Tanks in series versus compartmental model configuration: Considering hydrodynamics helps in parameter estimation for an N_2O model

Giacomo Bellandi[1,2], Chaïm De Mulder[2], Stijn Van Hoey[3], Usnam Rehman[4], Youri Amerlinck[2], Lisha Guo[5,8], Peter A. Vanrolleghem[6], Stefan Weijers[7], Riccardo Gori[1] and Ingmar Nopens[2]

[1] Department of Civil and Environmental Engineering, University of Florence, via di S. Marta 3, 50139 Florence, Italy (*email: Giacomo.Bellandi@dicea.unifi.it, Riccardo.Gori@dicea.unifi.it)*
[2]BIOMATH, Department of Mathematical Modelling, Statistics and Bioinformatics, Ghent University, Coupure Links 653, B-9000 Gent, Belgium (*email: Chaim.DeMulder@UGent.be, Ingmar.Nopens@UGent.be*)
[3]INBO, Kliniekstraat 25, 1070 Brussels (Anderlecht), Belgium (*email: stijn.vanhoey@inbo.be*)
[4]AM-TEAM, Hulstbaan 63, 9100 Sint Niklaas, Belgium (*email*: *Usman.Rehman@AM-TEAM.com*)
[5]Ryerson University, 350 Victoria Street, Toronto M5B 2K3, ON, Canada (*email: lisha.guo@ryerson.ca*)
[6]modelEAU, Université Laval, 1065, avenue de la Médecine Québec G1V 0A6, QC, Canada (*email: Peter.Vanrolleghem@gci.ulaval.ca*)
[7]Waterschap De Dommel, Bosscheweg 56, 5283 WB Boxtel, The Netherlands (*email: SWeijers@dommel.nl*)
[8]Trojan Technologies, 3020 Gore Road, London N5V 4T7, ON, Canada (*email: lguo@trojanuv.com*)

Abstract
The choice of the spatial submodel of a WRRF model should be one of the primary concerns in WRRF modelling. However, currently used mechanistic models are too often limited by a too simplified representation of local conditions. This is illustrated by the general difficulties in calibrating the latest N_2O models and the large variability in parameter values reported in the literature. The use of CM developed on the basis of accurate hydrodynamic studies using CFD can much better take into account local conditions and recirculation patterns in the AS tanks that are important with respect to the modelling objective. The conventional 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 at the steady state phase, the CM was found to provide better defined parameter ranges than TIS. Dynamic simulations further confirmed the CM capability to work in a more realistic parameter domain, avoiding unnecessary calibration to compensate for flaws in the spatial submodel.

Keywords
compartmental model; tanks in series; ASMG2d; parameter domain, N2O, model layout

INTRODUCTION

N_2O emissions are of great concern in WRRFs and modelling tools have been largely used to date in order to understand its production and define possible reduction strategies. The heterotrophic denitrification pathway model from Hiatt and Grady (2008) is currently the only generally accepted model. However, the pathways responsible for N_2O production are different and contributing to different extents to the emission depending on wastewater characteristics, plant dynamics and environmental conditions (Ahn et al., 2010; Daelman et al., 2015). Especially in full-scale applications, modelling is a fundamental tool for understanding N_2O production and emission dynamics. Mechanistic models have been applied to define general operational recommendations aimed at N_2O reduction (Ni and 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_2O models describing very detailed biological processes have recently been developed (Mannina et al., 2016; Ni and Yuan, 2015; Pocquet et al., 2015). In particular, models describing both AOB pathways (i.e. AOB denitrification and incomplete NH_2OH oxidation) have shown important advances in unfolding the contribution to N_2O production of the different consortia in laboratory controlled conditions (Ni et al., 2014; Pocquet et al., 2015; Spérandio et al., 2016). These mechanistic models are highly descriptive of the known biological processes responsible for N_2O production and 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). In particular, 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_2O and NO emission results. In one case, the ηAOB has been set to a high value making K_{FNA} poorly identifiable, while the opposite has been observed for another full-scale application. These large variations of parameters from one system to another are likely the result of concurring reasons e.g. micro-organisms history and adaptation, defaults in the structure of the models, undescribed local heterogeneities in reactor (Spérandio et al., 2016).

The large variations of parameters 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 reduced predictive power will also hamper the usage of such models in search for mitigation strategies. Given the detailed structure of available models with regards to the conversion processes involved, the considerable differences in parameters values among different (full-scale) applications are likely due to an unrealistic representation of local conditions in AS tanks, to which these conversion processes are highly sensitive (much more than the traditional ASM processes).

The design of proper WRRF layouts (with respect to spatial submodel) is an important step in plant-wide modelling and for understanding complex process dynamics such as the ones responsible for N_2O production (Rehman et al., 2014a). In current TIS configurations, recirculation and more detailed local concentrations were assumed to be negligible, and the use of plug-flow-CSTR configurations was preferred to reduce overall model complexity and computational demand. In view of the latest issues in N_2O modelling in WRRFs, it is to date 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 for resembling more accurately hydrodynamic behavior of the internal volume layout, is currently bringing an additional level of detail that can reflect in improved predictive power of available mechanistic models, which is key in optimization and control. Currently, the use of CMs developed upon detailed CFD studies is gaining interest from the modelling community (Le Moullec et al., 2010; Rehman et al., 2017, 2015, 2014b).

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 sensitive parameters using the ASMG2d model (Guo, 2014; Guo and Vanrolleghem, 2014) is provided. Based on literature, each model parameter was sampled in a specific range for generating a number of simulation scenarios. 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 (Amerlinck, 2015; Cierkens et al., 2012; De Keyser et al., 2014). 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 for resembling 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.

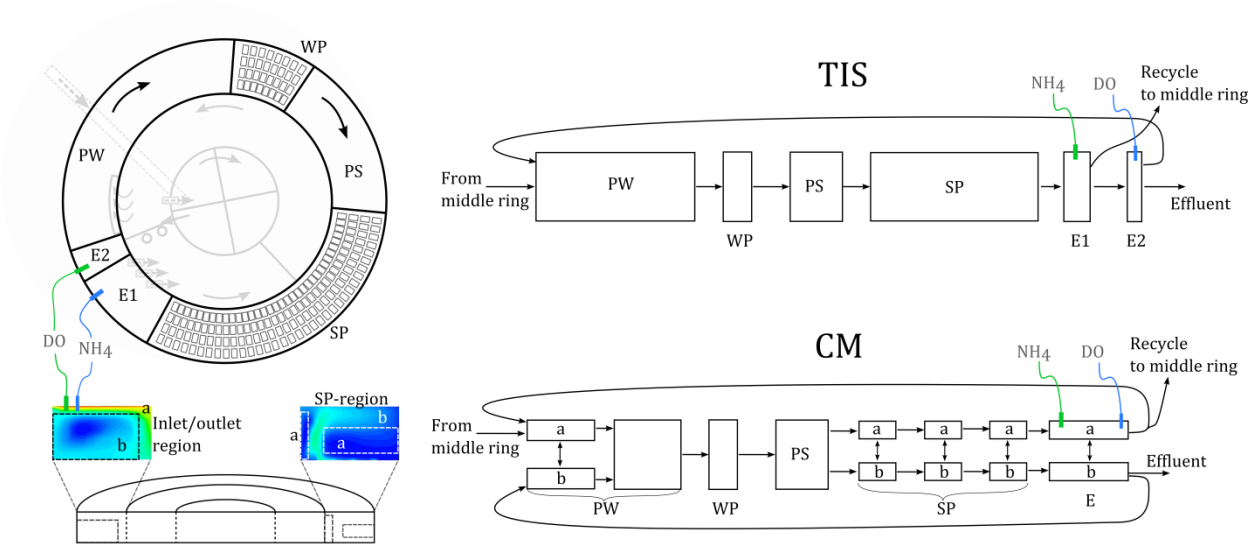


Figure 1 – Schematic representation of the partitioning of the AS tank volume according to the TIS (top) and CM (bottom) layouts. The planar representation of the AS tank (top left) is divided for the TIS (top right) in pre-winter (PW), winter package (WP), pre-summer (PS), summer package (SP), effluent (E1 and E2) zones. The CM follows the same concept of TIS in the general division of the volumes, but includes *a* and *b* recirculation zones according to Rehman (2016).

For comparing the two model layouts, a common mechanistic model was chosen with which comparison of the results was performed. Seen the efforts on calibrating the ASMG1 and ASMG2d on the same plant, the biokinetic model chosen for this work was the ASMG2d (Guo, 2014). This model is one of the most popular in full-scale applications and is also implemented in the WEST® platform. In addition to this, the ASMG2d has been considered in other studies in literature, representing an added value for further comparison of the results (Spérandio et al., 2016). It must be specified that, as other N_2O mechanistic models, the ASMG2d is far from being widely applicable to full-scale WRRFs due to the discussed difficulties that these models show in the calibration step. However, for the purpose of this study and for the application to this plant, the ASMG2d represents the most suitable choice.
As input to both the TIS and CM models, a dataset of validated SCADA data from May 2016 was used during which also N_2O measurements in the liquid (Unisense Environment, Denmark) were available. For the steady state simulations a period of 100 days was simulated and the last 30 days were used for averaging output variables. For the dynamic simulations, a 24h dataset of validated input data was used. In order to compare simulation output with measured values, dissolved N_2O measurements and SCADA data from the sensors present on

the AS tank were used. The output data of the simulations were taken from the (CSTR) model block resembling most closely the location of the relative sensor in the reality.
For the comparison of the two model layouts, three fundamental steps were followed: I) parameter selection and definition of parameters ranges, and ranking; II) steady state simulation of n-sampled parameters sets to confirm or redefine current parameter ranges; III) dynamic simulations of n-sampled parameters sets to evaluate whether CM can better define the parameter domain than TIS. Throughout steady state and dynamic simulations, 12 model fit metrics were assessed to evaluate the quality of the model output.

Parameter selection and sensitivity ranking (Step I)
A literature selection of the most influencing parameters for N_2O production contained in ASMG2d was performed. Screening the literature, a first set of 25 most uncertain parameters was selected (Gernaey and Jørgensen, 2004; Guo, 2014; Hiatt, 2006; Mampaey et al., 2013; Ni et al., 2013b; Spérandio et al., 2016; Van Hulle et al., 2012) and is reported in Table 1. Some of the parameters show up to 140% deviation from different calibration exercises (Spérandio et al., 2016).

Table 1 – Initial parameter selection showing extreme values of the domain used in literature.

Parameter	Description	Minimum value	Maximum value
$K_{O\_A1Lysis}$	Sat/inhibition coefficient for O_2 in lysis, AOB	0.2	1.6
$K_{O\_A2Lysis}$	Sat/inhibition coefficient for O_2 in lysis, NOB	0.2	0.69
b_{A1}	Rate constant for lysis of X_BA1	0.028	0.28
b_{A2}	Rate constant for lysis of X_BA2	0.028	0.28
$n_{NOx\_A1\_d}$	Anoxic reduction factor for decay, AOB	0.006	0.72
K_{FA}	Half-saturation index for Free Ammonia	0.001	0.005
K_{FNA}	Half-saturation index for FNA	5.00E-07	5.00E-06
K_{I10FA}	FA inhibition coefficient, NO_2 oxidation by NOB	0.5	1
K_{I10FNA}	FNA inhibition coefficient, NO_2 oxidation by NOB	0.036	0.1
K_{I9FA}	FA inhibition coefficient, NH_4 oxidation by AOB	0.1	1
K_{I9FNA}	FNA inhibition coefficient, NH_4 oxidation by AOB	0.001	0.1
K_{OA1}	O_2 half-saturation index for AOB	0.4	0.6
K_{OA2}	O_2 half-saturation index for NOB	1	1.2
Y_{A1}	Yield for AOB	0.15	0.24
Y_{A2}	Yield for NOB	0.06	0.24
K_{FA_AOBden}	NH half-saturation for AOB denit	0.001	1

K_{FNA_AOBden}	FNA half-saturation for AOB denit	1.00E-06	0.002
K_{IO_AOBden}	Inhibition coefficient for O_2 in AOB denit	0	10
K_{SNO_AOBden}	NO saturation coefficient for AOB denit	0.1	3.91
K_{SO_AOBden}	O_2 sat coefficient for AOB denit	0.13	12
$n1_{AOB}$	Growth factor for AOB in denitr step 1	0.08	0.63
$n2_{AOB}$	Growth factor for AOB in denitr step 2	0.08	0.63
K_{A1}	SA sat coefficient for heterotrophs aerobic growth	4	20
K_{F1}	SF sat coefficient for heterotrophs aerobic growth	4	20
K_{O1_BH}	Sat/inhibition coefficient for heterotroph growth	0.2	1

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 GSA was performed on this set of parameters using the 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 (De Pauw and 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 of 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 sensitive parameters resulting from Step I, the scenarios for the analysis in Step II and III were created. For each case 2k points on the domain of each parameter were uniformly sampled.

Step II
Steady state simulations were used to compare the model output concentrations with known normal operation conditions in the biological tank. This allowed to make a first ranking of the scenarios based on the proximity of the model output and the known measured values of NH_4, DO and TSS. As a result, this allowed to evaluate the domain of each parameter considered and eventually provide adjustments repeating the steady state simulations. This iterative approach allowed to define a domain for each parameter with “good” parameter values, so that no possibly good parameters values were left out and, at the same time, excluding zones of undoubtedly bad parameter values in order to proceed with Step III.

Step III
Once the last parameter domains after the steady state were defined, the LH-OAT sampling on 2k points was repeated for creating the scenarios for the dynamic simulations. Parameters were uniformly sampled on the eventually reduced domain after the steady state analysis. In this case, the outputs of the model were compared with a day of measured SCADA data (i.e. DO, NO_3, NH_4) and liquid N_2O measurements.

Scenario ranking using 12 different metrics

Different metrics can be used to score a model fit according to a variety of methods to describe the similarity between a modelled and an objective function, therefore, different are the criterion with which scores are assigned. Dissimilarity between metrics depends not only on their mathematical structure but also on the system behavior and objective. Hence, the need of an assortment of criteria to evaluate the performance of a model from different perspectives. For instance, RMSE is a commonly chosen metric to evaluate a model fit, however, it gives emphasis to the fit of peaks and high 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 (Table 2). 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, 2016a). All metrics were chosen also based on their response range of values, all metrics (including RVE) indicate the best fit possible with 0. The metrics were selected based on their input requirements so that only values of observed and modelled results could be used as input. In this way, the response value of each metric chosen, can be rescaled based on its output from a minimum of 0 (best fit) to a maximum of 1 (worst fit).

Table 2 – Summary table of the metrics considered for scenario ranking (Hauduc et al., 2015; Van Hoey, 2016a).

Metric	Category	Output range	Main feature
MAE	Absolute	[0, inf]	Indicates the average magnitude of the model error
RMSE	Absolute	[0, inf]	Emphasizes large errors
MSE	Absolute	[0, inf]	Emphasizes high errors
MSLE	Absolute	[0, inf]	Emphasizes low magnitude errors
RRMSE	Absolute	[0, inf]	Low values suggest good agreement
SSE	Absolute	[0, inf]	Low values suggest good agreement
AMRE	Relative	[0, inf]	Low values suggest good agreement
MARE	Relative	[0, inf]	Low values suggest good agreement
SARE	Relative	[0, inf]	Low values suggest good agreement
MeAPE	Relative	[0, inf]	Less affected by outliers and errors distribution
MSRE	Relative	[0, inf]	Emphasizes larger relative errors
RVE	Total Relative error	[-inf, inf]	Measures an overall adequacy

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 addressing 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*.

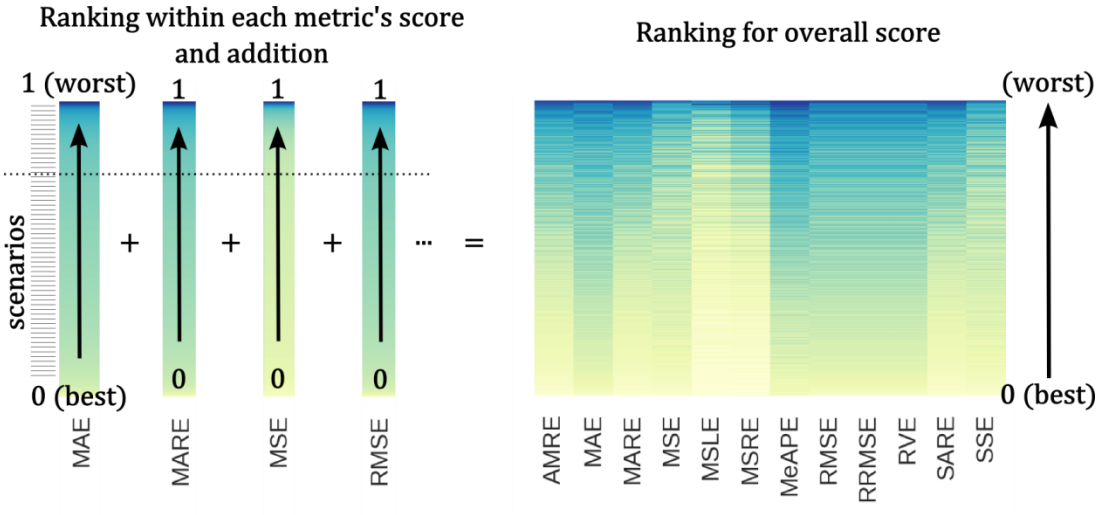


Figure 2 – Schematic representation of the scenario ranking method used in Step II and Step III. The initial ranking according to the single metric (left) allows to sum the scores of all metrics for each scenario and have a final score that is used for the overall ranking (right).

For the steady state case, the average output of the last part of the 100 days simulation (about 100 data points), were compared against an objective 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 NH_4, DO and TSS.

For the case of dynamic simulations, the model output of NH_4, NO_3, N_2O, 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 a ranking strategy summarizing the different aspects of the evaluation of a fit.

RESULTS AND DISCUSSION

Parameter ranking (Step I)

After the selection of the parameters and the definition of the respective range from the literature, a ranking exercise was done. A GSA was performed on this set of parameters using the LH-OAT approach with different perturbation factors. Different magnitudes were investigated (De Pauw and Vanrolleghem, 2006) resulting in a good performance of a perturbation factor of 10^{-3} for all the parameters. The suggested minimum sample size in the parameter space is in the range of 10k samples which makes the experiment highly time demanding (Van Hoey, 2016b).

For the case of the TIS layout, the GSA results provided a ranking of the most influential parameters for N_2O, O_2, NO_3, NH_4, TSS, X_{BA1}, X_{BA2} and X_H. The influential parameters seemed to be the same among the different variables tested overall showing very comparable importance with few variations in the ranking.
The GSA exercise was repeated in the same fashion 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.
In order to define an overall set of parameters suitable for both the TIS and the CM cases, it was decided to give a score to each parameter according to its position in the tornado plot of each variable (i.e. position 1 scores 1, position 2 scores 2, etc.). Initially, 10 parameters were selected according to the score and visual analysis of the tornado plots (i.e. b_{A1}, b_{A2}, K_{OA1}, K_{FA}, K_{FNA}, K_{F1}, K_{OA2}, K_{O1_BH}, Y_{A1}, $n_{NOx\_A1\_d}$). Given the presence of Y_{A1}, the proximity in the ranking of Y_{A2}, and the attention that this parameter received in literature, it was chosen to include also Y_{A2}. In a similar fashion, $K_{O\_A1Lysis}$ and the respective $K_{O\_A2Lysis}$ were included given the importance in the literature and their proximity to the cut-off threshold. Finally, given that K_{I9FA} was not the worst positioned in this new GSA ranking, it was also included. This selection resulted in a total of 14 parameters to be passed to step II and III.
In general, it is interesting that decay parameters for autotrophs are the most sensitive, and that a relevant quantity of half-saturation indexes (K-values) are present in relevant positions of 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 verify that the domain chosen for the different parameters was still valid, i.e. not indicating clear clues of a need for a modification of the domain.
Steady state simulations of 100 days were run. The output of the simulations was compared against average typical concentrations of NH_4, DO and TSS at the end of the summer package aeration compartment (1.01 mg N/L, 1.02 mg/L and 3200 g/m^3 respectively) obtained from averaging measured data of known good plant operation in dry conditions during summer 2012. 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. Each metric returns an internal ranking according to the score given to each scenario. An overall ranking among the scenarios is possible summing up the contribution of all metrics for each scenario.

TIS
The steady state simulations performed with the TIS model were ranked for the average output of NH_4, DO and TSS against an objective value. The ranking strategy used places the scenario with the lowest score overall (best performing) at the bottom of the graph (near value 0) in Figure 3, while the worst performing are ranked towards the top (near value 1). In particular, Figure 3 indicates that, for the scenario tested, the variation of the output TSS (Figure 3, right) as compared to the objective value is more pronounced than for DO or NH_4 (Figure 3, center and left respectively), the latter showing the smallest variations. Therefore, TSS seems particularly sensitive to variations in the selected parameters values as compared to DO and at last to NH_4, i.e. a deviation in color to the darker tones is visible already close to the bottom of Figure 3 (right). It must be also pointed out that, among the scenarios, the magnitude of variation between NH_4, DO and TSS values is largely different due to the different units. Therefore, the transition in color must be considered only as an indication of how fast the outputs of the different scenarios go far from the objective values and what is the contribution of each metric chosen. As an example, from the ranking graph relative to NH_4 (Figure 3, left), given the small absolute differences between the model output and the

objective value, it can be noticed that the most sensitive metric is MSLE. MSLE is very sensitive to smaller differences as compared to the rest of the metrics due to the fact that it treats both modelled and objective values with a logarithm, thus emphasizing small differences and particularly values smaller than 1. On the other hand, the rest of the metrics need bigger absolute differences between modelled and objective values. Hence, the need of a variety of metrics evaluating a model fit from different points of view. In particular, for the case of NH_4, MSLE provides most of the input for the final ranking as the rest of the metrics are showing very little variation.

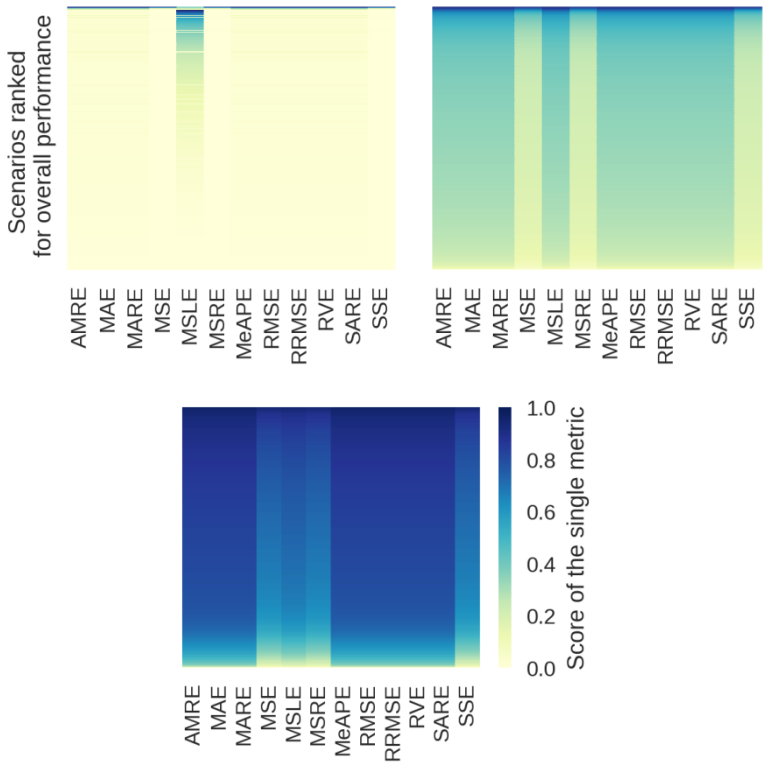


Figure 3 – Ranking of the scenarios (rows) according to the 12 metrics (columns) from the best performing (bottom) to the worst (top) (for NH_4, DO and TSS respectively from left to right). Each metric is colored according to its internal ranking from 0 (bright) to 1 (dark).

At this point the best performing scenarios (Figure 3, light colors) were selected from each of the cases, i.e. NH_4, DO and TSS. Here, only the most relevant results are summarized and discussed.

Distribution plots of the parameter values relative to the best performing scenarios, selected according to NH_4 ranking, showed that b_{A1} and K_{FA} appear to perform the best in the higher range of the respective domains (Figure 4). The rest of the parameters did not return a particular shape suggesting that there is not a preferred subrange in the tested range. This is an indication for possible reduction or modification to the conservative parameter ranges adopted for these simulations before moving to Step III.

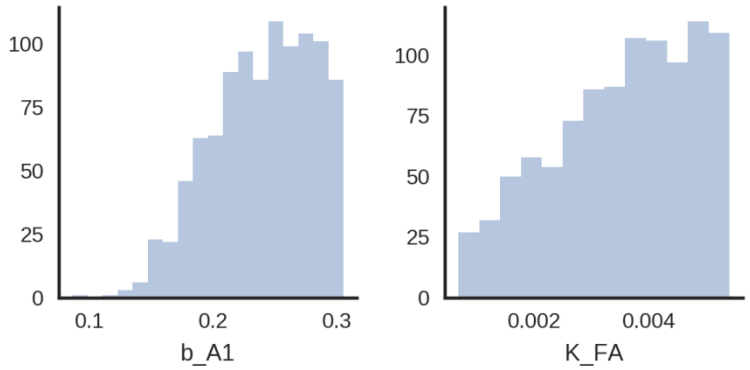


Figure 4 – Distribution of the parameter values resulting from the selection of the best performing scenarios for NH_4 in the TIS layout.

For the case of DO there is a confirmation of the good performance returning from the use of the higher range of b_{A1}, once again suggesting that for Step III a reconsideration of the sampling range for this parameter is useful (Figure 5, upper left). Interestingly, the range of b_{A1} is also corroborating with the one observed for NH_4. Also b_{A2} shows a defined tendency in its distribution, showing a relevant preference for values in the lowest range of its domain (Figure 5, upper right). In addition to this, K_{F1} and K_{FNA} show a higher density of good performing scenarios close to zero (Figure 5, bottom left and right graphs). This reflects 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} shows a perceivable preference towards lower values in its range, although less pronounced than for the previous cases (Figure 5, bottom center) and with opposite tendency.

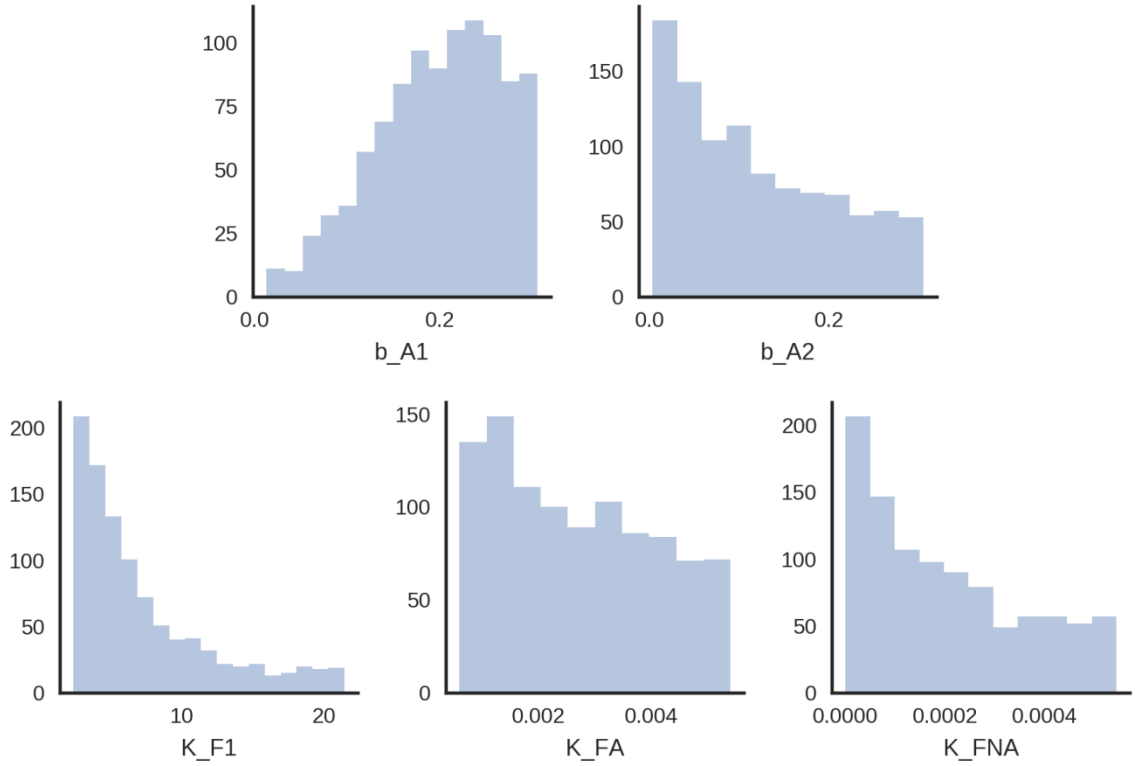


Figure 5 – Distribution of the parameter values resulting from the selection of the best performing scenarios for DO in the TIS layout.

From the isolation of the best performing scenarios according to the analysis of the modeled TSS, in contrast to the previous results, parameter b_{A1} shows the highest frequency peak in the

lowest range of its domain (data not shown). However, this distribution appears to be approaching a bimodal case as a noticeable peak in frequency is also visible in the highest part of the b_{A1} domain. This is particularly interesting as the peak on the right occurs very similar to the cases observed for NH_4 and DO. This confirms the necessity of shifting the parameter range towards bigger values for b_{A1} for the model to both comply for DO and NH_4.
By merging the three groups of best performing scenarios (i.e. for NH_4, DO, and TSS) it was possible to obtain an overall group containing all selected best performing scenarios. This overall group was used as ultimate check as including the parameter domains isolated for NH_4, DO, and TSS, helps in defining whether the information gathered from the singular 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 one observed for the case of the TIS layout (Figure 3). This means that the absolute variation of the model output for the different scenarios from the objective value, are similar for the two layouts for NH_4, DO and TSS.
Similarly to what was observed in the results of the TIS layout, distribution plots of the parameter values relative to the best performing scenarios for NH_4 ranking, showed that b_{A1} and K_{FA} perform the best in the higher range of their domains (Figure 6). This is an important aspect in view of identifying which parameter's domain needs adjustment before passing to the dynamic simulations (Step III).

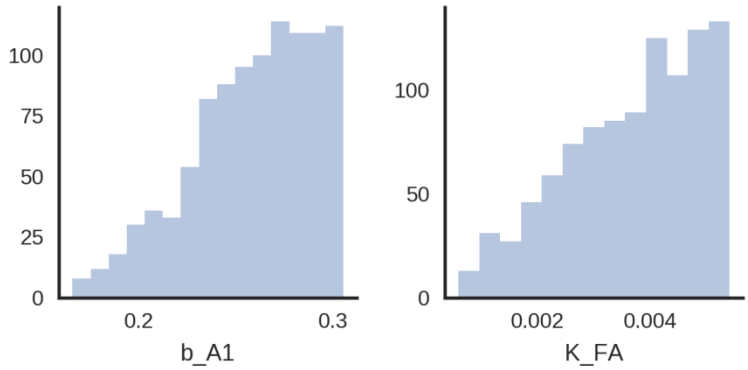


Figure 6 – Distribution of the parameter values resulting from the selection of the best performing scenarios for NH_4 in the CM layout.

Plotting the best performing scenarios relative to DO ranking (Figure 7), there is again full agreement with what already observed in the case of the TIS layout. The frequency of the best performing scenarios is highest in the higher range of b_{A1}'s domain, while b_{A2}, K_{F1} and K_{FNA}, show a clear preference for their lowest limit. A similar pattern can be observed for K_{FNA} (Figure 7, bottom center) although with less definite shape than for the other parameters, and very similar to what observed in the TIS results. In the same way, K_{FA} shows a perceivable preference towards lower values in its range, although opposite and less definite than was observed for NH_4, matching the results of the TIS layout.
Again these are important clues for the modification of the domain of certain parameters before passing to Step III. At the moment all results between TIS and CM seem to corroborate rather closely and no clear difference can be noticed.

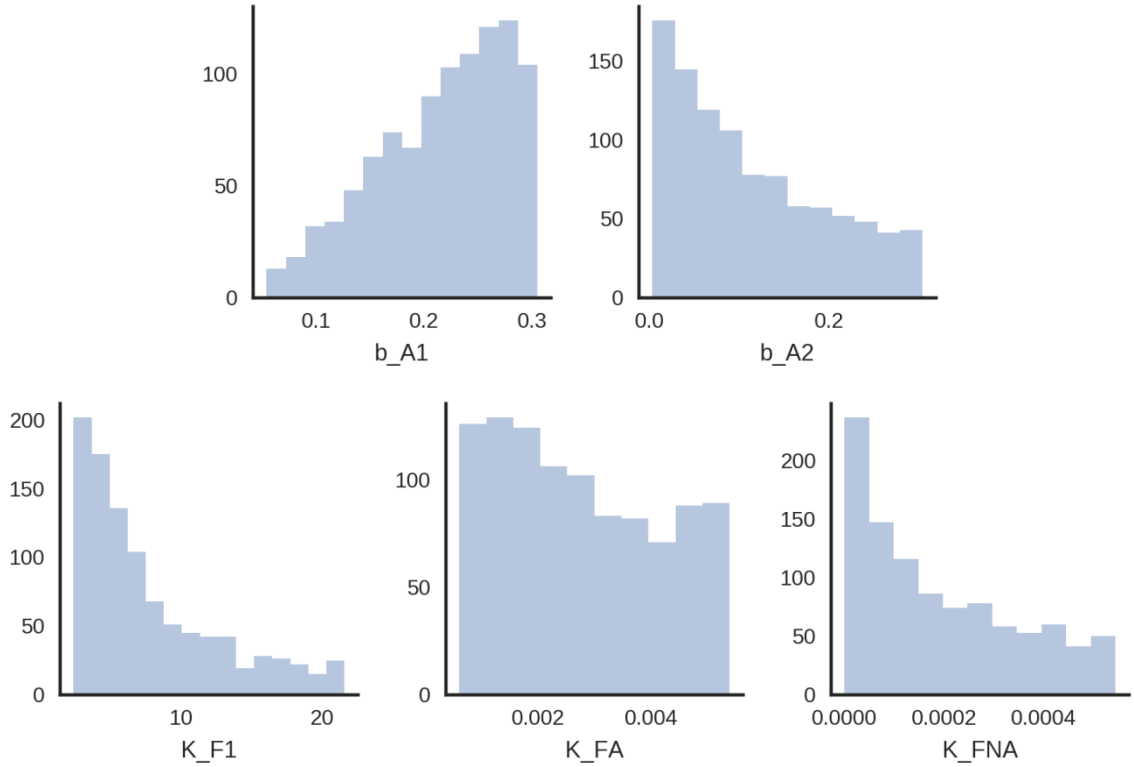


Figure 7 – Distribution of the parameter values resulting from the selection of the best performing scenarios for DO in the CM layout.

The best performing scenarios ranked according to TSS returned a well-defined shape of the b_{A1} distribution with high frequency value at the highest edge of its domain (data not shown). In the case of the TIS layout, i.e. where two peaks were observed in correspondence of both the highest and the lowest edges of the domain, this was less clear than for the CM. This is an important point in the evaluation of the performances of the two layouts as the case of b_{A1} for the TSS ranking is the first clue of the gain in definition of the CM as compared to the TIS layout.
In the overall view, merging the three groups of best performing scenarios resulting from the ranking for NH_4, DO, and TSS, a clear tendency of b_{A1} to show higher frequency in the highest part of its domain was noticed. This was a clear indication of the need for a redefined domain for b_{A1} before passing to Step III.
Similarly, b_{A2}, K_{F1}, and K_{FNA}, returned a clear preference of the highest frequency of their distribution plot for the lower edge of their domain.

Redefinition of parameter domains
According to the results of Step II for the cases of the TIS and CM layouts, some of the parameters show a clear potential for the 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 experiments that likely result in a less good prediction and are not very useful in the analysis anyway. The domains of b_{A1}, Y_{A1}, K_{FA}, and K_{F1}, were modified as indicated in Table 3.

Table 3 – New domains for the selected parameters derived from the results of both the TIS and CM layouts results.

Parameter	Minimum value	Maximum value
b_{A1}	0.15	0.40
Y_{A1}	0.04	0.20
K_{FA}	0.001	0.005
K_{F1}	2.4	21.6

Given the known tendency reported in literature for abating K_{FNA} values close to zero in order to accomplish a model fit, and given that those values are recognized to be unrealistic, the domain of K_{FNA} was not modified. In addition to this, the modification of four parameters domains could already have a positive effect on K_{FNA}.

Dynamic simulations (Step III)

The model was initialized with a steady state simulation of 100 days for performing the dynamic simulations. For doing this, the choice of a scenario for initialization was needed. Using the intersection of the three groups of best performing scenarios, i.e. the scenarios considered the best at the same time for NH_4, DO, and TSS cases, the best scenario according to all three cases could be identified.

Step III was targeted at defining the best performing scenarios analyzing the dynamic simulations output against measured data in specific locations on the bioreactor. The aim of this phase was to compare the capabilities of the TIS and CM layouts in defining a good 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 the capability of providing a realistic and observable parameters range of best performing values. Therefore, the ranking used the same method as for Step II, but using online measured data as objective functions of the metric comparison (i.e. NH_4, DO, N_2O and NO_3).

It must be pointed out that for the case of N_2O it was not possible to use all 12 metrics due to the fact that some metrics use the value of the objective function 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 N_2O output.
For the TIS layout, the ranking according to the measured NH_4 (Figure 8) 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 lays 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.

Concerning the ranking according to DO, all metrics resulted behaving similarly and overall agreeing in a common final ranking.

Although the ranking according to N_2O was forced to have fewer metrics, those metrics used were still coming from different categories, thus ensuring a ranking according to different approaches. All metrics appear to rank accordingly, although the fast transition towards the

darkest colors suggests the presence of few scenarios performing significantly better than the rest. In a similar picture the ranking according to NO_3 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 objective measured dataset.

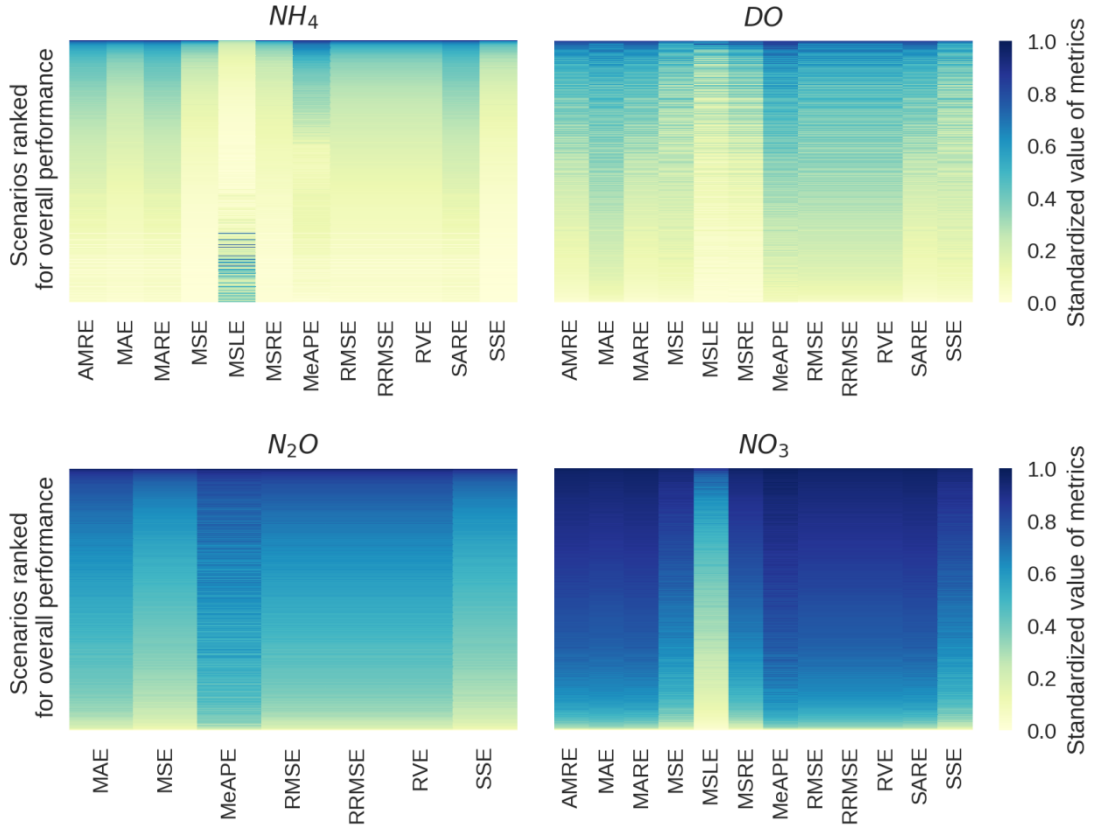
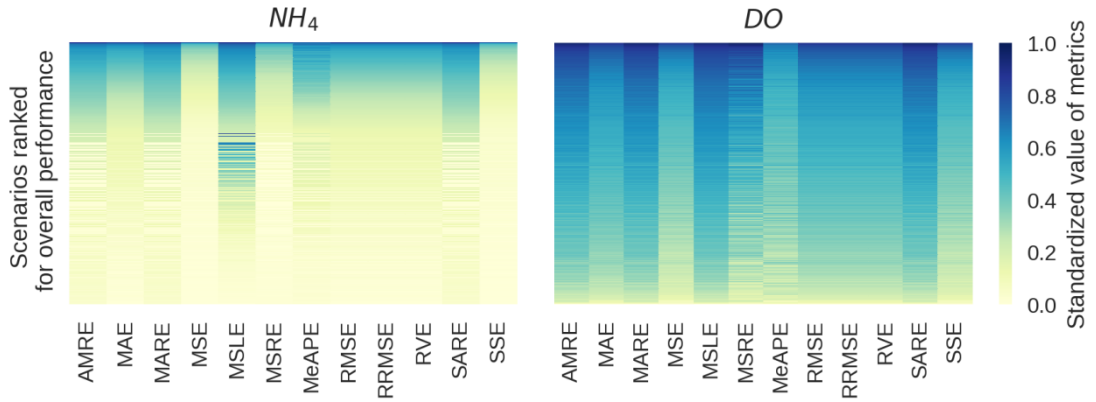


Figure 8 - Ranking of the scenarios (rows) according to the metrics (columns) from the best performing (bottom) to the worst (top). Each metric is colored according to its relative ranking from 0 to 1. Results of the TIS layout.

Figure 9 shows the ranking for the scenarios of the CM layout. Small differences can be observed among the metrics for the ranking according to NH_4 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 few scenarios are providing an output close to the measured dataset while the rest is quickly deviating away of it.



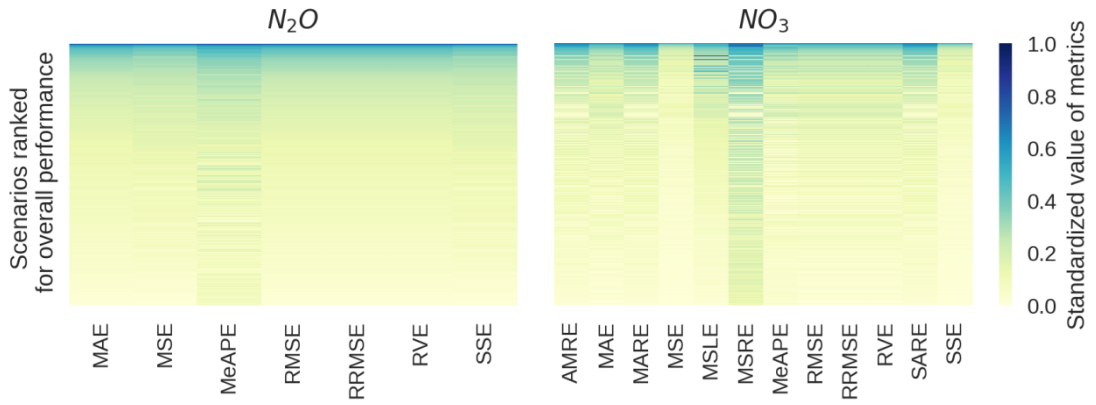


Figure 9 – Ranking of the scenarios (rows) according to the metrics (columns) from the best performing (bottom) to the worst (top). Each metric is colored according to its relative ranking from 0 to 1. Results of the CM layout.

Differently from the case of DO, the case of N_2O and NO_3 present a very gradual shift away from the objective function making all metrics generally providing the same ranking (for N_2O fewer metrics are considered).

Comparison between TIS and CM

The overall distributions of the parameter values for the best performing scenarios derived from the ranking for NH_4, DO, N_2O, and NO_3 are reported to make a global comparison of the performances of both model layouts in defining ranges of parameter values that are best performing.

For the case of Y_{A1} (Figure 10), the CM configuration (right) returned a clearly defined range of acceptable parameter values as compared to the case of the TIS layout. The Y_{A1} distribution of the CM appears to define a normally shaped curve which encounters a maximum frequency around the value of 0.1 g COD/g N. The TIS model (Figure 10, left) identifies the best performing scenarios in the lowest range of Y_{A1}, which are less realistic values as compared to the case of the CM.

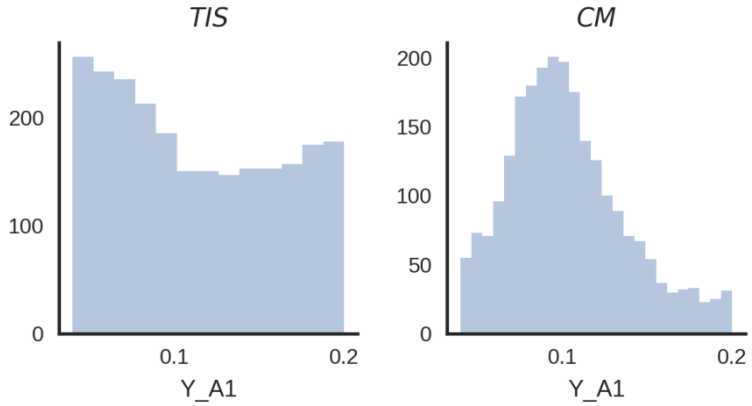


Figure 10 – Distributions of overall best performing scenarios for the case of the parameter values of Y_{A1} in the dynamic simulations with the TIS (left) and CM (left) layouts.

K_{FNA} (Figure 11), is a known difficult parameter to calibrate which is often abated to values very close to zero to force calibration fit (Spérandio et al., 2016). The CM results (Figure 11, right) show a more pronounced shape of a distribution as compared to the TIS, peaking in frequency around the value of 3E-4 g/m^3. This is an important indication finally proposing more realistic values for this parameter and to revert the general tendency of abating this parameter down to 1E-6. On the other hand, the TIS layout does not show a definite

distribution having almost everywhere the same frequency. However, it must be pointed out how the far right edge of the distribution for the TIS is slightly increasing in frequency suggesting the possibility of a need for a modification of the K_{FNA} domain.

In this view, it is interesting to consider that, despite the literature studies generally reporting very low values of K_{FNA}, the TIS layout reverts this tendency showing this time a propensity for more realistic values. 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.

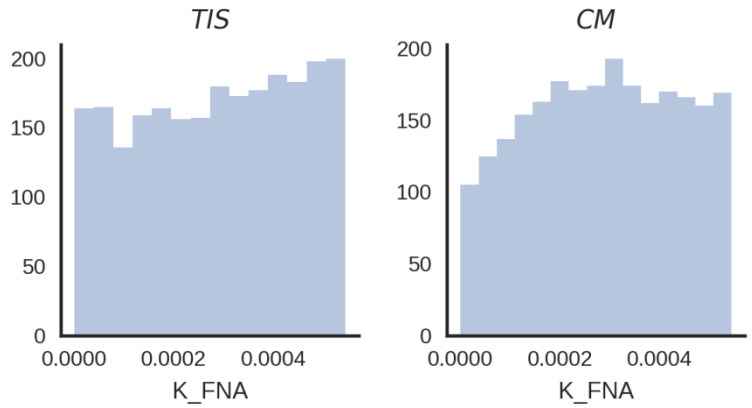


Figure 11 – Distributions of overall best performing scenarios for the case of the parameter values of K_{FNA} in the dynamic simulations with the TIS (left) and CM (left) layouts

Finally, looking at K_{F1}, it is interesting how both distributions have a similar shape (Figure 12), though more pronounced for the case of the CM layout. The distribution of K_{F1} returned by the TIS layout is noticeably flatter than the one returned by the CM. This can be considered another indication of the increased identifiability of the K_{F1} by means of the CM layout.

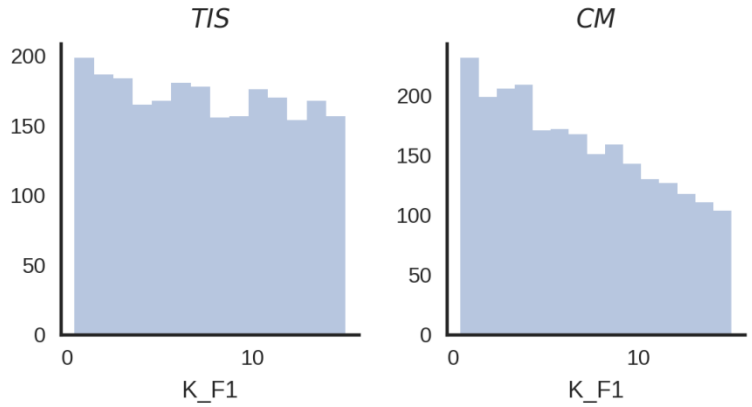


Figure 12 – Distributions of overall best performing scenarios for the case of the parameter values of K_{F1} in the dynamic simulations with the TIS (left) and CM (left) layouts

Model fits

The fitting of a model against measured values is the primary target of every calibration and modelling exercise. However, before that, the modeler should consider to be working as close as possible to an accurate representation of the reality, regarding both physical and biological aspects. Thus, firstly aiming to working with realistic parameter values.

It must be pointed out that the present work does not represent a calibration exercise and model fits against measured data were not shown with the purpose of focusing the reader's

attention in evaluating the ability of a model layout to select its suitable parameter domains. However, for completeness, model fits using the best performing scenario are shown in this section for a qualitative assessment of the model performances.

The best performing scenario was chosen according to all variables considered in Step III (NH_4, DO, N_2O, and NO_3) using the intersection of the four groups of best performing scenarios. In the end, the scenario selected for a layout is not the best fitting according to the single variable, but the best performing overall.

For the case of TIS, the intersection group resulted empty, and NO_3 had to be excluded from the intersection of the four groups. The best scenario of this selection, was used to show the results of the TIS (Figure 13, left).

For the case of CM, the selection of an overall best performing scenario was less problematic. The different groups of best performing scenarios from for NH_4, DO, N_2O and NO_3, were in accordance for 39 scenarios.

Table 4 reports the parameter values for the scenario selected for the TIS and CM layouts respectively. It is interesting to notice that the difference in Y_{A1} corroborating with what observed earlier in the parameter distribution. The same seems to be shown for the K_{FNA} parameter.

Table 4 – Parameter values for the scenario performing the best according to all variables of comparison for the TIS and for the CM layout.

Parameter	TIS	CM
$K_{O\_A1Lysis}$	0.116767	0.103970
$K_{O\_A2Lysis}$	0.319499	0.623322
b_{A1}	0.254775	0.177371
b_{A2}	0.211606	0.031296
Y_{A1}	0.06097	0.103379
Y_{A2}	0.2377	0.126444
$n_{NOx\_A1\_d}$	0.498316	0.565986
K_{FA}	0.005617	0.003866
K_{FNA}	0.000292	0.000206
K_{I9FA}	0.660808	0.703306
K_{OA1}	0.394944	0.550270
K_{OA2}	1.217972	1.145907
K_{F1}	2.230151	6.928328
K_{O1_BH}	0.758667	0.505318

In Figure 13 are reported the model results of the TIS (left) and CM (right) in comparison with the measured time series. The TIS model initially accumulates NH_4 until DO reaches a reasonably high concentration, not matching the measured values. On the other hand, although the CM over predicts DO while the TIS gives its best fit with it, the CM maintains NH_4 levels closer to what are the measured values.
In terms of N_2O, the CM misses the first peak but maintains a concentration that seems to resemble the measured one with reasonable accuracy. The TIS model layout gives the worst performance in terms of NO_3, where modelled concentrations remain at very low values as compared to the measured profile. The CM layout appears to provide modelled NO_3 results closer to the reality. Despite the fact that NO_3 levels are resulting pretty constant from the CM

and the peak in the measured NO_3 is not detected, the NO_3 concentration appears closer to the measured NO_3 than for the case of the TIS layout.

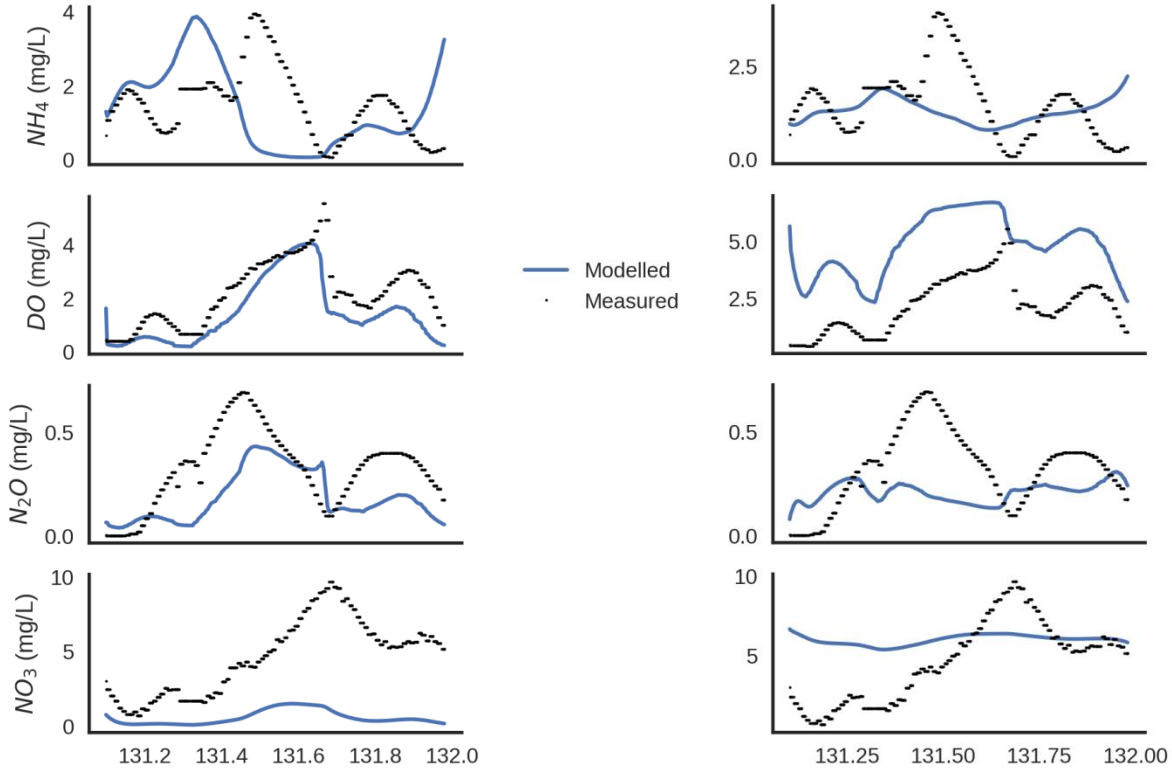


Figure 13 – Model outputs of the TIS (left) and CM (right) layouts in comparison with the full-scale measurements.

As specified earlier, this work does not want to be a calibration exercise, and model fits are reported for a more complete view. However, the benefits of the use of the CM layout have been evidenced once more.

Considering that the CM is a recent development of the more exploited TIS layout, the performances of the CM are showing an important potential and the advantage of increasing the level of accuracy of local recirculation and concentrations with respect to the reality.

CONCLUSIONS

In the present work, a ranking method and a visualization were proposed for selecting the best performing scenarios and providing a qualitative indication of the performance and contribution of each metric used.

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 of different nature. Also, the performance of each metric was highlighted in its ranking and the relative effect on the overall arrangement 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 results significantly. This implies a relevant gain in accuracy that allowed to redefine some of the key parameters to acceptable values and obtain more defined distributions.

The CM generally returned a more narrow parameter domain of good values with respect 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.
Further developments of the CM will consider the possibility of varying the volumes of the different compartments according to the influent flow.

REFERENCES

Ahn, J.H., Kim, S., Park, H., Katehis, D., Pagilla, K., Chandran, K., 2010. Spatial and temporal variability in atmospheric nitrous oxide generation and emission from full-scale biological nitrogen removal and non-BNR processes. Water Environ. Res. a Res. Publ. Water Environ. Fed. 82, 2362–72.

Amerlinck, Y., 2015. Model refinements in view of wastewater treatment plant optimization: improving the balance in sub-model detail. Ghent University.

Arnaldos, M., Amerlinck, Y., Rehman, U., Maere, T., Van Hoey, S., Naessens, W., Nopens, I., 2015. From the affinity constant to the half-saturation index: Understanding conventional modeling concepts in novel wastewater treatment processes. Water Res. 70, 458–470.

Cierkens, K., Nopens, I., De Keyser, W., Van Hulle, S.W.H., Plano, S., Torfs, E., Amerlinck, Y., Benedetti, L., Van Nieuwenhuijzen, A., Weijers, S., de Jonge, J., 2012. Integrated model-based optimisation at the WWTP of Eindhoven. Water Pract. Technol. 7, 1–9.

Daelman, M.R.J., van Voorthuizen, E.M., van Dongen, U.G.J.M., Volcke, E.I.P., van Loosdrecht, M.C.M., 2015. Seasonal and diurnal variability of N2O emissions from a full-scale municipal wastewater treatment plant. Sci. Total Environ. 536, 1–11.

De Keyser, W., Amerlinck, Y., Urchegui, G., Harding, T., Maere, T., Nopens, I., 2014. Detailed dynamic pumping energy models for optimization and control of wastewater applications. J. Artic. 5, 299–314.

De Pauw, D.J.W., Vanrolleghem, P. a, 2006. Practical aspects of sensitivity function approximation for dynamic models. Math. Comput. Model. Dyn. Syst. 12, 395–414.

Gernaey, K. V., Jørgensen, S.B., 2004. Benchmarking combined biological phosphorus and nitrogen removal wastewater treatment processes. Control Eng. Pract. 12, 357–373.

Guo, L., 2014. Greenhouse gas emissions from and storm impacts on wastewater treatment plants: Process modelling and control. LAVAL University.

Guo, L.S., Vanrolleghem, P. a., 2014. Calibration and validation of an activated sludge model for greenhouse gases no. 1 (ASMG1): Prediction of temperature-dependent N2O emission dynamics. Bioprocess Biosyst. Eng. 37, 151–163.

Hauduc, H., Neumann, M.B., Muschalla, D., Gamerith, V., Gillot, S., Vanrolleghem, P.A., 2015. Efficiency criteria for environmental model quality assessment: A review and its application to wastewater treatment. Environ. Model. Softw. 68, 196–204.

Hiatt, W.C., 2006. Activated sludge modeling for elevated nitrogen conditions. Clemson University.

Hiatt, W.C., Grady, C.P.L., 2008. An updated process model for carbon oxidation, nitrification, and denitrification. Water Environ. Res. 80, 2145–2156.

Le Moullec, Y., Gentric, C., Potier, O., Leclerc, J.P., 2010. Comparison of systemic, compartmental and CFD modelling approaches: Application to the simulation of a biological reactor of wastewater treatment. Chem. Eng. Sci. 65, 343–350.

Mampaey, K.E., Beuckels, B., Kampschreur, M.J., Kleerebezem, R., van Loosdrecht, M.C.M., Volcke, E.I.P., 2013. Modelling nitrous and nitric oxide emissions by autotrophic ammonium oxidizing bacteria. Environ. Technol. 34, 1555–66.

Mannina, G., Ekama, G., Caniani, D., Cosenza, A., Esposito, G., Gori, R., Garrido-Baserba, M., Rosso, D., Olsson, G., 2016. Greenhouse gases from wastewater treatment - A review of modelling tools. Sci. Total Environ.

Ni, B.J., Peng, L., Law, Y., Guo, J., Yuan, Z., 2014. Modeling of Nitrous Oxide Production by Autotrophic Ammonia-Oxidizing Bacteria with Multiple Production Pathways. Env. Sci Technol 48, 3916–3924.

Ni, B.J., Ye, L., Law, Y., Byers, C., Yuan, Z., 2013a. Mathematical modeling of nitrous oxide (N2O) emissions from full-scale wastewater treatment plants. Environ. Sci. Technol. 47, 7795–803.

Ni, B.J., Yuan, Z., 2015. Recent advances in mathematical modeling of nitrous oxides emissions from wastewater treatment processes. Water Res.

Ni, B.J., Yuan, Z., Chandran, K., Vanrolleghem, P. a., Murthy, S., 2013b. Evaluating four mathematical models for nitrous oxide production by autotrophic ammonia-oxidizing bacteria. Biotechnol. Bioeng. 110, 153–163.

Pocquet, M., Wu, Z., Queinnec, I., Spérandio, M., 2015. A two pathway model for N2O emissions by ammonium oxidizing bacteria supported by the NO/N2O variation. Water Res. 88, 948–959.

Rehman, U., 2016. Next generation bioreactor models for wastewater treatment systems by means of detailed

combined modelling of mixing and biokinetics. Ghent University.
Rehman, U., Audenaert, W., Amerlinck, Y., Maere, T., Arnaldos, M., Nopens, I., 2017. How well-mixed is well mixed? Hydrodynamic-biokinetic model integration in an aerated tank of a full-scale water resource recovery facility. Water Sci. Technol. 76, 1950–1965.
Rehman, U., Maere, T., Vesvikar, M., Amerlinck, Y., Nopens, I., 2014a. Hydrodynamic-biokinetic model integration applied to a full-scale WWTP, in: 9th IWA World Water Congress and Exhibition. Lisbon.
Rehman, U., Maere, T., Vesvikar, M., Amerlinck, Y., Nopens, I., 2014b. Hydrodynamic - biokinetic model integration applied to a full-scale, in: IWA World Water Congress & Exhibition. Lisbon, Portugal.
Rehman, U., Vesvikar, M., Maere, T., Guo, L., Vanrolleghem, P.A., Nopens, I., 2015. Effect of sensor location on controller performance in a wastewater treatment plant. Water Sci. Technol. 71, 700.
Spérandio, M., Pocquet, M., Guo, L., Ni, B.J., Vanrolleghem, P.A., Yuan, Z., 2016. Evaluation of different nitrous oxide production models with four continuous long-term wastewater treatment process data series. Bioprocess Biosyst. Eng. 39, 493–510.
van Griensven, A., Meixner, T., Grunwald, S., Bishop, T., Diluzio, M., Srinivasan, R., 2006. A global sensitivity analysis tool for the parameters of multi-variable catchment models. J. Hydrol. 324, 10–23.
Van Hoey, S., 2016a. Development and application of a framework for model structure evaluation in environmental modelling. Ghent University.
Van Hoey, S., 2016b. Development and application of a framework for model structure evaluation in environmental modelling.
Van Hulle, S.W.H., Callens, J., Mampaey, K.E., van Loosdrecht, M.C.M., Volcke, E.I.P., 2012. N2O and NO emissions during autotrophic nitrogen removal in a granular sludge reactor--a simulation study. Environ. Technol. 33, 2281–90.