without interfering with the influent in the event of process disruption. Temperature and pH were also assumed constant over any given anaerobic-aerobic cycle.

Further, this work did not consider an anoxic phase, for which metabolic models have been formulated previously, e.g., in Oehmen et al. (2010b). This was for practical reasons, as the additional burden of quantifying the effect of uncertainty in parameters relevant to anoxic conditions would lead to an exponential increase in the number of model evaluations required to converge to the true value of (variance-based) sensitivity indices. Nevertheless, given the growing interest in anaerobic-aerobic-anoxic EBPR, owing to the potential of further energy savings from reduced aeration requirements, as well as simultaneous removal of nitrogen alongside phosphorus, future work may first consider the evaluation of parameter importance using computationally less expensive screening methods, e.g., the elementary effects (EE) method, where the convergence of sensitivity indices is less dependent of the number of model evaluations (Saltelli et al., 2008). In this way, non-influential parameters would be identified prior to more detailed investigations of the interactions between those input parameters which contribute most to the variance in model predictions. To account for interactions between different inputs, higher-order sensitivities should be accounted for. This is typically achieved by variance-based methods e.g., Sobol indices (Saltelli et al., 2008) and Mara's (2009) extension of RBD-FAST, or density-based methods such as those detailed in Plischke et al. (2013) and Pianosi and Wagener (2015).

Finally, the driving force for the development of metabolic models has been the prediction of conditions more favorable to PAO than GAO. As such, despite the absence of evidence suggesting co-metabolism, i.e., the lack of dependence of one organism on metabolites synthesized by the other, it would

be worthwhile to complement the results obtained in this work with uncertainties in the prediction of concentrations for both *Competibacter* and *Defluvicoccus*-related GAO. This, as well as the differentiation between different PAO clades would merit further investigation to understand the uncertainty not only in the prediction of individual metabolic models, but also how it would propagate in studying the competition between organisms relevant to EBPR.

5. CONCLUSIONS

This work conducted an uncertainty and global sensitivity analysis of a metabolic model describing the behavior of PAO in alternating anaerobic and aerobic EBPR. The input uncertainty was characterized based on the relative abundance of data in the literature for each input parameter required to initialize the model. The input uncertainty was propagated to the output using the Monte Carlo method. Differences in the variance of the mean concentration profiles indicated that concentrations of phosphate in the bulk liquid phase and biomass were the most uncertain, whereas the concentrations of intra-cellular PHV, poly-P and PH₂MV were the least uncertain.

The global sensitivity analysis was conducted using SRC and Sobol sensitivity indices. The analysis comprised a total of 39 input parameters: 12 metabolic parameters, 10 kinetic parameters, 6 Arrhenius temperature coefficients and 9 initial conditions (influent), including temperature and pH. SRC were found to be an inadequate sensitivity measure for this model due to the low degree of linearity between the input parameters and the output variables. Consequently, reduction of model complexity by linearization of its parts is not feasible in the general case, and limited to particular subsets of output variables for more specific scenarios with well-defined initial conditions. Differences between first and total-order Sobol indices indicated that the variance in model predictions was mainly due to interaction effects between combinations of input parameters rather than the uncertainty of individual ones.

The contribution of each of the input parameters on the uncertainty of model predictions varied with the output variable in question. For the prediction of liquid phase concentrations and intra-cellular fractions of glycogen and poly-phosphate, the vast majority of the uncertainty could be attributed to a smaller subset of input parameters (64%). For the prediction of intra-cellular PHA constituents, the contribution was nearly uniformly distributed among all input parameters, indicating a high-degree of interaction. Although the contribution could not be isolated to any particular group of inputs (metabolic parameters, kinetic parameters, temperature coefficients, process conditions), the initial fractions of PHV, PH₂MV and glycogen ranked consistently among the most influential factors, both in terms of direct as well as total effect, suggesting that the value of these parameters should be carefully measured when applying EBPR metabolic models.

This work contributed a step toward a more complete understanding of the uncertainties associated with EBPR

metabolic model predictions, and how to address these uncertainties on an individual basis given knowledge of the corresponding input uncertainty. Possible approaches and prerequisite conditions to simplify metabolic models for PAO, both structurally via linearization, as well as by reducing the number of non-influential variables were illustrated based on the results of the sensitivity analyses. Parameters requiring further experimental consideration were highlighted. This will translate to more accurate prediction of PAO behavior, thus facilitating process monitoring and control. Further, the findings of this work will lead to more informed decision-making in model building and in fundamental investigations of organisms relevant to EBPR systems.

AUTHOR CONTRIBUTIONS

MN formulated and implemented the model, performed the numerical simulations, analyzed the data, and drafted the manuscript. AL and TR contributed to experimental design. All

authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

MN acknowledges support from the Engineering and Physical Sciences Research Council (EPSRC) in the form of his Ph.D. studentship. TR acknowledges support from the Royal Society.

ACKNOWLEDGMENTS

We would like to thank Dr. A. Oehmen and J. Santos for their input in formulating the metabolic model used in this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fbioe.2019.00234/full#supplementary-material>

REFERENCES

- Acevedo, B., Borrás, L., Oehmen, A., and Barat, R. (2014). Modelling the metabolic shift of polyphosphate-accumulating organisms. *Water Res.* 65, 235–244. doi: 10.1016/j.watres.2014.07.028
- Acevedo, B., Murgui, M., Borrás, L., and Barat, R. (2017). New insights in the metabolic behaviour of PAO under negligible poly-P reserves. *Chem. Eng. J.* 311, 82–90. doi: 10.1016/j.cej.2016.11.073
- Albertsen, M., McIlroy, S. J., Stokholm-Bjerregaard, M., Karst, S. M., and Nielsen, P. H. (2016). “*Candidatus* Propionivibrio aalborgensis”: a novel glycogen accumulating organism abundant in full-scale enhanced biological phosphorus removal plants. *Front. Microbiol.* 7:1033. doi: 10.3389/fmicb.2016.01033
- Brđjanovic, D., Logemann, S., van Loosdrecht, M. C., Hooijmans, C. M., Alaerts, G. J., and Heijnen, J. J. (1998). Influence of temperature on biological phosphorus removal: process and molecular ecological studies. *Water Res.* 32, 1035–1048. doi: 10.1016/S0043-1354(97)00322-9
- Brđjanovic, D., Loosdrecht, M. C. v., Hooijmans, C. M., Alaerts, G. J., and Heijnen, J. J. (1997). Temperature effects on physiology of biological phosphorus removal. *J. Environ. Eng.* 123, 144–153. doi: 10.1061/(ASCE)0733-9372(1997)123:2(144)
- Brđjanovic, D., van Loosdrecht, M. C., Versteeg, P., Hooijmans, C. M., Alaerts, G. J., and Heijnen, J. J. (2000). Modeling COD, N and P removal in a full-scale WWTP Haarlem Waarderpolder. *Water Res.* 34, 846–858. doi: 10.1016/S0043-1354(99)00219-5
- Campolongo, F., and Saltelli, A. (1997). Sensitivity analysis of an environmental model: an application of different analysis methods. *Reliabil. Eng. Syst. Safety* 57, 49–69. doi: 10.1016/S0951-8320(97)00021-5
- Carvalho, M., Oehmen, A., Carvalho, G., Eusébio, M., and Reis, M. A. (2014a). The impact of aeration on the competition between polyphosphate accumulating organisms and glycogen accumulating organisms. *Water Res.* 66, 296–307. doi: 10.1016/j.watres.2014.08.033
- Carvalho, M., Oehmen, A., Carvalho, G., and Reis, M. A. (2014b). The effect of substrate competition on the metabolism of polyphosphate accumulating organisms (PAOs). *Water Res.* 64, 149–159. doi: 10.1016/j.watres.2014.07.004
- Carvalho, M., Oehmen, A., Carvalho, G., and Reis, M. A. (2014c). Survival strategies of polyphosphate accumulating organisms and glycogen accumulating organisms under conditions of low organic loading. *Bioresour. Technol.* 172, 290–296. doi: 10.1016/j.biortech.2014.09.059
- Filipe, C. D., Daigger, G. T., and Grady, C. (2001a). Effects of pH on the rates of aerobic metabolism of phosphate-accumulating and glycogen-accumulating organisms. *Water Environ. Res.* 73, 213–222. doi: 10.2175/106143001X139191
- Filipe, C. D., Daigger, G. T., and Grady, C. (2001b). A metabolic model for acetate uptake under anaerobic conditions by glycogen accumulating organisms: stoichiometry, kinetics, and the effect of pH. *Biotechnol. Bioeng.* 76, 17–31. doi: 10.1002/bit.1022
- Filipe, C. D., Daigger, G. T., and Grady, C. (2001c). Stoichiometry and kinetics of acetate uptake under anaerobic conditions by an enriched culture of phosphorus-accumulating organisms at different pHs. *Biotechnol. Bioeng.* 76, 32–43. doi: 10.1002/bit.1023
- Flowers, J. J., He, S., Yilmaz, S., Noguera, D. R., and McMahon, K. D. (2009). Denitrification capabilities of two biological phosphorus removal sludges dominated by different ‘*Candidatus* Accumulibacter’ clades. *Environ. Microbiol. Rep.* 1, 583–588. doi: 10.1111/j.1758-2229.2009.00090.x
- Fukushima, T., Uda, N., Onuki, M., Satoh, H., and Mino, T. (2007). Development of the quantitative PCR method for *Candidatus* ‘Accumulibacter phosphatis’ and its application to activated sludge. *J. Water Environ. Technol.* 5, 37–43. doi: 10.2965/jwet.2007.37
- Herman, J., and Usher, W. (2017). SALib: an open-source Python library for sensitivity analysis. *J. Open Sour. Softw.* 2:97. doi: 10.21105/joss.00097
- Huber, D., von Voithenberg, L. V., and Kaigala, G. (2018). Fluorescence *in situ* hybridization (FISH): history, limitations and what to expect from micro-scale FISH? *Micro Nano Eng.* 1, 15–24. doi: 10.1016/j.mne.2018.10.006
- Kent, E., Neumann, S., Kummer, U., and Mendes, P. (2013). What can we learn from global sensitivity analysis of biochemical systems? *PLoS ONE* 8:e79244. doi: 10.1371/journal.pone.0079244
- Kuba, T., Murnleitner, E., Van Loosdrecht, M., and Heijnen, J. (1996). A metabolic model for biological phosphorus removal by denitrifying organisms. *Biotechnol. Bioeng.* 52, 685–695.
- Lanham, A. B., Oehmen, A., Saunders, A. M., Carvalho, G., Nielsen, P. H., and Reis, M. A. (2013). Metabolic versatility in full-scale wastewater treatment plants performing enhanced biological phosphorus removal. *Water Res.* 47, 7032–7041. doi: 10.1016/j.watres.2013.08.042
- Lanham, A. B., Oehmen, A., Saunders, A. M., Carvalho, G., Nielsen, P. H., and Reis, M. A. (2014). Metabolic modelling of full-scale enhanced biological phosphorus removal sludge. *Water Res.* 66, 283–295. doi: 10.1016/j.watres.2014.08.036
- Lanham, A. B., Ricardo, A. R., Coma, M., Fradinho, J., Carvalho, M., Oehmen, A., et al. (2012). Optimisation of glycogen quantification in mixed microbial cultures. *Bioresour. Technol.* 118, 518–525. doi: 10.1016/j.biortech.2012.05.087
- Liu, W.-T., Nakamura, K., Matsuo, T., and Mino, T. (1997). Internal energy-based competition between polyphosphate-and glycogen-accumulating bacteria in

- biological phosphorus removal reactors—effect of pc feeding ratio. *Water Res.* 31, 1430–1438. doi: 10.1016/S0043-1354(96)00352-1
- López-Vázquez, C. M., Hooijmans, C. M., Brdjanovic, D., Gijzen, H. J., and van Loosdrecht, M. (2007b). A practical method for quantification of phosphorus- and glycogen-accumulating organism populations in activated sludge systems. *Water Environ. Res.* 79, 2487–2498. doi: 10.2175/106143007X220798
- Lopez-Vazquez, C. M., Hooijmans, C. M., Brdjanovic, D., Gijzen, H. J., and van Loosdrecht, M. C. (2008a). Factors affecting the microbial populations at full-scale enhanced biological phosphorus removal (EBPR) wastewater treatment plants in The Netherlands. *Water Res.* 42, 2349–2360. doi: 10.1016/j.watres.2008.01.001
- Lopez-Vazquez, C. M., Oehmen, A., Hooijmans, C. M., Brdjanovic, D., Gijzen, H. J., Yuan, Z., et al. (2009). Modeling the PAO–GAO competition: effects of carbon source, pH and temperature. *Water Res.* 43, 450–462. doi: 10.1016/j.watres.2008.10.032
- López-Vázquez, C. M., Song, Y.-I., Hooijmans, C. M., Brdjanovic, D., Moussa, M. S., Gijzen, H. J., et al. (2007a). Short-term temperature effects on the anaerobic metabolism of glycogen accumulating organisms. *Biotechnol. Bioeng.* 97, 483–495. doi: 10.1002/bit.21302
- Lopez-Vazquez, C. M., Song, Y.-I., Hooijmans, C. M., Brdjanovic, D., Moussa, M. S., Gijzen, H. J., et al. (2008b). Temperature effects on the aerobic metabolism of glycogen-accumulating organisms. *Biotechnol. Bioeng.* 101, 295–306. doi: 10.1002/bit.21892
- Mara, T. A. (2009). Extension of the RBD-FAST method to the computation of global sensitivity indices. *Reliabil. Eng. Syst. Safety* 94, 1274–1281. doi: 10.1016/j.res.2009.01.012
- Maszenan, A., Seviour, R., Patel, B., Schumann, P., Burghardt, J., Tokiwa, Y., et al. (2000). Three isolates of novel polyphosphate-accumulating gram-positive cocci, obtained from activated sludge, belong to a new genus, *Tetrasphaera* gen. nov., and description of two new species, *Tetrasphaera japonica* sp. nov. and *Tetrasphaera australiensis* sp. nov. *Int. J. Syst. Evol. Microbiol.* 50, 593–603. doi: 10.1099/00207713-50-2-593
- Meijer, S., Van Loosdrecht, M., and Heijnen, J. (2001). Metabolic modelling of full-scale biological nitrogen and phosphorus removing WWTP's. *Water Res.* 35, 2711–2723. doi: 10.1016/S0043-1354(00)00567-4
- Meijer, S., Van Loosdrecht, M., and Heijnen, J. (2002). Modelling the start-up of a full-scale biological phosphorus and nitrogen removing WWTP. *Water Res.* 36, 4667–4682. doi: 10.1016/S0043-1354(02)00192-6
- Melkumova, L., and Shatskikh, S. Y. (2017). Comparing ridge and LASSO estimators for data analysis. *Proc. Eng.* 201, 746–755. doi: 10.1016/j.proeng.2017.09.615
- Metcalf, E., and Eddy, H. (2003). *Wastewater Engineering: Treatment, Disposal and Reuse, 4th Edn.* New York, NY: McGraw-Hill.
- Mokhtari, A., and Frey, H. C. (2005). Sensitivity analysis of a two-dimensional probabilistic risk assessment model using analysis of variance. *Risk Anal.* 25, 1511–1529. doi: 10.1111/j.1539-6924.2005.00679.x
- Moter, A., and Göbel, U. B. (2000). Fluorescence *in situ* hybridization (FISH) for direct visualization of microorganisms. *J. Microbiol. Methods* 41, 85–112. doi: 10.1016/S0167-7012(00)00152-4
- Murnleitner, E., Kuba, T., Van Loosdrecht, M., Heijnen, Joseph., et al. (1997). An integrated metabolic model for the aerobic and denitrifying biological phosphorus removal. *Biotechnol. Bioeng.* 54, 434–450.
- Oehmen, A., Carvalho, G., Lopez-Vazquez, C., Van Loosdrecht, M., and Reis, M. (2010a). Incorporating microbial ecology into the metabolic modelling of polyphosphate accumulating organisms and glycogen accumulating organisms. *Water Res.* 44, 4992–5004. doi: 10.1016/j.watres.2010.06.071
- Oehmen, A., Lemos, P. C., Carvalho, G., Yuan, Z., Keller, J., Blackall, L. L., et al. (2007). Advances in enhanced biological phosphorus removal: from micro to macro scale. *Water Res.* 41, 2271–2300. doi: 10.1016/j.watres.2007.02.030
- Oehmen, A., Lopez-Vazquez, C., Carvalho, G., Reis, M., and Van Loosdrecht, M. (2010b). Modelling the population dynamics and metabolic diversity of organisms relevant in anaerobic/anoxic/aerobic enhanced biological phosphorus removal processes. *Water Res.* 44, 4473–4486. doi: 10.1016/j.watres.2010.06.017
- Oehmen, A., Saunders, A. M., Vives, M. T., Yuan, Z., and Keller, J. (2006). Competition between polyphosphate and glycogen accumulating organisms in enhanced biological phosphorus removal systems with acetate and propionate as carbon sources. *J. Biotechnol.* 123, 22–32. doi: 10.1016/j.jbiotec.2005.10.009
- Oehmen, A., Vives, M. T., Lu, H., Yuan, Z., and Keller, J. (2005a). The effect of pH on the competition between polyphosphate-accumulating organisms and glycogen-accumulating organisms. *Water Res.* 39, 3727–3737. doi: 10.1016/j.watres.2005.06.031
- Oehmen, A., Zeng, R. J., Yuan, Z., and Keller, J. (2005b). Anaerobic metabolism of propionate by polyphosphate-accumulating organisms in enhanced biological phosphorus removal systems. *Biotechnol. Bioeng.* 91, 43–53. doi: 10.1002/bit.20480
- Ong, Y. H., Chua, A. S. M., Fukushima, T., Ngho, G. C., Shoji, T., and Michinaka, A. (2014). High-temperature EBPR process: the performance, analysis of PAOs and GAOs and the fine-scale population study of *Candidatus* “accumulibacter phosphatis”. *Water Res.* 64, 102–112. doi: 10.1016/j.watres.2014.06.038
- Pianosi, F., and Wagener, T. (2015). A simple and efficient method for global sensitivity analysis based on cumulative distribution functions. *Environ. Model. Softw.* 67, 1–11. doi: 10.1016/j.envsoft.2015.01.004
- Pijuan, M., Baeza, J. A., Casas, C., and Lafuente, J. (2004). Response of an EBPR population developed in an SBR with propionate to different carbon sources. *Water Sci. Technol.* 50, 131–138. doi: 10.2166/wst.2004.0627
- Plischke, E., Borgonovo, E., and Smith, C. L. (2013). Global sensitivity measures from given data. *Eur. J. Operat. Res.* 226, 536–550. doi: 10.1016/j.ejor.2012.11.047
- Saltelli, A., and Annoni, P. (2010). How to avoid a perfunctory sensitivity analysis. *Environ. Model. Softw.* 25, 1508–1517. doi: 10.1016/j.envsoft.2010.04.012
- Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., and Tarantola, S. (2010). Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Comput. Phys. Commun.* 181, 259–270. doi: 10.1016/j.cpc.2009.09.018
- Saltelli, A., and Bolado, R. (1998). An alternative way to compute Fourier amplitude sensitivity test (FAST). *Comput. Stat. Data Anal.* 26, 445–460. doi: 10.1016/S0167-9473(97)00043-1
- Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., et al. (2008). *Global Sensitivity Analysis: The Primer*. Chichester: John Wiley & Sons.
- Saltelli, A., Ratto, M., Tarantola, S., Campolongo, Francesca., et al. (2006). Sensitivity analysis practices: strategies for model-based inference. *Reliabil. Eng. Syst. Safety* 91, 1109–1125. doi: 10.1016/j.res.2005.11.014
- Shen, N., Chen, Y., and Zhou, Y. (2017). Multi-cycle operation of enhanced biological phosphorus removal (EBPR) with different carbon sources under high temperature. *Water Res.* 114, 308–315. doi: 10.1016/j.watres.2017.02.051
- Sin, G., Gernaey, K. V., and Lantz, A. E. (2009). Good modeling practice for PAT applications: propagation of input uncertainty and sensitivity analysis. *Biotechnol. Prog.* 25, 1043–1053. doi: 10.1002/btpr.166
- Sin, G., Van Hulle, S. W., De Pauw, D. J., Van Griensven, A., and Vanrolleghem, P. A. (2005). A critical comparison of systematic calibration protocols for activated sludge models: A SWOT analysis. *Water Res.* 39, 2459–2474. doi: 10.1016/j.watres.2005.05.006
- Smolders, G., Van der Meij, J., Van Loosdrecht, M., and Heijnen, J. (1994). Model of the anaerobic metabolism of the biological phosphorus removal process: stoichiometry and pH influence. *Biotechnol. Bioeng.* 43, 461–470. doi: 10.1002/bit.260430605
- Smolders, G., Van der Meij, J., Van Loosdrecht, M., and Heijnen, J. (1995). A structured metabolic model for anaerobic and aerobic stoichiometry and kinetics of the biological phosphorus removal process. *Biotechnol. Bioeng.* 47, 277–287. doi: 10.1002/bit.260470302
- Sobol, I. M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math. Comput. Simulat.* 55, 271–280. doi: 10.1016/S0378-4754(00)00270-6
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B (Methodol.)* 58, 267–288. doi: 10.1111/j.2517-6161.1996.tb02080.x
- van Loosdrecht, M. C., Nielsen, P. H., Lopez-Vazquez, C. M., and Brdjanovic, D. (2016). *Experimental Methods in Wastewater Treatment*. London: IWA Publishing.
- Van Veldhuizen, H., van Loosdrecht, M. C., and Heijnen, J. (1999). Modelling biological phosphorus and nitrogen removal in a full scale activated sludge process. *Water Res.* 33, 3459–3468. doi: 10.1016/S0043-1354(99)00064-0

- Vargas, M., Yuan, Z., and Pijuan, M. (2013). Effect of long-term starvation conditions on polyphosphate- and glycogen-accumulating organisms. *Bioresour. Technol.* 127, 126–131. doi: 10.1016/j.biortech.2012.09.117
- Whang, L.-M., and Park, J. K. (2006). Competition between polyphosphate- and glycogen-accumulating organisms in enhanced-biological-phosphorus-removal systems: effect of temperature and sludge age. *Water Environ. Res.* 78, 4–11. doi: 10.2175/106143005X84459
- Yagci, N., Insel, G., Artan, N., and Orhon, D. (2004). Modelling and calibration of phosphate and glycogen accumulating organism competition for acetate uptake in a sequencing batch reactor. *Water Sci. Technol.* 50:241. doi: 10.2166/wst.2004.0382
- Zeng, R. J., Saunders, A. M., Yuan, Z., Blackall, L. L., and Keller, J. (2003a). Identification and comparison of aerobic and denitrifying polyphosphate-accumulating organisms. *Biotechnol. Bioeng.* 83, 140–148. doi: 10.1002/bit.10652
- Zeng, R. J., van Loosdrecht, M., Yuan, Z., and Keller, J. (2003b). Metabolic model for glycogen-accumulating organisms in anaerobic/aerobic activated sludge systems. *Biotechnol. Bioeng.* 81, 92–105. doi: 10.1002/bit.10455

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Nguyen Quang, Rogers, Hofman and Lanham. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

NOMENCLATURE

Symbol	Meaning	Units
EBPR	enhanced biological phosphorus removal	
AS	activated sludge	
ASM	activated sludge model	
SRC	standard regression coefficients	
OLS	ordinary least squares	
LASSO	least absolute shrinkage and selection operator	
RBD-FAST	Fourier amplitude sensitivity test via random balanced designs	
GSA	global sensitivity analysis	
PHA	poly- β -hydroxy-alkanoate	
PHB	poly- β -hydroxy-butyrate	
PHV	poly- β -hydroxy-valerate	
PH ₂ MV	poly- β -hydroxy-methylvalerate	
PAO	phosphorus accumulating organism	
GAO	glycogen accumulating organism	
poly-P	poly-phosphate	
VFA	volatile fatty acids	
HAc	acetate	
HPr	propionate	
SRT	solids retention time	
HRT	hydraulic retention time	
T	temperature	°C
X_j	intra-cellular concentration of component j	C-mol/l
$X_{j,i}$	initial intra-cellular concentration of component j	P or C-mol/l
f_j	intra-cellular fraction of component j per unit of biomass	P or C-mol/C-mol
f_{ji}	initial intra-cellular fraction of component j per unit of biomass	P or C-mol/C-mol
$f_{j,max}$	maximum intra-cellular fraction of component j per unit of biomass	P or C-mol/C-mol
S_j	concentration of j in the bulk liquid phase	P or C-mol/l
$S_{VFA,i}$	initial concentration of j in the bulk liquid phase	C-mol/l
$r_{P/C,i}$	ratio of the PO ₄ to VFA concentration in the influent	
$r_{HAc/HPr,i}$	ratio of acetate to propionate in the influent VFA	
q_j	maximum rate of conversion of component j	P or C-mol/C-molh ⁻¹
$m_{ATP,an}$	anaerobic maintenance coefficient	P-mol/C-molh ⁻¹
$m_{ATP,ox}$	aerobic maintenance coefficient	ATP-mol/C-molh ⁻¹
K_j	half-saturation constant of component j	C-mol/l
θ_j	Arrhenius temperature coefficient of component j	
δ	yield of ATP per unit of NADH oxidized (P/O ratio)	ATP-mol/NADH-mol
ε	aerobic PO ₄ transport coefficient	P-mol/NADH-mol
K_1	ATP required for biomass synthesis from Acetyl-CoA*	ATP-mol/C-mol
K_2	ATP required for biomass synthesis from Propionyl-CoA*	ATP-mol/C-mol