Tanks in series versus compartmental model configuration: considering hydrodynamics helps in parameter estimation for an N₂O model

Giacomo Bellandi, Chaïm De Mulder, Stijn Van Hoey, Usnam Rehman, Youri Amerlinck, Lisha Guo, Peter A. Vanrolleghem, Stefan Weijers, Riccardo Gori and Ingmar Nopens

ABSTRACT

The choice of the spatial submodel of a water resource recovery facility (WRRF) model should be one of the primary concerns in WRRF modelling. However, currently used mechanistic models are limited by an over-simplified representation of local conditions. This is illustrated by the general difficulties in calibrating the latest N₂O models and the large variability in parameter values reported in the literature. The use of compartmental model (CM) developed on the basis of accurate hydrodynamic studies using computational fluid dynamics (CFD) can take into account local conditions and recirculation patterns in the activated sludge tanks that are important with respect to the modelling objective. The conventional tanks in series (TIS) configuration does not allow this. The aim of the present work is to compare the capabilities of two model layouts (CM and TIS) in defining a realistic domain of parameter values representing the same full-scale plant. A model performance evaluation method is proposed to identify the good operational domain of each parameter in the two layouts. Already when evaluating for steady state, the CM was found to provide better defined parameter ranges than TIS. Dynamic simulations further confirmed the CM’s capability to work in a more realistic parameter domain, avoiding unnecessary calibration to compensate for flaws in the spatial submodel.

Key words | ASMG2d, greenhouse gas emissions, mixing model, model layout, parameter domain

Giacomo Bellandi (corresponding author) Department of Civil and Environmental Engineering, Environmental Section, Polytechnic of Milan, Piazza L. da Vinci, 32, 20133 Milan, Italy E-mail: giacomo.bellandi@polimi.it

Chaïm De Mulder Youri Amerlinck Ingmar Nopens BIOMATH, Department of Mathematical Modelling, Statistics and Bioinformatics, Ghent University, Coupure Links 653, B-9000 Gent, Belgium

Stijn Van Hoey INBO, Herman Terlincxgebouw, Havenlaan 88 bus 73, 1000 Brussel (Anderlecht), Belgium

Usnam Rehman AM-TEAM, Hulstbaan 63, 9100 Sint Niklaas, Belgium

Lisha Guo Ryerson University, 350 Victoria St, Toronto M5B 2K3, ON, Canada and Trojan Technologies, 3020 Gore Road, London N5 V 4T7, ON, Canada

Lisha Guo Peter A. Vanrolleghem modelEAU, Université Laval, 1065, avenue de la Médecine, Québec G1 V 0A6, QC, Canada

Stefan Weijers Waterschap De Dommel, Blosscherweg 56, 5283 WB Boxtel, The Netherlands

Riccardo Gori Department of Civil and Environmental Engineering, University of Florence, via di S. Marta 3, 50139 Florence, Italy

doi: 10.2166/wst.2019.024
INTRODUCTION

Nitrous oxide (N$_2$O) has a GWP 265–298 times that of CO$_2$ for a 100-year timescale (IPCC 2013). Its emissions are also of great concern for water resource recovery facilities (WRRFs) (inter alia: Ravishankara et al. 2009) and modelling tools have been largely used in order to understand its production and define possible reduction strategies. The heterotrophic denitrification pathway model of Hiatt & Grady (2008) is currently the only generally accepted model. However, the pathways responsible for N$_2$O production are different and contribute to different extents to the emission depending on wastewater characteristics, plant dynamics and environmental conditions (inter alia: Daelman et al. 2015). Especially in full-scale applications, modelling is a fundamental tool for understanding N$_2$O production and emission dynamics. Mechanistic models have been applied to define general operational recommendations aimed at N$_2$O reduction (Ni & Yuan 2015) but still case-specific recommendations are necessary and more in depth process understanding is needed for an effective minimization of emissions.

A number of kinetic N$_2$O models describing very detailed biological processes have recently been developed (inter alia: Ni & Yuan 2015). In particular, models describing both ammonia oxidizing bacteria (AOB) pathways (i.e. AOB denitrification and incomplete hydroxylamine oxidation) have shown important advances in unfolding the contribution to N$_2$O production of the different consortia in laboratory controlled conditions (inter alia: Spérandio et al. 2016). These highly descriptive mechanistic models have been calibrated and validated in laboratory controlled conditions. However, despite the suggestion of Ni et al. (2013b) for using the dual pathway AOB models, Ni et al. (2013a) discouraged this implementation due to the risk of over-parametrization of the model and the possible creation of strong parameter correlations. In addition to this, the application of both dual pathway and single pathway models in full-scale is still troublesome due to recognized difficulties in identifying proper parameter sets (Ni et al. 2013b; Spérandio et al. 2016). Spérandio et al. (2016) observed high variability of different parameters, among the different case studies and the different models applied, with related high influence on N$_2$O and NO emission results (e.g. the AOB reduction factor set to a high value making the half saturation index for free nitrous acid ($K_{FNA}$) poorly identifiable). These large variations of parameters from one model to another are likely the result of concurring reasons, e.g. micro-organisms’ adaptation to local environmental and process conditions, defaults in the structure of the models, undescribed local heterogeneities in reactors (Spérandio et al. 2016).

The large variations of parameter values among different full-scale case studies considerably limit the predictive power of the models, as parameters cannot be extrapolated to other plants, and probably not even for different periods in the same plant. This also reduces the effective search for greenhouse gas emission mitigation strategies. Given the detailed mechanistic structure of available N$_2$O models, the considerable differences in parameters values among different (full-scale) applications can be explained by an unrealistic representation of local conditions in activated sludge (AS) tanks, to which these models are much more sensitive than traditional activated sludge model (ASM) processes.

The design of proper water resource recovery facility (WRRF) layouts (with respect to their spatial submodel) is an important step in plant-wide modelling and for understanding complex process dynamics (e.g. N$_2$O production) (Rehman et al. 2014). In current tanks in series (TIS) configurations, local recirculations and concentration gradients are assumed to be negligible, and the use of plug-flow/completely-mixed tank configurations is preferred to reduce overall model complexity and computational demand. To date, it is necessary to analyze the possibility and effect of the inclusion of more detailed descriptions of local concentrations in AS tanks by means of more detailed spatial submodels. The development of layouts designed to more accurately describe the hydrodynamic behavior of the internal volume can result in improved predictive power of available mechanistic models, which is key in optimization and control. Currently, the use of compartmental models (CMs) developed upon detailed computational fluid dynamics (CFD) studies is gaining interest from the modelling community (Le Moullec et al. 2010; Rehman et al. 2017, 2015, 2014).

In this work, a comparison of the performance of a CM and a TIS spatial submodel of the same full-scale WRRF on identifying a domain of good parameters values for the most influential parameters using the ASMG2d model (Guo 2014; Guo & Vanrolleghem 2014) is provided. Based on literature, each model parameter was sampled in a specific range for generating a number of simulation scenarios, each with a different parameter set. Each simulation scenario was...
ranked for its performance in predicting measured variables based on different criteria suggested by Van Hoey (2016). The latter returns the good performing scenarios in the form of a distribution of parameter values for both the CM and TIS.

MATERIALS AND METHODS

Model layouts

Two model layouts of the WRRF of Eindhoven were used, differing in terms of spatial submodel (Figure 1). The TIS layout of the Eindhoven WRRF (Figure 1, top) is a well consolidated model obtained after years of research of the facility (De Keyser et al. 2014; Amerlinck 2015). On the other hand, the CM version (Figure 1, bottom) is a recent development of the WRRF model layout resulting from a thorough hydrodynamic study based on CFD simulations in a three-phase (i.e. gas, solid, liquid) model integrated with an ASM to describe the biological activity (Rehman 2016). In particular, the volumes in which the biological tank was initially divided for the case of the TIS, were further partitioned by means of the cumulative species distribution concept that led to the development of the compartmental network currently in use.

For comparing the two model layouts, a common mechanistic reaction model was chosen with which comparison of the results was performed. Given the important efforts deployed to calibrate the ASMG1 and ASMG2d models on the same plant, the biokinetic model chosen for this work was the ASMG2d (Guo 2014).

For the comparison of the two model layouts, three fundamental steps were followed: (I) parameter selection and sensitivity ranking; (II) steady state simulation of n-sampled parameter sets to confirm or redefine current parameter ranges; (III) dynamic simulations of n-sampled parameter sets to evaluate whether CM can better define the parameter domain than TIS. Throughout steady state and dynamic simulations, 12 different model fit metrics (see below) were assessed to evaluate the quality of the model output (see Supplementary Material (SM), available with the online version of this paper).

Parameter selection and sensitivity ranking (Step I)

A literature selection of the most influencing parameters for N₂O production contained in ASMG2d was performed. Screening the literature, a first set of 25 most uncertain parameters was selected (Gernaey & Jørgensen 2004; Hiatt 2006; Van Hulle et al. 2012; Mampaey et al. 2015; Ni et al. 2016; Guo 2014; Spérandio et al. 2016) and is reported in Table 1. Some of the parameters show up to 140% deviation in different calibration exercises (Spérandio et al. 2016).

In order to ensure a sampling of the entire domain without excluding the maximum and minimum limits of each parameter, the domains reported in Table 1 were enlarged by 10% of the difference between the relative maximum and minimum values.

A global sensitivity analysis (GSA) was performed on this set of parameters using the Latin Hypercube – one
sample at time (LH-OAT) approach (van Griensven et al. 2006) with different perturbation factors. As the choice of the perturbation factor can have an important effect on the numerical stability and thus on the sensitivity results, different magnitudes were investigated, i.e. from $10^{-2}$ to $10^{-6}$ (De Pauw & Vanrolleghem 2006). Also, the impact of the number of samples was observed in order to check whether the increase of one or two orders of magnitude impacted the final ranking. These tests resulted in consistent ranking of the outputs, with the only exception being the tests with the perturbation factors smaller than $10^{-5}$, which resulted in numerical instabilities.

**Simulations process**

By means of a LH-OAT sampling approach on the most influential parameters resulting from Step I, the scenarios for the analysis in Step II and III were created. For Step I 10,000 simulations were run given the suggested minimum sample size in the parameter space (Van Hoey 2016), while in Steps II and III, each parameter was uniformly sampled in 2,000 points of its domain.

**Step II**

Steady state simulations were used to compare the model output concentrations with known normal operation conditions in the biological tank. Steady state simulations of 100 days were run, using as input data averaged measurements collected over two months of well-known good plant operation in dry conditions in the summer of 2012 with 1 min frequency. The output of the simulations was compared against average typical concentrations of ammonia (NH$_4$), dissolved oxygen (DO) and total suspended solids (TSS) at the end of the summer package aeration compartment (1.01 mg NH$_4$N/L, 1.02 mgO$_2$/L and 3,200 gTSS/m$^3$, respectively). This allowed making a first ranking of the scenarios based on the proximity of the model outputs and the known measured values of NH$_4$, DO and TSS. As a result, this allowed evaluation of the domain of each parameter considered and eventually provide adjustments by repeating the steady state simulations. This iterative approach allowed definition of a domain for each parameter with ‘good’ parameter values, so that no possibly good parameters values were left out and, at the same time, exclude zones of undoubtedly bad parameter values in order to proceed with Step III. Results of Step II are mostly reported in SM.

**Step III**

Once the last parameter domains after the steady state were defined, the LH-OAT sampling on 2,000 points was repeated for creating the scenarios for the dynamic simulations. Parameters were uniformly sampled from the reduced parameter domain after the steady state analysis. One day of validated influent data with minute frequency was used as input to the model. The outputs were compared with a one minutes frequency dataset of measured of DO, NH$_4$, nitrate (NO$_3$) and liquid N$_2$O (Unisense Environment, Denmark).

**Scenario ranking using 12 different metrics**

Different metrics can be used to score a model fit according to a variety of methods assessing the similarity between a modelled and a measured dataset. The dissimilarity between the different metrics depends not only on their mathematical structure but also on the system behavior and the modelling objective (e.g. average and peak behavior). Hence, there is the need for an assortment of criteria to evaluate the performance of a model from different perspectives. For instance, the root mean square error (RMSE) is a commonly chosen metric to evaluate a model fit; however, it emphasizes the fit of peaks and extreme values. Therefore, its combination with RVE, from the total relative error category, is advisable when variables with a wide range of values are compared (Hauduc et al. 2015).

In this view, for both the steady state and the dynamic simulation step, the outputs were evaluated by means of 12 metrics, i.e. MAE, RMSE, MSE, MSLE, RRMSE, SSE, AMRE, MARE, SARE, MeAPE, MSRE, and RVE (see Table S1, available online). These metrics were selected based on the classification of Hauduc et al. (2015) as the combination of different metrics from different classes have been observed to be more effective than choosing metrics from one class only (Van Hoey 2016). In this way, each metric evaluates the proximity of two time series from a different perspective and the 12 metrics provide a full evaluation of the model performance against measured values. It is to be noted that the full informative potential of the 12 metrics is used in the dynamic simulation step. All metrics were chosen also based on their response range of values, and all indicate the best fit possible at 0. The metrics were selected based on their input requirements so that only values of observed and modelled results must be used as input. In this way, the response value of each metric can be rescaled based on its output from a minimum of 0 (best fit) to a maximum of 1 (worst fit).
Finally, the different scenarios were ranked based on the 0 to 1 value of each metric separately. Subsequently, an overall ranking can be derived based on the score that each scenario has in each of the metrics. In this way, each metric is scalable within its own domain to a 0 to 1 domain, and attaching to each scenario a value from 0 to 1 allows the ranking of the scenarios according to the single metric (Figure 2, left). The value that each scenario collects from each metric can then be summed up with the rest of the scores obtained from the rest of the metrics to obtain a final overall score used for the final ranking of a given scenario (Figure 2, right). The scenarios performing the best for all metrics, i.e. scoring nearly 0 for each different metric, result in the lowest overall score. The best one third of all the scenarios was selected as the good scenarios.

For the steady state case, the average output of the last part of the 100 days simulation (about 100 data points), were compared against the measured value. Therefore, for the evaluation of the steady state simulation outputs, the metric evaluation is only based on the proximity of two single values, i.e. the modelled mean and the relative reference value for NH₄, DO and TSS.

For the case of dynamic simulations, the model outputs of NH₄, NO₃, N₂O, and DO were compared against measured values. In this case, the metric evaluation becomes more complex due to the different nature of the metrics involved. Each metric will return an estimation of the performance of the model output giving more emphasis to different aspects of a model fit. Hence, the necessity of using the proposed ranking strategy summarizing the different aspects of the evaluation of a fit.

### Table 1 | Initial parameter selection showing extreme values of the domain used in literature

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Minimum value</th>
<th>Maximum value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kₒₐ₋ₐ₁ₙₐₖ</td>
<td>Saturation/inhibition coefficient for O₂ in lysis, AOB</td>
<td>0.2</td>
<td>1.6</td>
</tr>
<tr>
<td>Kₒₐ₋ₐ₂ₙₐₖ</td>
<td>Saturation/inhibition coefficient for O₂ in lysis, NOB</td>
<td>0.2</td>
<td>0.69</td>
</tr>
<tr>
<td>bₐ₁</td>
<td>Rate constant for lysis of X_BA1</td>
<td>0.028</td>
<td>0.28</td>
</tr>
<tr>
<td>bₐ₂</td>
<td>Rate constant for lysis of X_BA2</td>
<td>0.028</td>
<td>0.28</td>
</tr>
<tr>
<td>nₙₒₓₐ₋ₐ₁ₙ</td>
<td>Anoxic reduction factor for decay, AOB</td>
<td>0.006</td>
<td>0.72</td>
</tr>
<tr>
<td>Kₑₐ</td>
<td>Half-saturation index for Free Ammonia (FA)</td>
<td>0.001</td>
<td>0.005</td>
</tr>
<tr>
<td>Kₑₐₙ</td>
<td>Half-saturation index for Free Nitrous Acid (FNA)</td>
<td>5.00E-07</td>
<td>5.00E-06</td>
</tr>
<tr>
<td>Kₛₒₐ</td>
<td>FA inhibition coefficient, NO₂ oxidation by NOB</td>
<td>0.5</td>
<td>1</td>
</tr>
<tr>
<td>Kₛₒₐₐ</td>
<td>FNA inhibition coefficient, NO₂ oxidation by NOB</td>
<td>0.036</td>
<td>0.1</td>
</tr>
<tr>
<td>Kₖₙ</td>
<td>FA inhibition coefficient, NH₄ oxidation by AOB</td>
<td>0.1</td>
<td>1</td>
</tr>
<tr>
<td>Kₛₙₙ</td>
<td>FNA inhibition coefficient, NH₄ oxidation by AOB</td>
<td>0.001</td>
<td>0.1</td>
</tr>
<tr>
<td>Kₒₐ</td>
<td>O₂ half-saturation index for AOB</td>
<td>0.4</td>
<td>0.6</td>
</tr>
<tr>
<td>Kₒ₂</td>
<td>O₂ sat coefficient for NOB</td>
<td>1</td>
<td>1.2</td>
</tr>
<tr>
<td>Yₐ₁</td>
<td>Yield for AOB</td>
<td>0.15</td>
<td>0.24</td>
</tr>
<tr>
<td>Yₐ₂</td>
<td>Yield for NOB</td>
<td>0.06</td>
<td>0.24</td>
</tr>
<tr>
<td>Kₑₐₐ₋ₐ₁ₙₐₖ</td>
<td>NH₄ half-saturation index for AOB denitrification</td>
<td>0.001</td>
<td>1</td>
</tr>
<tr>
<td>Kₑₐₙ₋ₐ₁ₙₐₖ</td>
<td>FNA half-saturation index for AOB denitrification</td>
<td>1.00E-06</td>
<td>0.002</td>
</tr>
<tr>
<td>Kᵣₒ₋ₐ₁ₙₐₖ</td>
<td>Inhibition coefficient for O₂ in AOB denitrification</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>Kₛₙ₀₋ₐ₁ₙₐₖ</td>
<td>NO saturation coefficient for AOB denitrification</td>
<td>0.1</td>
<td>3.91</td>
</tr>
<tr>
<td>Kₛ₀₋ₐ₁ₙₐₖ</td>
<td>O₂ sat coefficient for AOB denitrification</td>
<td>0.13</td>
<td>12</td>
</tr>
<tr>
<td>n₁ₐ₀</td>
<td>Growth factor for AOB in denitrification step 1</td>
<td>0.08</td>
<td>0.63</td>
</tr>
<tr>
<td>n₂ₐ₀</td>
<td>Growth factor for AOB in denitrification step 2</td>
<td>0.08</td>
<td>0.63</td>
</tr>
<tr>
<td>Kₐ₁</td>
<td>S_A saturation coefficient for heterotrophs aerobic growth</td>
<td>4</td>
<td>20</td>
</tr>
<tr>
<td>Kₛ₁</td>
<td>S_p saturation coefficient for heterotrophs aerobic growth</td>
<td>4</td>
<td>20</td>
</tr>
<tr>
<td>Kₒ₁₋ₐ₁ₙ</td>
<td>Saturation/inhibition coefficient for heterotroph growth</td>
<td>0.2</td>
<td>1</td>
</tr>
</tbody>
</table>
RESULTS AND DISCUSSION

Parameter ranking (Step I)

For the GSA with LH-OAT approach, different perturbation factors were investigated (De Pauw & Vanrolleghem 2006) resulting in a good performance of a $10^{-3}$ perturbation factor for all the parameters (Figures S1 and S2, available with the online version of this paper).

For the case of the TIS layout, the GSA results provided a ranking of the most influential parameters for N$_2$O, O$_2$, NO$_3$, NH$_4$, TSS, X$_{BA1}$, X$_{BA2}$ and X$_I$. The influential parameters were the same among the different variables tested, showing very comparable overall importance with a few variations in the ranking (Table S2, available online).

The GSA exercise was repeated in the same way for the case of the CM layout. Interestingly, the relevant parameters were very similar to the case of the TIS showing very few variations and negligible differences from the previous ranking exercise.

Initially, 10 parameters (i.e. $b_{A1}$, $b_{A2}$, $K_{OA1}$, $K_{FA}$, $K_{FNA}$, $K_{F1}$, $K_{OA2}$, $K_{O1,BH}$, $Y_{A1}$, $n_{NOx,A1,d}$) were selected according to the score and visual analysis of the tornado plots (Rose 2013). Given the presence of $Y_{A1}$, the proximity in the ranking of $Y_{A2}$, and the attention that this parameter received in literature (inter alia: Pocquet et al. 2015; Spérandio et al. 2016), it was chosen to include also $Y_{A2}$. In a similar way, $K_{OA1,lysis}$ and $K_{OA2,lysis}$ were included given their strong link with $K_{OA1}$, their importance in the literature, and their proximity to the cut-off threshold of the sensitivity results. Finally, given the importance of $K_{FNA}$ in the process of N$_2$O production, it was also included. This selection resulted in a total of 14 parameters to be passed to Steps II and III.

In general, it is interesting that decay parameters for autotrophs are the most influential, and that a significant quantity of half-saturation indexes (K-values) are present in the ranking. This highlights the importance of the correct definition of half-saturation indexes (Arnaldos et al. 2015).

Steady state simulations (Step II)

The aim of Step II was to define the best scenario (i.e. set of parameters) for initializing the model for dynamic simulation (Step III) and to verify that the domain chosen for the different parameters was still valid, i.e. not indicating a need for a modification of the domain.

The output of the simulations, compared against average typical concentrations of NH$_4$, DO and TSS, making use of the selected metrics. This is thus a point-to-point comparison to make a preliminary selection of parameters ranges. Model outputs were scored from 0 (best) to 1 (worst) using the 12 metrics described and ranked accordingly in order to isolate the best performing scenarios.

TIS

Distribution plots of the parameter values relative to the best performing scenarios were used as an indication for possible reduction or modification of the parameter ranges adopted for these simulations before moving to Step III.

The results reflected the general tendency of abating $K_{FNA}$ to very low values (normally in the order of $10^{-6}$) and confirms the reported difficulties in the calibration of this parameter (Spérandio et al. 2016). Similarly, $K_{FA}$ showed a perceivable preference towards lower values in its range.
From the isolation of the best performing scenarios according to the analysis of the modelled TSS, NH₄ and DO confirms the necessity of shifting the parameter range towards bigger values for bA₁.

By merging the three groups of best performing scenarios (i.e. for NH₄, DO, and TSS) it was possible to obtain an overall group of best performing scenarios. This overall group was used as ultimate check to include the parameter domains isolated for NH₄, DO, and TSS, and to help in defining whether the information gathered from the individual cases still holds when considering multiple parameters simultaneously.

**CM**

The visual ranking of the scenarios for the steady state simulations with the CM layout resembles very closely the ranking observed for the TIS layout (SM). This means that the absolute variations of the model output for the different scenarios are similar for the two layouts for NH₄, DO and TSS.

From the NH₄, DO, and TSS rankings individually and overall (merging the three groups of best performing scenarios), a clear tendency of bA₁ to show a higher frequency in the highest part of its domain was noticed (Figure S9, available online). This was a clear indication of the need for a redefined domain for bA₁ before passing to Step III. Similarly, bA₂, KF₂, and KFNA returned a clear preference of the highest frequency of their distribution plot for the lower part of their domain.

**Redefinition of parameter domains**

According to the results of Step II for the TIS and CM layouts, some of the parameters would benefit from a modification of their sampling domain before passing to Step III. For those parameters showing truncated distributions and high frequency of best performing values close to an edge of their domain, the modification was considered. This reduces the number of simulations that likely result in a less good prediction and are not very useful in the analysis anyway. The number of simulations with the CM layout resembles very closely the full-scale measured data. In this view, the scenarios were ranked according to the 12 metrics and compared, as in Step II, in terms of their capability of providing a realistic parameter range of best performing values. Therefore, the ranking used the same method as for Step II, but using online measured data for the metric comparison (i.e. NH₄, DO, N₂O and NO₃).

It must be pointed out that in the case of N₂O it was not possible to use all 12 metrics due to the fact that some metrics use the value at the denominator of a fraction returning an infinite solution if a variable reaches zero. AMRE, MARE, MSLE, MSRE, and SSE were not considered for ranking the scenarios according to the N₂O output.

For the TIS layout, the ranking according to the measured NH₄ (Figure 5) showed an interesting behavior of the MSLE metric which, at first sight, seems to rank the scenarios inversely to the rest of the metrics. This is true for some of the worst performing scenarios for MSLE (darker color), which are not considered as bad by the rest of the metrics. The reason lies in the high sensitivity of the MSLE to small differences between modelled and measured values. In particular, when both measured and modelled variables are smaller than 1, the discrepancy is enhanced by the effect of the logarithm and the quadratic term in the MSLE. Thus, the importance of using multiple metrics is illustrated once more. Using multiple metrics of different nature allows to analyze and rank the scenarios from different points of view, but also to compensate for particular behavior of a single metric. Nonetheless, the visualization proposed in this work highlights the contribution of the single metric and relative potential limits.

Dynamic simulations (Step III)

The model was initialized with a steady state simulation of 100 days for performing the dynamic simulations. A best scenario for initialization was thus needed. Using the intersection of the three groups of best performing scenarios, i.e. the scenario considered the best at the same time for NH₄, DO, and TSS could be identified.

Step III was targeted at defining the best performing scenarios by analyzing the dynamic simulation outputs against measured data in specific locations of the bioreactor. The aim of this phase was to compare the capabilities of the TIS and CM layouts in defining a set of scenarios best resembling the full-scale measured data. In this view, the scenarios were ranked according to the 12 metrics and compared, as in Step II, in terms of their capability of providing a realistic parameter range of best performing values. Therefore, the ranking used the same method as for Step II, but using online measured data for the metric comparison (i.e. NH₄, DO, N₂O and NO₃).

Concerning the ranking according to DO, all metrics resulted in well-behaving simulations (small values for the 12 metrics) and overall agreeing in a common final ranking.

Although the ranking according to N₂O was forced to have fewer metrics (see above), the seven metrics retained
were still coming from different categories, thus ensuring a ranking according to different approaches. All metrics appear to rank similarly, although the fast transition towards the darkest colors suggests the presence of only a few scenarios performing significantly better than the rest.

In a similar picture, the ranking according to NO3 can be observed, where, for most of the metrics, a fast transition to darker colors indicates a fast deviation of the modelled results away from the measured dataset.

Figure 4 shows the ranking for the scenarios of the CM layout. Small differences can be observed among the metrics for the ranking according to NH4 in which MSLE seems to behave slightly different from the rest of the metrics, although generally agreeing with the rest of the metrics for the best performing scenarios (lighter colors).

For the case of DO there is faster transition to the darker tones of the ranking for all metrics, indicating probably that only a few scenarios are providing an output close to the measured dataset while the rest is quickly deviating away from it.

Differently from the case of DO, the case of N2O and NO3 present a very gradual shift away from the objective function making all metrics generally provide the same ranking (note again that for N2O fewer metrics could be considered).

**Comparison between TIS and CM**

The overall distributions of the parameter values for the best performing scenarios are derived from the ranking for NH4, DO, N2O, and NO3, and selecting the best third of all the scenarios. These distributions are reported to make an overall comparison of the performances of both model layouts in defining ranges of parameter values that are best performing.

For the case of Y_{AI} (Figure 5), the CM configuration (right) returned a clearly defined range of acceptable parameter values as compared to the case of the TIS layout. The Y_{AI} histogram of the CM appears to recall a defined distribution curve which encounters a maximum frequency around the value of 0.1 g COD/g N. The TIS model (Figure 5, left) identifies the best performing scenarios in the lowest range of Y_{AI}, which are less realistic values as compared to the case of the CM.

K_{FNA} (Figure 6), is a parameter that is known to be difficult to calibrate, often leading to values very close to zero to force a fit (Spérandio et al. 2016). The CM results (Figure 6,
right) again show a more pronounced shape of a distribution as compared to the TIS. This is an important result as it proposes more realistic values for this parameter and reverts the general tendency of moving this parameter down to 1E-6 g/m³. On the other hand, the TIS layout does not show a definite distribution, having the same frequency almost everywhere. However, it must be pointed out that the frequency on the right edge of the distribution for the TIS is slightly increasing, suggesting the possibility of a need for a modification of the K_{FNA} domain.

In this view, it is interesting to consider that, despite literature studies generally reporting very low values of K_{FNA}, the TIS layout appears to revert this tendency showing this time a propensity for more realistic values. This might be because a rather consolidated TIS configuration is used for this plant, which has been calibrated and validated in detail in multiple past studies on this WRRF. Furthermore, it is interesting to point out how the CM model confirms the same tendency but with a more pronounced shape of the distribution. This is another confirmation that the
higher hydrodynamic accuracy of the CM significantly increases the identifiability of some parameters.

CONCLUSIONS

In the present work, a ranking method and a visualization approach were proposed for selecting the best performing model parameter sets (scenarios) and providing a qualitative indication of the performance and contribution of each of 12 metrics used to express model fit. It is important to remember that this was not a calibration exercise and that model fits are deliberately not shown to emphasize the core messages.

The visual representation of the ranking of the different scenarios in both steady state and dynamic simulations returned interesting clues on the necessity of considering multiple metrics that assess different natures of model fit. Also, the performance of each metric was highlighted in its ranking and in the relative effect on the overall ranking of the scenarios, providing information on the contribution of the single metric and as a whole.

Plotting the parameter values obtained through the selection of the best performing scenarios, it was possible to directly compare the performance of the CM and TIS layouts in terms of parameter identification. The use of the CM increases the level of detail in the representation of local concentrations. The volume containing the sensors (and therefore providing local concentrations) is much better represented in the case of the CM. This improves calibration results significantly. This implies a relevant gain in model accuracy that redefined some of the key parameters to acceptable values and obtained more defined distributions.

The CM generally returned a more narrow parameter domain of good values compared to the TIS layout. This indicates that the more detailed description of local concentrations helps in defining a narrower domain of key parameters, which will, upon calibration, improve the model predictive power.

ACKNOWLEDGEMENTS

Peter A. Vanrolleghem, I. Nopens and L. Guo acknowledge the financial support obtained through the TECC project of the Québec Ministry of Economic Development, Innovation and Exports (MDEIE) and the joint research project funded by the Flemish Fund for Scientific Research (FWO-G.A051.10). Peter Vanrolleghem holds the Canada Research Chair on Water Quality Modeling.

REFERENCES

Amerlinck, Y. 2015 Model Refinements in View of Wastewater Treatment Plant Optimization: Improving the Balance in sub-Model Detail. Ghent University, Gent, Belgium.
Guo, L. 2014 Greenhouse gas Emissions From and Storm Impacts on Wastewater Treatment Plants: Process Modelling and Control. LAVAL University, Quebec City, QC, Canada.
Hiatt, W. C. 2006 Activated Sludge Modeling for Elevated Nitrogen Conditions. Clemson University, Clemson, SC, USA.


First received 7 August 2018; accepted in revised form 7 January 2019. Available online 17 January 2019