Marine Pollution Bulletin 127 (2018) 217–224

Contents lists available at ScienceDirect

Marine Pollution Bulletin

journal homepage: www.elsevier.com/locate/marpolbul

Modeling marine oily wastewater treatment by a probabilistic agent-based approach

Liang Jing[a], Bing Chen[a,b,*], Baiyu Zhang[a], Xudong Ye[a]

[a] Northern Region Persistent Organic Pollution Control (NRPOP) Laboratory, Faculty of Engineering and Applied Science, Memorial University of Newfoundland, St. John's, NL A1B 3X5, Canada

[b] College of Environmental Science and Engineering, Peking University, Beijing, China, 100871

ARTICLE INFO

Keywords:
Marine oily wastewater treatment
Agent-based modeling
Probability-based simulation
Reaction competition

ABSTRACT

This study developed a novel probabilistic agent-based approach for modeling of marine oily wastewater treatment processes. It begins first by constructing a probability-based agent simulation model, followed by a global sensitivity analysis and a genetic algorithm-based calibration. The proposed modeling approach was tested through a case study of the removal of naphthalene from marine oily wastewater using UV irradiation. The removal of naphthalene was described by an agent-based simulation model using 8 types of agents and 11 reactions. Each reaction was governed by a probability parameter to determine its occurrence. The modeling results showed that the root mean square errors between modeled and observed removal rates were 8.73 and 11.03% for calibration and validation runs, respectively. Reaction competition was analyzed by comparing agent-based reaction probabilities, while agents' heterogeneity was visualized by plotting their real-time spatial distribution, showing a strong potential for reactor design and process optimization.

1. Introduction

Modeling of wastewater treatment processes has been of great importance due to the need for a full understanding of complex treatment systems and the optimization of their practical applications. Numerous modeling techniques, such as reaction kinetics and equilibrium (Mbamba et al., 2015; Eglal and Ramamurthy, 2015), computational fluid dynamics (Wols et al., 2015; Santoro et al., 2015) and artificial neural networks (Jing et al., 2015; Jing et al., 2016) have been extensively documented in the literature. An interesting note is that most of the existing techniques are designed to simulate bulk properties and rely much on the understanding of population-level dynamics rather than individual-level responses, such as heterogeneity among individuals, local interactions, and adaptive decision making (Schuler et al., 2011). The lack of individual-level information may result in an incomplete understanding of treatment processes, especially when competition and complex interactions among different components exist (Schuler, 2005, 2006; Hellweger and Bucci, 2009). For example, Hellweger (2007) argued that traditional lumped modeling approaches can introduce significant errors when simulating individual phytoplankton growth and trace metal transformation during eutrophication. Further, knowledge on the adsorption competition for photocatalytic active sites (Wang et al., 2009) and heavy metal removal (Eglal and Ramamurthy, 2015), and microbial competition in biological wastewater treatment systems (Albuquerque et al., 2013) has not been made explicit by population-level techniques.

Recently, the importance of individual variations in wastewater treatment processes has favored the advancement of process-based models, particularly agent-based models (ABMs). These bottom-up models are particularly known for describing the behavior of individual system components from a micro point of view and yielding different predictions of bulk behaviors than conventional macro-scale approaches where population heterogeneity and intra-population variability are usually not appreciated. The core of ABMs emphasizes on the autonomous and adaptive nature of individuals such that each individual has its own characteristics/goals and can make its own decisions according to certain rules. By capturing the interactions of individuals with each other and with their surrounding environment, the behavior of the population can be simulated as a whole (Wilensky and Rand, 2015). Given most scientific knowledge exists at either individual- or population-level instead of both, ABMs can bridge this gap via exploring the effects of individual decisions on collective behavior and predicting how populations will change across time and space. Such an intuitive concept makes ABMs well-suited to model highly

* Corresponding author at: Northern Region Persistent Organic Pollution Control (NRPOP) Laboratory, Faculty of Engineering and Applied Science, Memorial University of Newfoundland, St. John's, NL A1B 3X5, Canada.
E-mail address: bchen@mun.ca (B. Chen).

https://doi.org/10.1016/j.marpolbul.2017.12.004
Received 11 June 2017; Received in revised form 26 October 2017; Accepted 2 December 2017
Available online 08 December 2017
0025-326X/

complex dynamics using only simpler rules and assumptions. Nonetheless, it is worth noted that the efficacy of agent-based modeling may be compromised by its underdeveloped theoretical framework and high computational demands (Esser et al., 2015). ABMs have been applied to a variety of problems such as human and animal migration, traffic control, stock trading, and land development (Bonabeau, 2002; Matthews et al., 2007; Hellweger and Kianirad, 2007; Crooks et al., 2008; Farmer and Foley, 2009; Schreinemachers and Berger, 2011; Filatova et al., 2013; Müller et al., 2013; Klabunde and Willekens, 2016). It has been recently argued that ABMs should have high potential for wastewater treatment modeling. The movement, behavior, and spatial distributions of different agents in wastewater, such as pollutants, reactive species, and microorganisms can be well simulated by ABMs. Subsequently, population-level behaviors (e.g., concentrations of pollutants) would emerge as a result of cumulative actions of agents (Schuler et al., 2011). In addition, ABMs do not have the same restrictions as population-level differential equations and can easily account for the numbers, distributions, and time delays of different molecules (Pogson et al., 2006). Xavier et al. (2007) applied agent-based simulation to describe the complex dynamics of four bacterial groups in an aerobic granular sludge sequencing batch reactor. Pereda and Zamarreno (2012) proposed a Matlab-based ABM for modeling activated sludge process in a batch reactor and obtained a better understanding of this phenomena. Bucci et al. (2012) developed an ABM to predict the heterogeneity of an enhanced biological phosphorus removal process. The biological variability in individual cell behavior and states was predicted by randomizing model parameters and state variables, respectively. However, to the best of our knowledge, the applications of ABMs in modeling wastewater treatment processes are still limited in the literature (Gernaey et al., 2004; Schuler et al., 2011). Most of the relevant studies have either used macro-scale behavior rules such as partial differential equations to approximate individual-level responses, or focused on simplified processes without considering reaction kinetics. In addition, the calibration of existing ABMs has mostly been done by trial and error attempts, which relies mainly on expert knowledge and may encounter problems when a large amount of experimental data is available.

Therefore, to help fill these gaps, the objective of this study was to establish a novel agent-based probabilistic approach for modeling of wastewater treatment processes by 1) developing a probability-based agent simulation model and 2) improving its performance through a global sensitivity analysis and a genetic algorithm-based calibration. The removal of naphthalene (NAP) from marine oily wastewater by UV induced photodegradation (experimental results adopted from Jing et al., 2014a) was modeled as a demonstrative example to examine the applicability and accuracy of the proposed modeling approach.

2. Methodology

2.1. Experimentation data

The original work of Jing et al. (2014a) was comprised of a factorial experimental design to study the removal of NAP from marine oily wastewater. In brief summary, photodegradation of NAP was carried out in a two-layer cylindrical reactor (Fig. S1 in Supplementary material). The outer body and inner section are made of aluminum and clear fused quartz, respectively, with eight 18.4 W low-pressure UV lamps (254 nm peak, full width half maximum of 15 nm, Atlantic Ultraviolet, Canada) evenly positioned between them. The inner quartz beaker (11.1 cm internal radius, 11.5 cm external radius, and 20 cm height) features a polycarbonate lid where a stainless steel paddle agitator, a 50 W heater, and a thermometer are mounted. The experimental system was constantly stirred in order to ensure a well-mixed solution. NAP (> 99%) and NAP D8 (> 99%, internal standard) were obtained from Aldrich, Canada. Dichloromethane and acetone were purchased from Honeywell Burdick and Jackson (USA) for stock solution preparation and extraction. Seawater was obtained from a clean coastal site in St. John's, Canada and filtered through a 5 μm filter to remove organisms and debris. UV lamps were turned on once NAP stock solution was spiked into 6 L seawater to simulate the discharge of oily wastewater from offshore industries (Jing et al., 2015). Samples were taken half-hourly and sent for GC-MS analysis after pre-treatment. More details about the experiments are available in Jing et al. (2014a).

In this study, a probability-based agent simulation model was first developed to simulate the removal of NAP and then calibrated with data from 8 calibration runs (30 min interval from 0.5 to 4 h, excluding the start point; 8 samples per run, 64 data points in total). Data from another 4 validation runs (32 data points in total) were used to evaluate the performance of the calibrated model. Details about the calibration and validation runs can be found in Tables S1 and S2 in Supplementary material.

2.2. The agent-based simulation model

A probability-based agent simulation model for NAP removal was developed in the free software platform NetLogo 5.3.1 because it is an intuitive and well-documented programming tool with high flexibility and simplicity. NetLogo is a modified version of the Logo programming language and is particularly suitable for modeling complex systems which change over time. It employs a graphical environment to support numerous breeds of programmable agents to wander and interact in a grid of patches. All types of agents can interact with each other and perform multiple tasks concurrently by receiving pre-set instructions on their behaviors from the user. Detailed introduction to NetLogo can be found in Wilensky and Evanston (1999) and Wilensky and Rand (2015).

In this study, UV induced photodegradation of NAP in seawater can be described by the following simplified reaction scheme (Miller and Olejnik, 2001). It is assumed that radical reactions of chloride and nitrate are not involved given their insignificant contributions (Lair et al., 2008; Fang et al., 2017). NAP can absorb one photon ($h\nu$ in Eq. 1) to transit to its excited state NAP* (Bayrakceken et al., 2012), which can further return to the ground state (Eq. 2) by dissipating energy via various ways such as internal conversion, energy transfer to other molecules, and emitting its characteristic fluorescence. Oxygen molecule (O_2), which is known to be an efficient quencher of excited state PAHs, in the ground triplet state can be excited into a reactive singlet state ($^1O_2$) by encountering a NAP* (Eq. 3). $^1O_2$ can then attack NAP and form products such as peroxides and hydroperoxides (Eq. 4) or return to the ground state by emitting energy (Eq. 5). In addition, humic substances (HS), which are the main components of natural organic matter in seawater, can also absorb one photon to transit to its excited state HS* (Eq. 6), return to the ground state (Eq. 7), react with O_2 to generate $^1O_2$ (Eq. 8), or be degraded by $^1O_2$ (Eq. 9) to form products. All photodegradation products are assumed to further react with $^1O_2$ and generate one or two products (Eqs. 10 and 11). It should be noted that Eqs. 6–11 are a lumped approximation of many photoreactions associated with HS and photochemical products.

$$\mathrm{NAP} + h\nu \xrightarrow{p_1} NAP^* \quad (1)$$

$$NAP^* \xrightarrow{p_2} \mathrm{NAP} + \mathrm{energy} \quad (2)$$

$$\mathrm{NAP}^* + \mathrm{O_2} \xrightarrow{p_3} NAP + {}^1O_2 \quad (3)$$

$$\mathrm{NAP} + {}^1\mathrm{O_2} \xrightarrow{p_4} \mathrm{Product} \quad (4)$$

$${}^1\mathrm{O_2} \xrightarrow{p_5} \mathrm{O_2} + \mathrm{energy} \quad (5)$$

$$\mathrm{HS} + h\nu \xrightarrow{p_6} \mathrm{HS}^* \quad (6)$$

$$HS^* \xrightarrow{p_7} \mathrm{HS} + \mathrm{energy} \quad (7)$$

$$HS^* + O_2 \xrightarrow{p_8} HS + {}^1O_2 \quad (8)$$

$$HS + {}^1O_2 \xrightarrow{p_9} \text{Product} \quad (9)$$

$$\text{Product} + {}^1O_2 \xrightarrow{p_{10}} \text{Product} \quad (10)$$

$$\text{Product} + {}^1O_2 \xrightarrow{p_{11}} \text{Product} + \text{Product} \quad (11)$$

where p are set as fixed values in order to control the occurrence of the reactions based on probability theory (see Step 3 below). According to Eqs. 1–11, there are 8 different types of agents, including NAP, NAP*, photon, O_2, $^1O_2$, product, HS, and HS*, and 1 type of patch. By default, a two-dimensional simulation environment is set to approximate the vertical cross section of the reactor with a grid of patches that had 33 rows and 33 columns (1089 patches in total) where each square patch is 13 pixels wide (Fig. S5 in Supplementary material). The two-dimensional projection of the three-dimensional cylindrical reactor is used for simplicity, given the UV irradiation and the solution are homogeneous and the reactor is symmetric. The space taken by the agitator, thermometer and heater is assumed to be negligible. The simulation steps can be summarized as follows:

Step 1: Initialization

Prior to running the model, the values of all model parameters (Table 1) and the initial numbers of NAP, photon, O_2, HS, and temperature need to be specified. The position (i.e., x-y coordinates in pixels) and heading direction of each agent are randomized. All agents are given a size of 1 by 1 pixel. Once the parameter initialization is completed, the model is executed to run tick by tick (i.e., NetLogo timer function, 1 tick = 1 simulation step).

Step 2: Move

During each tick, all agents are first asked to wander by turning left a random positive degree less than d_L, turning right a random positive degree less than d_R, and then moving a distance of D (pixel) along the final direction (Table 2). To consider the effect of temperature, if temperature changes from T_1 to T_2 (°C), then D is assumed to change as suggested by a temperature enhancement parameter m:

$$D_2 = D_1 \times ((T_2 + 273.15)/(T_1 + 273.15))^m \quad (12)$$

Step 3: Reactions

The second law of photochemistry defines that for each photon of light absorbed, only one molecule is activated for a photochemical reaction (Roberts et al., 2017). During each tick, as indicated in Eq. 1, when a NAP moves to the same position as a photon dose, this particular NAP has a certain probability to partner with this photon and be converted to a NAP* while the photon subsequently diminishes. This probability is expressed by comparing a random number generated within a certain range (e.g., 0–100) with a preset probability parameter p_1 (e.g., 50). If the random number is less than the value of p_1, then the reaction can proceed; otherwise it cannot and the two agents will move apart from each other. In other words, the use of probability-based interpretation here aims at differentiating reaction tendencies. An assumption here is that a temperature change can also affect the probability of all reactions (i.e., parameter n) as shown below:

$$Random\ number/((T_2 + 273.15)/(T_1 + 273.15))^n < p_1 \quad (13)$$

The rest of the reactions (Eqs. 2–11) can be explained by the same concept. By assigning different p values to individual reactions, the simulation model is able to mimic the difference in reaction tendencies using a grey-box approach. At the end of each tick, the numbers of different agents can be recorded and plotted.

Step 4: Termination

Repeat steps 2–3 for a preset number of ticks and then the simulation is automatically terminated. The performance of the simulation was evaluated by comparing simulation results with experimental observations in remaining NAP percentage (%) at multiple time points. It should be noted that the proposed agent-based simulation model can be modified and used for many other treatment processes.

Table 2
Baseline values, variation ranges, main effects and calibrated values for all parameters.

Parameter	Unit	Baseline value	Variation range	Overall effect index (%)	Calibrated value
p_1	–	50	1–100	36.1	8.4
p_2	–	50	1–100	0.1	50
p_3	–	50	1–100	0.2	50
p_4	–	50	1–100	27.1	39.1
p_5	–	50	1–100	0.1	50
p_6	–	50	1–100	0.9	1.0
p_7	–	50	1–100	0.6	41.2
p_8	–	50	1–100	1.1	13.4
p_9	–	50	1–100	8.1	94.8
p_{10}	–	50	1–100	17.3	4.3
p_{11}	–	50	1–100	3.1	39.2
m	–	25	0–50	0.1	25
n	–	25	0–50	4.9	7.2
D	pixel	5	1–10	0.1	5
d_L	degree	180	0–360	0.1	180
d_R	degree	180	0–360	0.1	180

Note: p_1, p_2, and p_8 are fixed at 50 during calibration according to sensitivity analysis results.

2.3. Model initialization and settings

According to Tables S1 and S2, given the initial concentration of NAP was 10 μg L^{-1}, the number of NAP molecules spiked in 6 L seawater was calculated as 2.8×10^{17}. The initial number of oxygen molecules (i.e., assumed at 4.5 mg L^{-1}) was estimated as 5×10^{20}. The number of photons was calculated from the UV fluence rate. For fluence rates of 2.88 and 8.27 mW cm^{-2}, the number of photons (5.44×10^{-19} J per photon) penetrating through the vertical cross section (1083 cm^2) was estimated as 5.73×10^{18} and 1.65×10^{19} per second, respectively. By assuming that one tick in the proposed model corresponded to 1 min in the experiment, then the number of photons was calculated as 3.44×10^{20} and 9.88×10^{20}, respectively.

It should be noted that running an agent-based simulation model

Table 1
Optimization of model initial conditions using all 8 calibration runs.

Trial #	Initial conditions[a]							*RMSE*[b] (%)	
	$Photon_1$	$Photon_2$	Oxygen	HS_1	HS_2	T_1	T_2	Mean	Std
1	172	494	250	125	200	23	40	35.69	0.05
2	860	2740	1250	625	1000	23	40	27.38	0.15
3	1720	4940	2500	1250	2000	23	40	25.41	0.07
4	4300	12,350	6250	3125	5000	23	40	25.56	0.09
5	8600	24,700	12,500	6250	10,000	23	40	25.26	0.34
6	17,200	49,400	25,000	12,500	20,000	23	40	25.56	0.21

[a] NAP is set to 1,400; $photon_1$ and $photon_2$ correspond to 2.88 and 8.27 mW cm^{-2} fluence rate, respectively; HS1 and HS2 refer to 25 and 40 psu salinity (Table S1).
[b] Each trial was repeated 5 times for all calibration runs. Model parameters were set to baseline values as shown in Table 2.

with large numbers of agents could be computationally expensive and time consuming (Ho et al., 2015). However, too small numbers of agents may result in some agents theoretically staying unselected for a long time, thereby leading to large fluctuations in modeling results (Sweda and Klabjan, 2014). As the main scope of this paper was to examine the potential of the proposed agent-based probabilistic approach for modeling of wastewater treatment processes, to reduce the computational burden, the initial numbers of various agents were selected based on their influence on the modeling performance using root mean square error (*RMSE*). Each agent can then be interpreted as a cluster of certain species of molecules. The simulated remaining NAP (%) at 30-tick intervals were compared with the experimental observations (i.e., 8 data points per run, 8 calibration runs, 64 data points in total) to calculate the *RMSE*. Given the stochastic nature of agent-based modeling, each combination of initial numbers of agents was repeated 5 times to get a mean *RMSE* and a standard deviation. As shown in Table 1, the number of NAP was set as 1400 to reduce simulation fluctuation. As the numbers of other agents increased, a decreasing trend was observed for the mean *RMSE* until it reached a plateau value around 25.41% in trial 3. The standard deviation was 0.07%, indicating that the numbers of agents were not too small to unveil fluctuations in modeling results. Therefore, the number of NAP and oxygen were set as 1400 and 2500, respectively. The numbers of photon were set as 1720 and 4940 for the measured average fluence rates of 2.88 and 8.27 mW cm^{-2}, respectively. In addition, the number of HS was set as 1250 and 2000 for salinity levels of 25 and 40 practical salinity unit, respectively. T_1 and T_2 were set as 23 and 40 °C, respectively.

The maximum number of ticks was set to 240 in accordance with the experimental data (i.e., 240 min duration). The baseline values of p and other model parameters were all set to the mid points of their preset ranges (Table 2). At the end of each simulation tick, the singlet excited state agents (i.e., NAP^*, $^1O_2$, and HS^*) that did not react with other agents at the same position were set to diminish because singlet excited state species usually have a drastically short half-life time. In addition, the numbers of photon and oxygen were readjusted to their initial values after each tick because UV fluence rate (i.e., photon supply) was kept constant whereas the available oxygen in seawater, as stated above, far exceeded the demand during the experiments (Jing et al., 2014a).

2.4. Baseline simulation and global sensitivity analysis

Due to the random nature of agent behavior, for each of the 8 calibration runs listed in Table S1, the baseline simulation was repeated 100 times by employing the baseline parameter settings. As all runs were sampled at 30-min intervals, therefore, the simulated remaining NAP percentages at 30-tick intervals were first averaged and then compared with the experimental observations to calculate the overall *RMSE* for all calibration runs.

To find out which model parameters were most influential to the overall *RMSE*, the Fourier amplitude sensitivity test (FAST) was conducted for all 16 model parameters (Table 2). FAST is a variance-based global sensitivity analysis method that can assign characteristic frequency to each parameter through a search function, and then single out each parameter's first-order (i.e., main effect) contribution to model output by its characteristic frequency. A total-effect index can be calculated to accommodate contributions from high order interactions. It is computationally efficient and can be used for nonlinear and non-monotonic models such as chemical reaction models (Haaker and Verheijen, 2004). More details about FAST can be found in Xu and Gertner (2008) and Vanuytrecht et al. (2014). In this study, all parameters were simultaneously sampled 500 times within their preset ranges via Matlab (Table 2). The overall importance of each parameter was then normalized by calculating its global contribution to the total variance of the overall *RMSE*. The most influential parameters (i.e., with normalized effect index > 0.5%) were then selected to be included in model calibration.

2.5. Model calibration using genetic algorithm

In order to improve model performance, model calibration was implemented by tuning selected parameters and matching simulated to experimental results. As mentioned earlier, NetLogo's own "BehaviorSpace" module can only provide iterative simulation results by changing one factor at a time. Users need to manually find the best parameter combination including environmental settings and the logic of agents' behaviors through trial and error. Therefore, to reduce calibration time and to enhance performance, this study adopted genetic algorithm (GA) to optimize the most influential parameters through the integration of Matlab and NetLogo. GA is a powerful probabilistic global optimization technique. It combines the "survival of the fittest" principle of natural evolution with a randomized information exchange which helps to form a stochastic search routine and produce new individuals with higher fitness. More details about GA can be found in Arad et al. (2013). By taking computation time and resource constraints into account, population size (N_p) and maximum generation count (N_g) were set as 50 and 20, respectively (Jing et al., 2015).

As shown in Fig. S2 in Supplementary material, model calibration started from creating a random initial population according to N_p. Each individual in the population had random and fixed values assigned to the most influential parameters and those non-influential ones, respectively. Then each individual was used as inputs to run the agent-based simulation model in headless mode, which can be directly called in Matlab using Java scripts. Simulation results for all calibration runs were averaged for overall *RMSE* calculation as stated in Section 2.4. The population was then ranked prior to generating its offspring population via GA operators including reproduction, mutation, and crossover. This process continued until any one of the convergence criteria of GA was met and then the top ranked individual was obtained as the final solution.

2.6. Model validation

Data from 6 extra validation runs (Table S2) were used to test if the calibrated model was able to well predict NAP removal using inputs that were different from the calibration runs. The prediction accuracy was evaluated using overall *RMSE* as stated above.

3. Results and discussion

3.1. Baseline simulation and global sensitivity analysis

A baseline simulation was conducted using the baseline parameter settings and repeated 100 times for each calibration run, from which an overall *RMSE* of 25.44% was obtained, indicating that the simplified reaction scheme (Eqs. 1–11) was reasonably justified. As shown in Fig. S3 in Supplementary material, in spite of the small range, the 100-time repeated simulations were still able to reflect the stochastic feature of agent heterogeneity and the randomized nature of agent-based modeling. For calibration runs 1–4, the modeled results were within reasonable agreement with the observed results. For example, in run 1, the observed remaining percentages at all eight sampling points were 47.2, 35.8, 19.5, 10.0, 4.9, 3.1, 1.7, and 1.4%, respectively. The average modeled remaining percentages were 22.5, 17.6, 15.7, 14.6, 13.7, 13.0, 12.3, and 11.9%, respectively, with a *RMSE* of 13.14% as compared to the observed results. However, for calibration runs 5–8, the modeled results largely deviated from observations with *RMSEs* of 36.01, 40.07, 23.90, and 26.89%, respectively. Note that the only difference between runs 1–4 and 5–8 was the change of the number of photons from 4940 to 1720. This may be attributed to the fact that p_1–p_{11} were not at their optimal values to balance the reactions. In addition, the temperature

factor n may also need to be adjusted in order to reduce such deviation.

The normalized overall importance of each parameter given by FAST is shown in Table 2. By adopting a threshold level at 0.5%, p_2, p_3, p_5, m, d_R, and d_R were determined to be non-influential; therefore, all other parameters were included in model calibration. The two most influential ones were found to be p_1 (36.1%) and p_4 (27.1%), which controls how fast NAP molecules can be excited and degraded, respectively. Interestingly, the temperature-related reaction probability enhancement factor n was also found to be influential. As suggested by Jing et al. (2014a), fluence rate and temperature were the two most important factors for the removal of NAP. Given fluence rate was already taken into account as the number of photons and was not included in FAST, the sensitivity analysis results therefore agreed with the experimental findings. The exclusion of p_2 and p_3 meant that the relaxation of NAP* was not significant to the removal process because a large amount of HS (Table 1) was also available for radical-induced reactions. The reason why p_5 was also excluded from calibration may be attributed to the fact that the relaxation of $^1O_2$ was mandatorily executed after each simulation tick.

3.2. Model calibration and validation

Fig. 1 shows that the modeling results after calibration were in better agreement with the experimental observations as compared to the trends before calibration (Fig. S3 in Supplementary material). The overall *RMSE* of all calibration runs after calibration was determined as 8.73%, which was lower than the baseline results (before calibration) of 25.44%, indicating the effectiveness of model calibration. For instance, in run 8, the modeled results were in good agreement with the observed results at a small average deviation of 4.52% and a *RMSE* of 5.16%. Fig. S4 (Supplementary material) further supports this argument by highlighting that the modeled results after calibration were closer to the observed results with a slope of 0.86 for the linear regression line, as compared to a linear slope of 0.19 for the modeled results before calibration. The calibrated model was further tested with 4 validation runs. As depicted in Fig. 2, the modeled results agreed reasonably well with the observed results with an overall *RMSE* of 11.03%. The individual *RMSE* for each of the validation run were 8.28, 7.82, 8.56, and 16.74%, respectively.

The calibrated parameters are listed in Table 2. It can be seen that the value of m an n are all far > 1, indicating that temperature increase had a positive effect on both the movement of agents and the probability of reaction occurrence. This was in accordance with Jing et al. (2014a) that increasing temperature from 23 to 40 °C seemed to stimulate the removal of naphthalene by exciting the collision between photons and molecules. The final values of most p were lower than the midpoint (i.e., 50) of their allowable ranges, especially for p_1, p_4, p_6, p_{10} and p_{11} that represented the excitation of NAP and HS, and the degradation of NAP and product. This finding suggested that, from a probability point of view, two different agents moving to the same position did not necessarily lead to reactions. Such an interesting finding could be caused by the settings of the maximum number of simulation ticks and the fact that the number of photons and oxygens were greater than the numbers of NAP and HS. Within 240 simulation ticks, the number of meetings between photons/oxygens and NAP/HS was greater than what would be required to degrade a limited amount of NAP. Therefore, degradation had to be somehow suppressed in order to fit the experimental results. As a direct competition, the absorption of photon by HS (p_6) appeared to be less favored than that by NAP (p_1). This may be interpreted as an outcome of the greater number of HS since more HS* may lead to a large number of singlet oxygen for NAP consumption, which needs to be suppressed as stated above. Interestingly, p_{11} was far greater than p_{10}, implying that the generation of multiple products was more competitive in order to increase the consumption of singlet oxygen. And p_9 was the only one that was higher than the midpoint and close to the upper bound, implying that the occurrence of HS degradation was not only fairly important (8.1% overall effect in Table 2) but also more favored to balance the

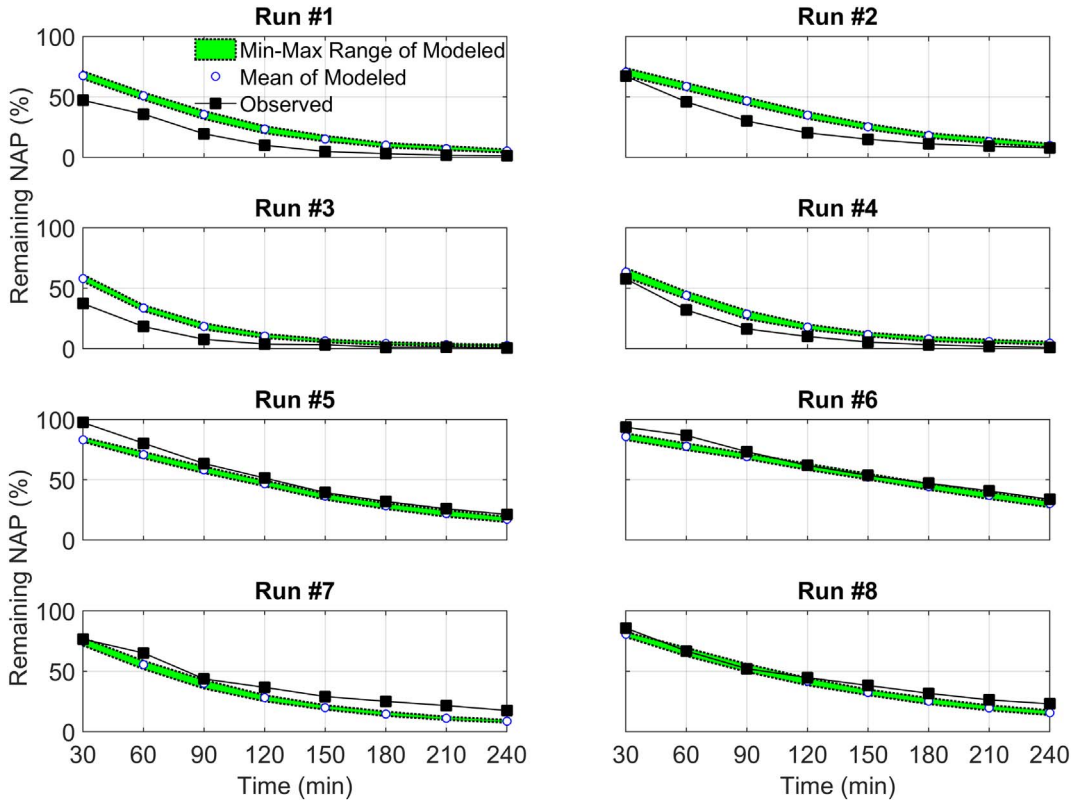

Fig. 1. Comparison between modeled and observed remaining NAP (%) for all 8 calibration runs after calibration.

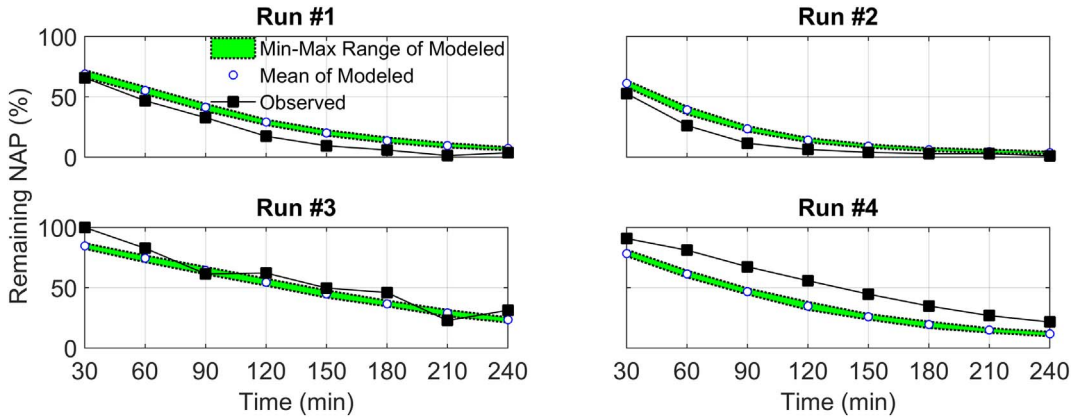

Fig. 2. Comparison between modeled and observed remaining NAP (%) for all 4 validation runs.

concentration of $^1O_2$. It also indicated that the concentration of HS was a driving factor in the photodegradation of naphthalene, given its well-known filter effect of UV light at 254 nm with extinction coefficient around 3.15 (L m^{-1} mg^{-1}) as reported by Kong et al. (2016).

Fig. S5 in Supplementary material depicts the modeling results obtained from using the calibrated parameters (Table 2) and the initial settings of calibration run #4 (Table S1). The visualization of the simulation environment (Figs. S5a–S5h) shows the variation of the numbers of selected agents during the entire simulation period. It can be seen from Fig. S5i and j that the numbers of NAP and HS kept decreasing while the number of products maintained a steady increasing trend. Interestingly, the numbers of reactive species NAP* and HS* were kept at a low level, while the number of singlet oxygen first increased and then started to decrease, possibly due to the increasing number of reaction products. Such a trend agrees well with the literature, in spite of a much coarser time-scale in this study (Alarcón et al., 2010; Pyryaeva et al., 2014).

3.3. Spatial distributions of agents

The proposed agent-based simulation model can also report the spatial distribution of any type of agent at selected time points. Fig. 3 plots the interpolated spatial distributions of NAP at eight different time points for calibration run #4 (Table S1). The overall shapes of the distributions in all subplots seem, to some extent, similar but not identical throughout the time points evaluated. This can be explained by the fact that the moving distance (D) of all agents was set at 5 pixels/tick, which was significantly smaller than the side length of the square simulation environment (33 patches × 13 pixels/patch = 429 pixels per side). Therefore, the moving range of agents including NAP was limited. Nonetheless, it can be seen from the subplots that as simulation time progresses, the total number of NAP gradually decreases. At 30 min, there are multiple areas that have 10–15 NAP, whereas at 240 min, the highest density of NAP revolves around 4–5. It should also be noted that, due to the two-dimensional projection of the reactor chamber and the random initial agent positions, the spatial distribution of NAP appears to be somewhat inhomogeneous and slightly different from the well-mixed condition. However, such difference can be accounted for by readjusting the position of each agent after each simulation tick. If the experimental conditions were more heterogeneous (e.g., non-symmetrical or non-uniform UV fluence rate, gas/liquid injection, and multi-layer non-symmetric reactor) or the distributions and

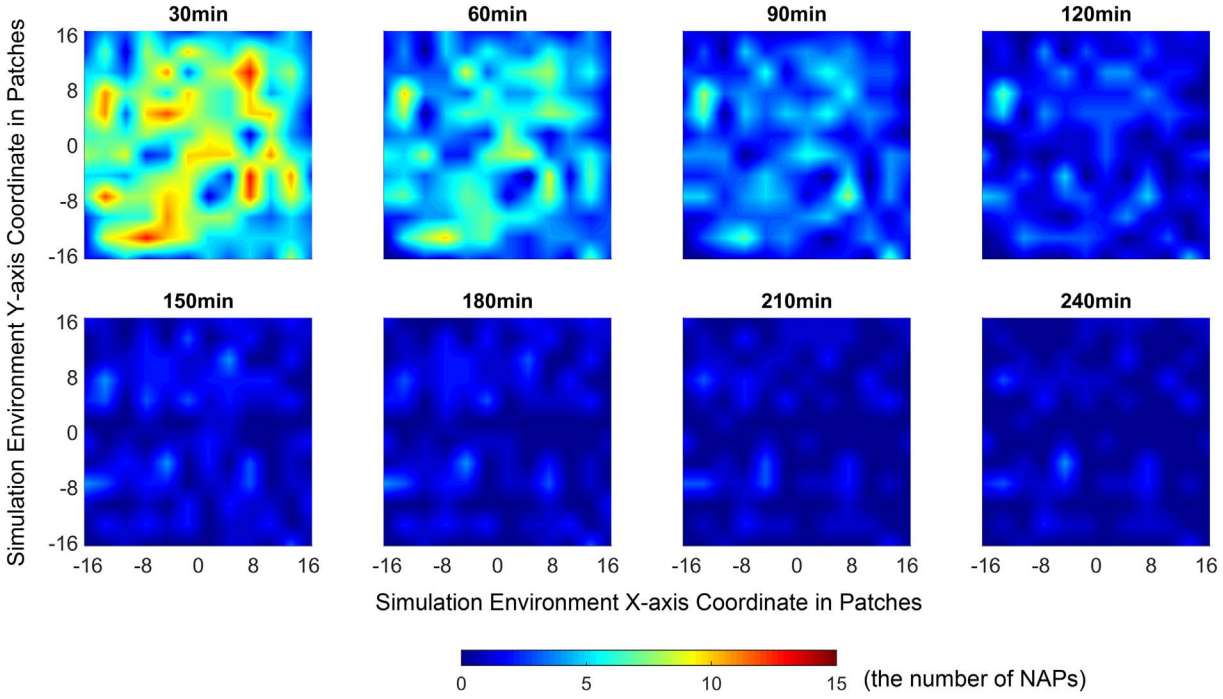

Fig. 3. The spatial distribution of NAP at different time points (based on calibration run #4).

movement of agents were constrained, then the agent-based modeling may be carried out in a 3D NetLogo environment and the results may be more valuable for reactor design and optimization.

3.4. Advantages and limitations of the agent-based modeling approach

The proposed agent-based modeling approach is an intuitive tool bridging the void between kinetic models depicted by reaction pathways and artificial intelligence-based models depicted by a black-box nature where no physical insight is available. By setting the behavior rules for each type of agent, the interactions among individuals and with their surrounding environment can be successfully captured to understand how their autonomous decisions and competition could affect the population-level outcomes. The use of global sensitivity analysis and model calibration can help model development by identifying parameters that contributed significantly to variations and improving modeling accuracy, respectively. In this study, the proposed approach was able to accurately predict the UV-induced degradation of NAP in marine oily wastewater. It could simultaneously track the populations and locations of multiple types of agent including contaminants, radicals, and products in a pre-defined simulation environment. The use of probability-based reaction pathways can be a good approximation of reaction kinetics with controllable reaction rates. On the other hand, it has adaptive learning ability and good generalization property for unseen data. Its performance (*RMSE* of 8.73% after calibration) was comparable to that of artificial neural networks (*RMSE* of 6.68% adopted from Jing et al., 2014b). The proposed agent-based modeling approach can also provide users more physical insight through a graphical user interface, especially the spatial distributions of selected agents that can be used for reactor system design and process analysis.

In spite of its technical advantages, however, there are several other important factors that need to be considered for further improvement. Firstly, the partnership between two different agents moving to the same position was evaluated one by one following the reaction sequence (i.e., Eqs.1–11). Competition rules may be applied here when numerous agents move to the same position. In other words, preferences may be assigned to partnerships between certain types of agent (e.g., known high reaction rates) such that their combinations may be more likely than other low priority partnerships. Secondly, the occurrence of a particular reaction was determined by a probability-based interpretation as a lumped approximation of its reaction rates under different conditions. Such an approximation can remarkably simplify the modeling process but at the expense of less physical insight. To remediate this, the relationships among reaction rates may either be sampled from given probability distributions or incorporated into the proposed modeling approach as calibration constraints (e.g., $p_1 > p_2$ because the rate of the first reaction is generally higher). Thirdly, the initial numbers of some agents (e.g., oxygen, photon and HS) were arbitrarily set in order to reduce computational burden. Further optimization may be carried out to find the most appropriate initial settings on how many molecules should be clustered as one agent. Fourthly, in this study, the number of NAP was measured and modeled while the numbers of many other substances were not able to be verified. By experimentally measuring more substances (e.g., different intermediates and products) and adopting multi-objective GA in model calibration, the modeling accuracy and applicability may be further improved. Last but not the least, agent behavior rules were fixed in this study. However, they can be set as time- or position-dependent during the simulation in order to accommodate the complexity of the treatment system.

4. Conclusions

This study shed light on the modeling of wastewater treatment processes using a novel agent-based probabilistic approach. Most existing techniques focus on population-level dynamics rather than individual-level behaviors, which may impede the understanding of heterogeneity within populations, reaction competition, and complex interactions among individuals. The use of agent-based modeling may help overcome such barriers by capturing the autonomous and adaptive nature of individuals and then exploring how the population would change across time and space. The developed approach was tested for its applicability in modeling the removal of naphthalene from marine oily wastewater by UV irradiation.

The removal of naphthalene was described by an agent-based simulation model using 8 types of agents and 11 reactions. Each reaction was governed by a probability parameter to determine its occurrence. Based on the available experimental results from 8 calibration runs and baseline model settings, a global sensitivity analysis was conducted to identify the most influential parameters. The modeling accuracy in terms of root mean square error (*RMSE*) was then reduced from 25.44 to 8.73% after a genetic algorithm-based calibration. Temperature was found to have positive impact on the movement of agents and the probability of reactions. All probability parameters were lower than their upper bounds, suggesting that two agents moving to the same position may not lead to reactions. The calibrated model was further tested with 4 validation runs with a *RMSE* of 11.03%. In addition, the spatial distributions of agents can be predicted at any given time point and may be used to complement computational fluid dynamics tools for reactor design and process optimization purposes. In spite of its limitations on reaction mechanisms and computational demands, the proposed agent-based modeling approach was able to introduce a novel way of describing wastewater treatment processes by investigating how individual behavior can affect population dynamics. It has a high potential to be applied, with certain modifications, to other environmental research areas, such as water treatment, biosurfactants production, groundwater remediation and environmental emergency responses.

Acknowledgements

Special thanks go to Natural Sciences and Engineering Research Council of Canada (NSERC), Research & Development Corporation of Newfoundland and Labrador (RDC NL), Canada Foundation for Innovation (CFI), Canada-China Scholars Exchange Program (CCSEP), and Foreign Affairs, Trade and Development Canada for supporting this research.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.marpolbul.2017.12.004.

References

Alarcón, E., Edwards, A.M., Aspee, A., Moran, F.E., Borsarelli, C.D., Lissi, E.A., Gonzalez-Nilo, D., Poblete, H., Scaiano, J.C., 2010. Photophysics and photochemistry of dyes bound to human serum albumin are determined by the dye localization. Photochem. Photobiol. Sci. 9 (1), 93–102.

Albuquerque, M.G., Carvalho, G., Kragelund, C., Silva, A.F., Crespo, M.T.B., Reis, M.A., Nielsen, P.H., 2013. Link between microbial composition and carbon substrate-uptake preferences in a PHA-storing community. ISME J. 7 (1), 1–12.

Arad, J., Housh, M., Perelman, L., Ostfeld, A., 2013. A dynamic thresholds scheme for contaminant event detection in water distribution systems. Water Res. 47 (5), 1899–1908.

Bayrakceken, F., Yegin, K., Korkmaz, E., Bakis, Y., Unal, B., 2012. Optical Energy Transfer Mechanisms: From Naphthalene to Biacetyl in Liquids and from Pyrazine to Biacetyl. Int. J. Photoenergy 2012, 239027.

Bonabeau, E., 2002. Agent-based modeling: Methods and techniques for simulating human systems. Proc. Natl. Acad. Sci. 99 (Suppl. 3), 7280–7287.

Bucci, V., Majed, N., Hellweger, F.L., Gu, A.Z., 2012. Heterogeneity of intracellular polymer storage states in enhanced biological phosphorus removal (EBPR)–observation and modeling. Environ. Sci. Technol. 46 (6), 3244–3252.

Crooks, A., Castle, C., Batty, M., 2008. Key challenges in agent-based modelling for geo-spatial simulation. Comput. Environ. Urban. Syst. 32 (6), 417–430.

Eglal, M.M., Ramamurthy, A.S., 2015. Competitive adsorption and oxidation behavior of heavy metals on nZVI coated with TEOS. Water Environ. Res. 87 (11), 2018–2026.

Esser, D.S., Leveau, J.H., Meyer, K.M., 2015. Modeling microbial growth and dynamics. Appl. Microbiol. Biotechnol. 99 (21), 8831–8846.
Fang, C., Lou, X., Huang, Y., Feng, M., Wang, Z., Liu, J., 2017. Monochlorophenols degradation by UV/persulfate is immune to the presence of chloride: illusion or reality? Chem. Eng. J. 323, 124–133.
Farmer, J.D., Foley, D., 2009. The economy needs agent-based modelling. Nature 460 (7256), 685–686.
Filatova, T., Verburg, P.H., Parker, D.C., Stannard, C.A., 2013. Spatial agent-based models for socio-ecological systems: challenges and prospects. Environ. Model. Softw. 45, 1–7.
Gernaey, K.V., van Loosdrecht, M.C., Henze, M., Lind, M., Jørgensen, S.B., 2004. Activated sludge wastewater treatment plant modelling and simulation: state of the art. Environ. Model. Softw. 19 (9), 763–783.
Haaker, M.P.R., Verheijen, P.J.T., 2004. Local and global sensitivity analysis for a reactor design with parameter uncertainty. Chem. Eng. Res. Des. 82 (5), 591–598.
Hellweger, F.L., 2007. Is it time to abandon the chemistry approach to biogeochemistry? (agent-based water quality modeling). Proc. Water Environ. Fed. 2007 (12), 5646–5665.
Hellweger, F.L., Bucci, V., 2009. A bunch of tiny individuals—individual-based modeling for microbes. Ecol. Model. 220 (1), 8–22.
Hellweger, F.L., Kianirad, E., 2007. Accounting for intrapopulation variability in biogeochemical models using agent-based methods. Environ. Sci. Technol. 41 (8), 2855–2860.
Ho, N.M., Thoai, N., Wong, W.F., 2015. Multi-agent simulation on multiple GPUs. Simul. Model. Pract. Theory 57, 118–132.
Jing, L., Chen, B., Zhang, B.Y., Zheng, J.S., Liu, B., 2014a. Naphthalene degradation in seawater by UV irradiation: the effects of fluence rate, salinity, temperature and initial concentration. Mar. Pollut. Bull. 81, 149–156.
Jing, L., Chen, B., Zhang, B.Y., 2014b. Modeling of UV-induced photodegradation of naphthalene in marine oily wastewater by artificial neural networks. Water Air Soil Pollut. 225 (4), 1–14.
Jing, L., Chen, B., Zhang, B.Y., Li, P., 2015. Process simulation and dynamic control for marine oily wastewater treatment using UV irradiation. Water Res. 81, 101–112.
Jing, L., Chen, B., Zhang, B.Y., Li, P., 2016. An integrated simulation-based process control and operation planning (IS-PCOP) system for marine oily wastewater management. J. Environ. Inform. 28 (2), 126–134.
Klabunde, A., Willekens, F., 2016. Decision-making in agent-based models of migration: state of the art and challenges. Eur. J. Popul. 32 (1), 73–97.
Kong, X., Jiang, J., Ma, J., Yang, Y., Liu, W., Liu, Y., 2016. Degradation of atrazine by UV/chlorine: efficiency, influencing factors, and products. Water Res. 90, 15–23.
Lair, A., Ferronato, C., Chovelon, J.M., Herrmann, J.M., 2008. Naphthalene degradation in water by heterogeneous photocatalysis: an investigation of the influence of inorganic anions. J. Photochem. Photobiol. A Chem. 193 (2), 193–203.
Matthews, R.B., Gilbert, N.G., Roach, A., Polhill, J.G., Gotts, N.M., 2007. Agent-based land-use models: a review of applications. Landsc. Ecol. 22 (10), 1447–1459.
Mbamba, C.K., Tait, S., Flores-Alsina, X., Batstone, D.J., 2015. A systematic study of multiple minerals precipitation modelling in wastewater treatment. Water Res. 85, 359–370.
Miller, J.S., Olejnik, D., 2001. Photolysis of polycyclic aromatic hydrocarbons in water. Water Res. 35 (1), 233–243.
Müller, B., Bohn, F., Dreßler, G., Groeneveld, J., Klassert, C., Martin, R., Schlüter, M., Schulze, J., Weise, H., Schwarz, N., 2013. Describing human decisions in agent-based models–ODD + D, an extension of the ODD protocol. Environ. Model. Softw. 48, 37–48.
Pereda, M., Zamarreno, J.M., 2012. An OOP agent-based model for the activated studge process using MATLAB. In: Distributed Computing and Artificial Intelligence. Springer, Berlin Heidelberg, pp. 241–248.
Pogson, M., Smallwood, R., Qwarnstrom, E., Holcombe, M., 2006. Formal agent-based modelling of intracellular chemical interactions. Biosystems 85 (1), 37–45.
Pyryaeva, A.P., Goldort, V.G., Kochubei, S.A., Baklanov, A.V., 2014. Singlet oxygen $O_2(a^1\Delta_g)$ formation via UV-excitation of isoprene-oxygen C_5H_8–O_2 encounter complexes in gas phase. Chem. Phys. Lett. 610, 8–13.
Roberts, A.P., Alloy, M.M., Oris, J.T., 2017. Review of the photo-induced toxicity of environmental contaminants. Comp. Biochem. Physiol., Part C: Toxicol. Pharmacol. 191, 160–167.
Santoro, D., Crapulli, F., Raisee, M., Raspa, G., Haas, C.N., 2015. Nondeterministic computational fluid dynamics modeling of *Escherichia coli* inactivation by peracetic acid in municipal wastewater contact tanks. Environ. Sci. Technol. 49 (12), 7265–7275.
Schreinemachers, P., Berger, T., 2011. An agent-based simulation model of human–environment interactions in agricultural systems. Environ. Model. Softw. 26 (7), 845–859.
Schuler, A.J., 2005. Diversity matters: dynamic simulation of distributed bacterial states in suspended growth biological wastewater treatment systems. Biotechnol. Bioeng. 91 (1), 62–74.
Schuler, A.J., 2006. Distributed microbial state effects on competition in enhanced biological phosphorus removal systems. Water Sci. Technol. 54 (1), 199–207.
Schuler, A.J., Majed, N., Bucci, V., Hellweger, F.L., Tu, Y., Gu, A.Z., 2011. Is the whole the sum of its parts? Agent-basedmodelling of wastewater treatment systems. Water Sci. Technol. 63 (8), 1590–1598.
Sweda, T.M., Klabjan, D., 2014. Agent-based information system for electric vehicle charging infrastructure deployment. J. Infrastruct. Syst. 21 (2), 04014043.
Vanuytrecht, E., Raes, D., Willems, P., 2014. Global sensitivity analysis of yield output from the water productivity model. Environ. Model. Softw. 51, 323–332.
Wang, R., Ren, D., Xia, S., Zhang, Y., Zhao, J., 2009. Photocatalytic degradation of Bisphenol A (BPA) using immobilized TiO2 and UV illumination in a horizontal circulating bed photocatalytic reactor (HCBPR). J. Hazard. Mater. 169 (1), 926–932.
Wilensky, U., Evanston, I., 1999. NetLogo: Center for Connected Learning and Computer-Based Modeling. Northwestern University, Evanston, IL, pp. 49–52.
Wilensky, U., Rand, W., 2015. An Introduction to Agent-Based Modeling: Modeling Natural, Social, and Engineered Complex Systems with NetLogo. MIT Press.
Wols, B.A., Harmsen, D.J.H., Wanders-Dijk, J., Beerendonk, E.F., Hofman-Caris, C.H.M., 2015. Degradation of pharmaceuticals in UV (LP)/H_2O_2 reactors simulated by means of kinetic modeling and computational fluid dynamics (CFD). Water Res. 75, 11–24.
Xavier, J.B., De Kreuk, M.K., Picioreanu, C., van Loosdrecht, M.C., 2007. Multi-scale individual-based model of microbial and bioconversion dynamics in aerobic granular sludge. Environ. Sci. Technol. 41 (18), 6410–6417.
Xu, C., Gertner, G.Z., 2008. A general first-order global sensitivity analysis method. Reliab. Eng. Syst. Saf. 93 (7), 1060–1071.