Microalgae and cyanobacteria modeling in water resource recovery facilities: A critical review
B.D. Shoener, S.M. Schramm, F. Béline, B.O. Martínez, Benedek G.Plosz, Spencer Snowling, Jean-Philippe Steyer, Borja Valverde-Pérez, Dorottya Wagner, Jeremy S. Guest

To cite this version:

HAL Id: hal-02268495
https://hal.archives-ouvertes.fr/hal-02268495
Submitted on 21 Aug 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Microalgae and cyanobacteria modeling in water resource recovery facilities: A critical review

Brian D. Shoener a, Stephanie M. Schramm a, Fabrice Béline b, Olivier Bernard c, Carlos Martínez c, Benedek G. Plósz d, Spencer Snowling e, Jean-Philippe Steyer f, Borja Valverde-Pérez b, Dorothy Wagner b, Jeremy S. Guest a,*

a Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, 205 N. Mathews Avenue, Urbana, IL 61801, USA
b IRSTEA, UR OPALE, F-35044, Rennes, France
c Université Côte d’Azur, INRA, Biocore, 2004, Route des Lucioles – BP 93, 06 902, Sophia Antipolis Cedex, France
d Department of Chemical Engineering, University of Bath, Claverton Down, Bath, BA2 7AY, UK
e Hydromantis Environmental Software Solutions, Inc., 407 King Street West, Hamilton, Ontario, L8P 1B5, Canada
f LBE, Univ. Montpellier, INRA, 102 Avenue des Etangs, 11100, Narbonne, France
g Department of Environmental Engineering, Technical Univ. of Denmark, Bygningstorvet, Building 115, 2800, Kgs. Lyngby, Denmark
h Chemistry and Bioscience, Aalborg University, Fredrik Bajers Vej 71H, 9220, Aalborg East, Denmark

A R T I C L E   I N F O

Article history:
Received 5 October 2018
Received in revised form 19 December 2018
Accepted 20 December 2018
Available online 28 December 2018

Keywords:
Growth
Nutrient uptake
Lipid storage
Starch storage
Wastewater treatment plant (WWTP)

A B S T R A C T

Microalgal and cyanobacterial resource recovery systems could significantly advance nutrient recovery from wastewater by achieving effluent nitrogen (N) and phosphorus (P) levels below the current limit of technology. The successful implementation of phytoplankton, however, requires the formulation of process models that balance fidelity and simplicity to accurately simulate dynamic performance in response to environmental conditions. This work synthesizes the range of model structures that have been leveraged for algae and cyanobacteria modeling and core model features that are required to enable reliable process modeling in the context of water resource recovery facilities. Results from an extensive literature review of over 300 published phytoplankton models are presented, with particular attention to similarities with and differences from existing strategies to model chemotrophic wastewater treatment processes (e.g., via the Activated Sludge Models, ASMs). Building on published process models, the core requirements of a model structure for algal and cyanobacterial processes are presented, including detailed recommendations for the prediction of growth (under phototrophic, heterotrophic, and mixotrophic conditions), nutrient uptake, carbon uptake and storage, and respiration.

© 2018 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Introduction

Nutrient removal requirements for water resource recovery facilities (WRRFs) are nearing the limit of current technologies (e.g., the limit of biological nutrient removal (BNR) is roughly 3 mg N L⁻¹ for total nitrogen and 0.1 mg P L⁻¹ for total phosphorus (USEPA, 2007)). As effluent requirements become more stringent, removal of both nitrogen and phosphorus past the current limit of technology requires the development of new technologies capable of reliably scavenging all forms of nutrients, including dissolved organic nitrogen and dissolved organic phosphorus (Bott and Parker, 2011). Microalgal resource recovery systems could significantly advance nutrient management of wastewaters by simultaneously achieving effluent concentrations of nitrogen and phosphorus below the current limit of technology and allowing for nutrient reuse (e.g., as fertilizer (Leow et al., 2015; Metting, 1996)). Although technical and economical bottlenecks still exist, the broad and sustained adoption of algal and cyanobacterial treatment processes is contingent upon the ability to reliably and accurately simulate full-scale performance in response to reactor and process design, influent composition, and environmental conditions. This ability is hindered by a lack of model fidelity and transparency regarding model structure and underlying science.

Some microalgae and cyanobacteria have the ability to utilize phototrophic, heterotrophic, or mixotrophic (i.e., phototrophic and heterotrophic simultaneously) metabolisms (e.g., Chlorella vulgaris...
A lack of clear direction for this suppositions that diminish model accuracy. These contrasting models of heterotrophic bacteria) or (ii) incorrect simplifying assumptions that diminish model accuracy. These contrasting approaches have resulted in hundreds of models for algae, indicating a lack of clear direction for this field.

Initial modeling efforts sought to understand phytoplankton behavior in natural ecosystems (e.g., (Jørgensen, 1976; Steele, 1962)), but translation of empirically derived models from nature to engineered systems requires verification and possibly modification. Additionally, disparate approaches to algal and cyanobacterial process modeling, highly variable experimental conditions (for model calibration and validation), and a lack of regard for existing chemotrophic model structures (e.g., ASMs, ADM1) have also impeded the development of generalizable model structures and well-defined parameters relevant to WRRFs (a.k.a. wastewater treatment plants, WWTPs). Recent review articles have summarized the breadth of models available to simulate algal growth (e.g., (Darvehei et al., 2018; Lee et al., 2015)), but there is no clear indication of when subcomponents should be considered or excluded (e.g., simulate organic carbon uptake or not) nor is there a rationale or guidance for choosing any particular equation to simulate each subcomponent. Recent process models developed by the authors (Baroukh et al., 2014; Guest et al., 2013; Wagner et al., 2016) have attempted to reconcile these differences, but an industry-wide, harmonized consensus is still lacking. To advance the broader implementation of algae and cyanobacteria process models by researchers and practitioners, it is critical to establish a unified modeling framework that is capable of accounting for relevant process and environmental conditions while simultaneously avoiding unnecessary complexity.

The objective of this work was to critically review approaches to algae and cyanobacteria modeling and propose a unified framework for phytoplankton process modeling in the context of WRRFs. As researchers attempt to balance model complexity with accuracy, the range of disparate phytoplankton wastewater treatment models continues to grow. To gain a better understanding of current approaches to modeling, a critical literature review was performed to characterize core components of modeling algal and cyanobacterial bioprocesses and elucidate their relative importance to the overall accuracy and complexity of wastewater models. Based on the available information, a modeling framework is proposed that can be used for future research and development in order to advance phytoplankton model fidelity and transparency as well as allow for its integration with current International Water Association (IWA) models (e.g., (Batstone et al., 2002; Henze et al., 2000)). This work synthesizes the findings and recommendations from an international collaboration of phytoplankton modelers working toward the development of a unified modeling framework for microalgal and cyanobacterial process models. Building on recent process models developed by the authors and on recent reviews, the results from an extensive critical literature review of 324 articles and conference proceedings presenting algal/cyanobacteria models is presented to identify state variables and processes that can serve as the core, unified modeling framework for phytoplankton-based bioprocesses.

**Methods**

A comprehensive review of algal and cyanobacterial modeling literature was conducted through Scopus based on the presence of search terms in the title, abstract, or keywords of research articles. The search terms for this review utilized “wildcards” to efficiently search for multiple variants of a word at once (e.g., model, modeling, and models are all found using the term “model” as well as a proximity search to ensure the word “model” was within 10 words of “grow” or “metabol” (i.e., “w/10”). The specific search used was: “title-abs-key(alga* or cyanobact* or phytoplank*) and (grow* or metabol*) w/10 model*” which yielded 2,402 research articles on January 26, 2018. Each paper was then screened to determine if it met any of the following exclusion criteria: (i) it did not model growth, (ii) it did not pertain to cyanobacteria or eukaryotic algae, (iii) there was no new or updated model presented, (iv) the model presented was a simple regression of (v) the paper was a review. Following screening, citations as well as citing papers were examined for each included paper in order to capture any research articles that may have been excluded from the Scopus search; these papers were included if they did not meet any of the exclusion criteria. The literature review yielded a total of 324 articles and conference proceedings that met the inclusion criteria (i.e., 2,078 did not meet inclusion criteria; a full list of the included papers can be found in Listing S1 of the supplementary information).
components were extracted and classified based on the processes being simulated — including the process rate equations for growth, nutrient uptake, and storage — as well as state variables (e.g., inorganic carbon, ammonium, nitrate, phosphate), main forcing variables (incident irradiance, temperature, background turbidity), and an external factor (e.g., organic carbon, nutrients, oxygen, stored carbon), as well as state variables (e.g., nutrient concentrations). Metabolic simulation considered (phototrophic, heterotrophic, and/or mixotrophic) metabolism, inclusion of photoacclimation, and how the photosynthesis-irradiance (PI) relationship was modeled. Finally, approaches to explicitly model pH, irradiance within the reactor, gas transfer, and other supporting processes were also evaluated to identify paths forward.

Mechanistic modeling of phytoplankton

Energy sources

Photoautotrophic growth of microalgae and cyanobacteria is the most frequently modeled metabolism (included in 93% of articles). Broadly speaking, phytoplankton photosynthesize CO2 into organic carbon using the energy garnered from light (Blankenship, 2002). Given this ability to convert inorganic carbon to organic carbon, phytoplankton are considered to be primary producers (Falkowski and Raven, 2007). As a result of their dependence on light, accurately simulating lighting conditions (e.g., continuous vs. diurnal, light intensity) and the response of phytoplankton to light (e.g., increased/decreased growth rate) is of utmost importance, which is rarely included as a consideration in other WRRF processes. In addition to light and inorganic carbon, photoautotrophic growth requires nutrients (namely nitrogen and phosphorus). In the absence of nutrients, algae produce storage compounds that can later be metabolized once nutrients are available (Guest et al., 2013). Discussions on how to model light, carbon, and nutrients can be found below.

Mixotrophic conditions (Bose and Chakraborty, 2016) are modeled in the same way as photoautotrophic growth and consumed during heterotrophic growth. The simplest way to simulate growth is with a logistic model (Verhulst, 1838; Table 1), which does not explicitly represent the interactions between an organism and its internal or external environment; growth is solely dependent on population. However, growth of microorganisms has long been observed to be affected by environmental conditions surrounding the cell, especially regarding substrate and nutrient concentrations (Monod, 1949). When modeling either of these growth regimes, the implications on dissolved oxygen (DO) need to be considered (e.g., produced during photoautotrophic growth and consumed during heterotrophic growth).

General model structures

When attempting to simulate growth in silico, models must possess adequate accuracy to represent the processes that are occurring while simultaneously minimizing complexity. The simplest way to simulate growth is with a logistic model (Verhulst, 1838; Table 1), which does not explicitly represent the interactions between an organism and its internal or external environment; growth is solely dependent on population. However, growth of microorganisms has long been observed to be affected by environmental conditions surrounding the cell, especially regarding substrate and nutrient concentrations (Monod, 1949). Though using the logistic model is not computationally intensive, its insensitivity to environmental conditions limits applicability to wastewater treatment. Nonetheless, 5.6% of papers reviewed used this model. The empirically-based Monod model (Monod, 1949; Table 1) — which is structurally the same as the mathematically-derived Michaelis-Menten model (Michaelis and Menten, 1913) — approximates kinetics as a hyperbolic relationship between growth and an external factor (e.g., organic carbon, nutrients, oxygen, and vice-versa (Nieva and Valiente, 1996). While mixotrophic growth does increase the number of uncertain parameters requiring calibration, this metabolism has the potential to improve algal productivity at WRRFs, lowering effluent nutrient concentrations and decreasing costs. Though interest in these growth conditions has increased, only 9.6% of models reviewed considered non-photoautotrophic metabolism. Initial heterotrophic and mixotrophic modeling efforts were developed to describe the production of specific molecules (e.g., astaxanthin (Zhang et al., 1998), phycocyanin (Zhang et al., 1998), and lutein (Zhang et al., 1999b)) or simulate high density monocultures (e.g., Haematococcus placentis (Bose and Chakraborty, 2016)). When modeling either of these growth regimes, the implications on dissolved oxygen (DO) need to be considered (e.g., produced during photoautotrophic growth and consumed during heterotrophic growth).

Fig. 1. Energy and carbon sources that are used by algae in each of the three metabolisms. Numbers in parentheses are manuscripts in the literature review that utilized that metabolism. Specific citations can be found in Listing S1 of the SI.
irradiance). This relationship is defined by the maximum specific growth rate and half-saturation constant (i.e., the concentration at which the growth rate is half the maximum rate; Fig. 2). The Monod model is widely used in the wastewater industry because it can be calibrated easily (due to having two parameters), closely approximates reality for many chemotrophic prokaryotes, and is accurate across a wide range of conditions for bacteria and archaea when substrate concentrations are below inhibitory levels. As a result, Monod has been adapted for use in the ASM models (Henze et al., 2000), ADM1 (Batstone et al., 2002), and number of custom models embedded in simulation platforms (e.g., GPS-X™ (Hydromantis, 2017), BioWin (Envirosim, 2017), and Sumo® (Dynamita, 2017)); of the phytoplankton models reviewed, 48% used a Monod equation.

A layer of complexity was added to the substrate-growth relationship when microalgae were observed to exhibit a notable lag between nutrient uptake and growth (Ketchum, 1939), suggesting these two processes may be partially decoupled (Droop, 2009, 1968). Further, microalgae were found to take-up and store nutrients (notably nitrogen and phosphorus) in excess of what was needed for growth to occur. Correlating the growth rate to internal substrate/nutrient content — the amount of which is determined by uptake and consumption — allows this model to decouple nutrient uptake and growth (Cunningham and Maas, 1978; Droop, 2009; Flynn, 2008; Stevenson et al., 1996; Sunda et al., 2009). Though this formulation is more appropriate for modeling phytoplankton growth, only 19% of the articles reviewed used a Droop equation.

Excessively high nutrient (e.g., ammonia) or substrate (e.g., acetate) concentrations have, at times, been shown to decrease growth rates due to inhibition (e.g., through ammonia toxicity or increased maintenance energy requirements (Chen and Johns, 1996)). High oxygen concentrations can also inhibit photosynthesis (Blankenship, 2002). Component concentrations that could be inhibitory may be modeled via the empirical Haldane model for enzymatic reactions ([Haldane, 1930]; Table 1) — also referred to as the Andrews equation given that it was first used for growth of microorganisms based on observations by Andrews (1968) and confirmed for phytoplankton in laboratory experiments by (Alba, 1982) — which is formulated similarly to the Monod model but includes an inhibition parameter (12% of models reviewed; Fig. 2).

When light reaches the cell, phytoplankton can grow photoautotrophically as defined by the relationship between photosynthesis and irradiance (PI). The simplest way to achieve this is with a Monod expression (18% of articles; e.g., (Béchet et al., 2015; Concasa et al., 2012)). Slightly more intricate models have been developed and implemented for engineered systems, including the Poisson single-hit model (5.0%; (Sakshaug et al., 1989; Skjelbred et al., 2012)), the Smith model (16%; (Broekhuizen et al., 2012; Kenny and Flynn, 2016; Smith, 1936)), and the Jassby-Platt model (3.7%; (Breuer et al., 2015; Jassby and Platt, 1976; Van Wagenen et al.,

### Table 1

<table>
<thead>
<tr>
<th>Model</th>
<th>Equation</th>
<th>Parameters</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic</td>
<td>$\mu = \frac{m_{\text{max}}}{1 + \frac{X}{X_{\text{max}}}}$</td>
<td>$\mu =$ specific growth rate [d⁻¹]; $m_{\text{max}} =$ maximum specific growth rate [d⁻¹]; $X =$ biomass concentration [g m⁻³]; $X_{\text{max}} =$ maximum biomass concentration [g m⁻³]</td>
<td>Verhulst (1838)</td>
</tr>
<tr>
<td>Monod</td>
<td>$\mu = \frac{m_{\text{max}} S}{K_s + S}$</td>
<td>$\mu =$ specific growth rate [d⁻¹]; $m_{\text{max}} =$ maximum specific growth rate [d⁻¹]; $S =$ substrate concentration [g m⁻³]; $K_s =$ half-saturation constant [g m⁻³]</td>
<td>Monod (1949)</td>
</tr>
<tr>
<td>Droop</td>
<td>$\mu = \frac{m_{\text{max}} q}{q}$</td>
<td>$\mu =$ specific growth rate [d⁻¹]; $m_{\text{max}} =$ maximum specific growth rate [d⁻¹]; $q =$ subsistence quota [g cell⁻¹]; $Q =$ cell quota [g cell⁻¹]; $K_i =$ inhibition constant [g m⁻³]</td>
<td>Droop (1968)</td>
</tr>
<tr>
<td>Haldane/Andrews</td>
<td>$\mu = \frac{m_{\text{max}} S}{K_s + S + \frac{S}{Q}}$</td>
<td>$\mu =$ specific growth rate [d⁻¹]; $m_{\text{max}} =$ maximum specific growth rate [d⁻¹]; $S =$ substrate concentration [g m⁻³]; $K_s =$ half-saturation constant [g m⁻³]; $Q =$ cell quota [g cell⁻¹]; $K_i =$ inhibition constant [g m⁻³]</td>
<td>(Andrews, 1968; Haldane, 1930)</td>
</tr>
</tbody>
</table>

**Fig. 2.** Conceptual representation of three most common growth rate equations used in algae modeling detailing how external substrate concentrations influence growth rates. Numbers in parentheses are the number of manuscripts that used that equation to model growth. If an article used multiple formulations, all were counted. Specific citations can be found in Listing S1 of the SI. Parameter definitions can be found in Table 1.
I metabolize multiple forms of nutrients (e.g., NO₃⁻, NH₄⁺, N₂H₄, N₂O₃, and DON for nitrogen (Dortch, 1990; Flynn et al., 1997; Liu et al., 2012): PO₄³⁻ and DOP for phosphorus (Brown and Shilton, 2014; Liu et al., 2012)). The respective rates of N and P uptake are not independent: rather, algae can modify their N:P ratio via growth rate adjustments to model nutrient uptake, the model requires two calibrated parameters (rather than one when utilizing a yield coefficient). This equation can be simplified if nutrient concentration is several orders of magnitude greater than the half saturation coefficient (e.g., when growing cultures on WRRF sidestreams such as anaerobic digester centrate). In this case, the uptake rate for that nutrient will be so close to the maximum specific uptake rate that the half-saturation coefficient can be considered negligible, simplifying the model.

Following uptake, nitrogen and phosphorus are metabolized to form intermediate biomass compounds when carbon is available (Blankenship, 2002). When luxury uptake occurs, algae have been shown to store phosphorus as polyphosphate (Brown and Shilton, 2014). Table 2). While more complex than Monod, these models do not account for the potential inhibitory effects of prolonged light exposure. The Steele model (10%; Drewry et al., 2015; Gonçalves et al., 2016; Steele, 1962; Wagner et al., 2016) and the Eilers and Peeters model (11%; Guest et al., 2013; Ketheesan and Nirmalakhandan, 2013; Peeters and Eilers, 1978) — similar in structure to Andrews (1968) — are able to account for photoinhibition. A further discussion of photoinhibition is included below.

When formulating a growth equation, these models can be multiplied (e.g., Droop for phosphorus multiplied by Andrews for nitrogen. Multiplication was first postulated by Baule (1917) and used in 44% of papers) or combined through a threshold formulation, where growth rates are only impacted by the most limited nutrient or substrate (von Liebig, 1841); used in 11% of papers). However, the multiplicative approach is only applicable when model components (e.g., light, nutrients, carbon, etc.) are independent, indicating that multiplying model components is overused. Thus, the threshold formulation should be used to account for scenarios where growth is dependent on multiple components (e.g., limitation by N and P). More complex relationships must be accounted for when the two compounds interact nonlinearly in the cell. For example, a phosphorus limitation reduces the nitrogen uptake efficiency (Bougaran et al., 2010).

Nutrient uptake, storage, and mobilization

In scenarios where phytoplankton are used for secondary or tertiary treatment of municipal wastewaters (i.e., when treating waters with dilute nutrient concentrations), phytoplankton growth can be dependent on either external or internal nutrients. In terms of macronutrients, phytoplankton are unique in their ability to metabolize multiple forms of nutrients (e.g., NO₃⁻, NH₄⁺, NO₂⁻, and DON for nitrogen (Dortch, 1990; Flynn et al., 1997; Liu et al., 2012): PO₄³⁻ and DOP for phosphorus (Brown and Shilton, 2014; Liu et al., 2012)). The respective rates of N and P uptake are not independent: rather, algae can modify their N:P ratio via growth rate adjustments to model nutrient uptake, the model requires two calibrated parameters (rather than one when utilizing a yield coefficient). This equation can be simplified if nutrient concentration is several orders of magnitude greater than the half saturation coefficient (e.g., when growing cultures on WRRF sidestreams such as anaerobic digester centrate). In this case, the uptake rate for that nutrient will be so close to the maximum specific uptake rate that the half-saturation coefficient can be considered negligible, simplifying the model.

Following uptake, nitrogen and phosphorus are metabolized to form intermediate biomass compounds when carbon is available (Blankenship, 2002). When luxury uptake occurs, algae have been shown to store phosphorus as polyphosphate (Brown and Shilton, 2014). Table 2). While more complex than Monod, these models do not account for the potential inhibitory effects of prolonged light exposure. The Steele model (10%; Drewry et al., 2015; Gonçalves et al., 2016; Steele, 1962; Wagner et al., 2016) and the Eilers and Peeters model (11%; Guest et al., 2013; Ketheesan and Nirmalakhandan, 2013; Peeters and Eilers, 1978) — similar in structure to Andrews (1968) — are able to account for photoinhibition. A further discussion of photoinhibition is included below.

When formulating a growth equation, these models can be multiplied (e.g., Droop for phosphorus multiplied by Andrews for nitrogen. Multiplication was first postulated by Baule (1917) and used in 44% of papers) or combined through a threshold formulation, where growth rates are only impacted by the most limited nutrient or substrate (von Liebig, 1841); used in 11% of papers). However, the multiplicative approach is only applicable when model components (e.g., light, nutrients, carbon, etc.) are independent, indicating that multiplying model components is overused. Thus, the threshold formulation should be used to account for scenarios where growth is dependent on multiple components (e.g., limitation by N and P). More complex relationships must be accounted for when the two compounds interact nonlinearly in the cell. For example, a phosphorus limitation reduces the nitrogen uptake efficiency (Bougaran et al., 2010).

Table 2 Models of photosynthesis-irradiance response with associated equations, citations, and a list of parameters.

<table>
<thead>
<tr>
<th>Model</th>
<th>Equation</th>
<th>Parameters</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eilers and Peeters</td>
<td>$\mu = \frac{2 \cdot \mu_{\text{max}} (1 + \beta) \cdot I}{(1 + \beta)^2 + 2 \cdot \frac{L}{L_{\text{s}}} \cdot I + 1}$</td>
<td>$\mu = \text{specific growth rate [d}^{-1}]$</td>
<td>Peeters and Eilers (1978)</td>
</tr>
<tr>
<td></td>
<td>$I = \text{light intensity [\mu E m}^{-2} \text{ s}^{-1}]$</td>
<td>$I_{\text{max}} = \text{maximum specific growth rate [d}^{-1}]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$L_{\text{s}} = \text{optimum irradiance [\mu E m}^{-2} \text{ s}^{-1}]$</td>
<td>$\beta = \text{attenuation coefficient [-]}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$K_{I} = \text{half saturation intensity [\mu E m}^{-2} \text{ s}^{-1}]$</td>
<td>$K_{I} = \text{light saturation [\mu E m}^{-2} \text{ s}^{-1}]$</td>
<td>Monod (1949)</td>
</tr>
<tr>
<td>Platt and Jassby</td>
<td>$\mu = \mu_{\text{max}} \cdot \tanh \left( \frac{z_{1}}{z_{\text{max}}} \right)$</td>
<td>$\alpha = \text{initial slope of PI curve [m}^{2} \text{ \mu E}^{-1}]$</td>
<td>Jassby and Platt (1976)</td>
</tr>
<tr>
<td>Poisson single hit</td>
<td>$\mu = \mu_{\text{max}} \left( 1 - e^{-K_{I}} \right)$</td>
<td>$K_{I} = \text{light saturation [\mu E m}^{-2} \text{ s}^{-1}]$</td>
<td>Poisson (1837)</td>
</tr>
<tr>
<td>Smith</td>
<td>$\mu = \mu_{\text{max}} \cdot \frac{z_{1}}{\sqrt{z_{\text{max}}^{2} + (zd)^{2}}}$</td>
<td>$\alpha = \text{initial slope of PI curve [m}^{2} \text{ \mu E}^{-1} \text{ s}]$</td>
<td>Smith (1936)</td>
</tr>
<tr>
<td>Steele</td>
<td>$\mu = \mu_{\text{max}} \cdot z_{1} \cdot e^{(1 - \alpha z_{1})}$</td>
<td>$\alpha = \text{initial slope of PI curve [m}^{2} \text{ \mu E}^{-1}]$</td>
<td>Steele (1962)</td>
</tr>
</tbody>
</table>
Additionally, nitrogen may be stored either as intracellular pools (Coppens et al., 2014; Thoresen et al., 2012) or as amino acids, though not as readily as phosphorus (Elrifii and Turpin, 1985). These nutrient stores will then be utilized during nutrient deplete conditions, if carbon is available. As a result, careful consideration must be taken when deciding how to incorporate the effect of nutrients on algal metabolism.

Carbon uptake, storage, and mobilization

In addition to nutrients, the uptake of organic and/or inorganic carbon should also be considered carefully (included in 26% of articles). Carbon uptake is commonly modeled using Michaelis-Menten or Haldane kinetics, depending on whether or not the compound can be inhibitory at higher concentrations (Baroukh et al., 2017) and on the amount of existing internal carbon stores (Concas et al., 2016; Guest et al., 2013). Apart from uptake, carbon utilization is dependent on several interrelated factors that must be considered concurrently. For example, under low light or dark conditions, algae cannot photosynthesize enough to meet maintenance requirements, so they will consume either external organic carbon (i.e., mixotrophic or heterotrophic metabolism) or stored carbohydrates/lipids (Baroukh et al., 2017; Guest et al., 2013). As a result, carbon usage should be modeled with switching functions (introduced in ASM1 (Henze et al., 1987)) to account for these factors.

In many algal species (e.g., Chlorella sorokiniana, Chlamydomonas reinhardtii, and Scenedesmus obliquus (Guschina and Harwood, 2006; Harwood and Guschina, 2009; Subramanian et al., 2013; Vitova et al., 2015)), carbon storage compounds (e.g., carbohydrates and lipids) accumulate in the cell in response to nutrient-deplete conditions (Fig. 3). Broadly, these two compounds have been modeled in numerous ways, most consistently as a function of biomass concentration using a yield coefficient (4.6% of articles; e.g., (Dilschneider and Posten, 2013; Mohammad Mirzaie et al., 2016)). More complex models for these compounds involve switching functions dependent on nutrient availability as well as the relative concentrations of each storage compound; these models do not have a consistent structure (7.1% of articles (Baroukh et al., 2014; Guest et al., 2013; Maitre et al., 2011)). Given the relatively limited knowledge of storage compound dynamics in the timescale of dynamic wastewater processes, there is no clear model that is optimal, but in general, carbohydrates and lipids should both be modeled separately (2.8% model carbohydrates only, 8.7% model lipids only, 3.7% model both together as a generic storage compound, 1.2% model both separately). Additionally, carbohydrates typically are produced and consumed faster than lipids, but also reach the maximum internal concentration sooner (Davis et al., 2016; Laurens et al., 2014). Lipids, conversely, are produced at a slower rate, but accumulate to a greater extent in many species of algae (Davis et al., 2016; Laurens et al., 2014). As an added layer of complexity, biomass composition can change both over diel cycles and also in response to design decisions (e.g., SRT (Gardner-Dale et al., 2017)), necessitating more mechanistic modeling of these metabolic processes. Oxygen consumption should also be accounted for using Monod kinetics when modeling stored carbon usage. In addition to this, stored carbohydrates have been shown to be interconverted to lipids, further increasing the potential intricacy of this sub-process (Pick and Avidan, 2017). Given the complexity of modeling carbohydrate and lipid accumulation, carbon storage should only be considered if nutrient-deplete conditions are expected to exist or if the cells are exposed to fluctuating lighting conditions (e.g., diel, natural lighting) in the scenario being modeled.

Light

Irradiance and dissipation

Light is an energy source during phototrophic and mixotrophic algal growth for photosynthesis. Accurately predicting light distribution is essential for modeling algal growth and metabolism. Photosynthetically active radiation (PAR) – the spectral range of radiation that photosynthetic cells can use – occurs between wavelengths of 400-750 nm (Wilhelm and Jakob, 2011). Three types of irradiance models were identified in (Bechet et al., 2013): type I models which use incident or average irradiance to predict the rate of photosynthesis for the entire culture, type II models which estimate overall reactor productivity as the sum of depth-resolved productivities within the system, and type III models which account for both light gradients and short light cycles for individual cells. Type I models are simplest, but type II models tend to exhibit greater accuracy for a small increase in complexity. Type III models are likely to be too complex for WRRFs given the need for individual-based (a.k.a. microscopic) models (Gujer, 2002). As light penetrates a phytoplankton cultivation system, it can be attenuated through absorption or scattering by cells or by the reactor itself (Posten, 2009; Wagner et al., 2018; Wang et al., 2014); Fig. 4). Light extinction can be exacerbated by high concentrations of highly diffusive particulate matter (Borowitzka, 1998). The Beer-Lambert law (Table 3) is widely used in phytoplankton modeling (35% of articles) and accounts for light extinction due to absorption by pigments and scattering by the cells as well as absorption and diffusion due to non-cellular components (Martinez et al., 2018). Often, the extinction coefficient is taken as the absorption rate (Koller et al., 2017). For dilute cultures, the attenuation coefficient has been proposed to be expressed as a nonlinear function of absorption and scattering coefficients (Kirk, 1984; Morel, 1988). Modifications to this equation may be required to simulate the effects of multiple scattering, when applicable (e.g., (Tam and Zardecki, 1982)), or to account for the disparity in coefficient values between natural and engineered systems.

Photoinhibition and photoacclimation

When modeling the effects of irradiance in engineered systems, temporal changes in light intensity need to be considered,
Safafar et al., 2015; Vaquero et al., 2014; Xie et al., 2016). This weeks (Combe et al., 2015) and contributes to chlorophyll:carbon photoacclimation strategy takes place at a time-scale of days or mass ratio variations typically within a factor of five (Falkowski, 1983), which has been represented by a model relating the chlorophyll quota to the current light irradiance and temperature. More complex models exist (Flynn, 2001; García-Camacho et al., 2012; Nikolaou et al., 2016), but the computational intensity of these models relative to the increase in fidelity precludes their use in the context of modeling WRRFs. Accounting for photoacclimation in model development is crucial to represent successive phases with low turbidity (where cells are subjected to high light) and growth periods in a highly turbid medium (for which average light is low); only 6.8% of models reviewed included this process. Specific citations can be found in Listing S1 of the SI.

Temperature

The influence of temperature on process dynamics can be tantamount to that of light (Ras et al., 2013); 34% of articles especially in naturally lit scenarios which will follow a diel cycle. Some algae have developed the ability to adapt in response to changing lighting conditions (Carvalho et al., 2011). Photoinhibition (i.e., the reduction in growth due to exposure to excess light) can begin affecting growth within 1 min if irradiance is high enough (Béchet et al., 2013; Fig. 5). Under high light intensities, photoacclimation processes are used to mitigate photoinhibition, whereby the chlorophyll production is suppressed and carotenoids (i.e., photo-protective pigments) are synthesized (Aburai et al., 2015; Anning et al., 2000; García-Camacho et al., 2012; Guihéneuf and Stengel, 2015; Koller et al., 2017; MacIntyre et al., 2002; Safar et al., 2015; Vaquero et al., 2014; Xie et al., 2016). This photoacclimation strategy takes place at a time-scale of days or weeks (Combe et al., 2015) and contributes to chlorophyll:carbon mass ratio variations typically within a factor of five (Falkowski, 1983). (Anning et al., 2000) studied the PI response curves of microalgae cultures pre-acclimated at different light irradiances and showed 65% higher growth rates and higher chlorophyll content for the cells pre-acclimated in the dark. (Bouterfas et al., 2002) compared multiple PI models and found that the Eilers and Peeters model is the most accurate when simulating growth in response to high irradiance. This phenomenon is of utmost importance for wastewater treatment given that particulate and dissolved materials increase medium turbidity, reducing the light penetration and maintaining a low average light in the water. On top of this, pigment synthesis can be strongly impacted by nitrogen limitation as the pigment content is related to the protein content, which is itself related to the nitrogen status (represented by the quota q in

### Table 3

Model of light penetration with associated equation and a list of parameters.

<table>
<thead>
<tr>
<th>Model</th>
<th>Equation</th>
<th>Parameters</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beer-Lambert</td>
<td>$I = I_0 e^{-k_0 z (X+k_0)^{1/2}}$</td>
<td>$I$ = light intensity [$\mu$E $m^{-2}$ $s^{-1}$]</td>
<td>(Beer, 1852)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$z$ = depth [m]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$I_0$ = incident light intensity [$\mu$E $m^{-2}$ $s^{-1}$]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$X$ = background turbidity [g $m^{-3}$]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$k_0$ = attenuation coefficient [m$^2$ g$^{-1}$]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$k_0$ = attenuation coefficient [m$^2$ g$^{-1}$]</td>
<td></td>
</tr>
</tbody>
</table>

**Fig. 4.** Simulation of light penetration into reactor (lower graphic) using Beer-Lambert as a function of irradiance (upper graphic, y-axis) and time (x-axis). Light is simulated using a 14-hour sinusoidal wave. Green shading represents light in reactor (i.e., darker colors correspond to lower light). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

**Fig. 5.** Conceptual representation of photoacclimation. Numbers in parentheses are manuscripts that included this process. Specific citations can be found in Listing S1 of the SI.
Table 4). This response is de
(Grimaud et al., 2017)). The impact of temperature on metabolism can be represented in mechanistic models accurately predicting the temperature dynamics over the course of the day (Béchet et al., 2011, 2010), but its effect has commonly been neglected or minimized in biological models, often by using an Arrhenius formulation ((Arrhenius, 1889); 40% of models considering temperature; Fig. 6 and Table 4).

The response of growth rate to temperature has been proposed to follow an asymmetric curve for most microorganisms ((Rosso et al., 1993); 3.6% of articles that included temperature; Fig. 6 and Table 4). This response is defined by three cardinal temperatures: \( T_{\text{min}} \), the minimum temperature that will support growth; \( T_{\text{opt}} \), the optimum temperature for growth; and \( T_{\text{max}} \), the maximum temperature that will support growth; and \( T_{\text{opt}} \). The asymmetry of the growth curve results from differential effects on cellular physiology at temperatures lower or higher than \( T_{\text{opt}} \). At low temperatures, the rates of enzymatic biochemical reactions are affected (Blankenship, 2002). At high temperatures, structure and stability of some cellular components, such as key enzymes or membrane components (mainly lipids or proteins), are denatured (Ras et al., 2013; Salvucci and Crafts-Brandner, 2004). The consequences on cell growth rates at temperatures greater than \( T_{\text{max}} \) are considered to be zero. The deleterious effects of high temperature exposure are also temporally dependent, and the accumulation of damage at high temperature (Béchet et al., 2017; Holcomb et al., 1999). Given the potential impacts that temperature can have on biomass, a temperature model should be carefully chosen and calibrated (a more detailed review of temperature models can be found in (Grimaud et al., 2017)).

Respiration and maintenance

While the main focus of this critical review is on phytoplankton growth and related processes, carbon loss through endogenous respiration and maintenance should not be neglected when constructing a model (included in 57% of models reviewed). These processes are often considered to occur at a constant rate (e.g., (Decostere et al., 2013); 47% of the models that included respiration); this assumption is valid for maintenance energy requirements, but endogenous respiration is impacted by multiple factors (e.g., light, temperature, pH, etc. (Béchet et al., 2017; Ippoliti et al., 2016)). Maintenance energy is dependent on the ATP requirements of the organism and, as such, does not change (Beefitink et al., 1990). The ATP demand can be met through multiple routes, depending on the situation being modeled (i.e., if carbon storage products are being formed or not). For instance, cells can meet maintenance requirements first through storage products and only rely on endogenous respiration in the absence of these internal stores (Beefitink et al., 1990). In addition to reducing biomass concentrations, respiration can also lead to DO consumption, which can have an impact on any aerobic chemotrophic organisms in the system. As such, incorporation of maintenance and endogenous respiration (as one mechanism to meet maintenance ATP demands (Guest et al., 2013)) is advisable.

Lumped pathway metabolic modeling

Metabolic modeling uses biochemical data representing a species’ or community’s metabolism to develop stoichiometric parameters for cellular operations (e.g., yield coefficients for substrates; 3.4% of the papers utilized metabolic reconfigurations). Metabolic models can incorporate different energy sources (e.g., photosynthetic, mixotrophic, and heterotrophic growth (Jüneja et al., 2016; Yang et al., 2000)), different metabolic pathways leveraged under a given set of environmental conditions (e.g., nutrient replete and deplete), and the accumulation of storage compounds (Guest et al., 2013; Radakovits et al., 2012). Full metabolic models typically involve entire genome reconstructions, comprised of individual reactions (i.e., between 200-3,000) and metabolic flux analyses (Baroukh et al., 2015; Chang et al., 2011). Although metabolic reconstructions provide a very thorough view of phytoplankton metabolism, the complexity hinders process engineers from implementing metabolic models of phototrophic technologies at WRRFs. Several models build upon the concept of metabolic reconstructions while taking a simpler approach that may be considered more accessible and valuable for developing processing parameters at WRRFs through metabolic modeling. In WRRF modeling of chemotrophic processes (e.g., enhanced biological phosphorus removal, EBPR), the concept of “lumped pathway metabolic modeling” has been applied (e.g., (Smolders et al., 1994)), which groups reactions based on whether intermediate compounds accumulate. These grouped (or “lumped”) reactions are assumed to occur simultaneously and are modeled as a function of one parameter (Roels, 1983). As such, lumped pathway metabolic modeling has the potential to simplify metabolic reconstructions, given that they have on the order of 10 reaction equations as opposed to >100 equations present in some reconstructions (Filipe and Daigger, 1999; Guest et al., 2013; Fig. 7).

WRRF modeling platforms typically use empirical yield coefficients for the ASMs to help describe sludge accumulation within the system ((Figuerola-Torres et al., 2017; Henze et al., 2007); Fig. 7), with the general recommendation to conduct experiments to develop yield coefficients for a specific WRRF (Henze et al., 2007; Rieger et al., 2012). Any variation in yield coefficients between WRRFs is generally believed to be caused solely by environmental influences such as pH, temperature, and fluctuating substrate concentrations (Henze et al., 2007). However, yield coefficients have a foundation in the metabolic pathways an organism is using (e.g., glycolysis, pentose phosphate pathway, etc.), indicating that values for these coefficients can be theoretically grounded. If
metabolic pathways are conserved, derived theoretical values can then be used regardless of situation being modeled. Variability in terms of growth rate will still exist due to environmental conditions, but this variation can be accounted for in associated sub-processes (e.g., maintenance) rather than by yield coefficients. Lumped pathway models also ground many parameters (e.g., P/O ratio) in fundamental, well-established ranges while decreasing the number of unknown stoichiometric parameters in the model, thus increasing the model’s innate mechanistic friction, lowering the degrees of freedom, and simplifying the final model structure (Baroukh et al., 2014; Guest et al., 2013; van Aalst-van Leeuwen et al., 1997). While there are many benefits to using full metabolic models, the additional complexity often precludes their inclusion in algal models.

Inclusion of other organisms

In the context of wastewater treatment, maintaining a community of one type of organism is challenging, if not entirely unrealistic. In the context of algae and cyanobacteria, chemotrophic bacteria (e.g., heterotrophic bacteria, nitrifiers) and predatory organisms (e.g., zooplankton) can be present, which impact the dynamics of the entire system being modeled ((Mehrabadi et al., 2016; Wolf et al., 2007); other organisms were included in 21% of models). The relationship between heterotrophic bacteria and photoautotrophic phytoplankton has long been known to be symbiotic, where bacteria and phytoplankton exchange products: chemotrophic bacteria produce CO2 and consume DO while phytoplankton take-up CO2 and produce DO (Rich, 1963). This interaction can be particularly beneficial for photoautotrophic phytoplankton because high DO concentrations can inhibit photosynthesis (Kaplan and Reinhold, 1999). However, higher biomass concentrations (stemming from chemotrophic organism growth) can also limit photosynthesis by reducing light to suboptimal levels. Additionally, phytoplankton and chemotrophs (including autotrophs) will be competing for substrates (N, P, inorganic carbon, etc.). If the phytoplankton are utilizing mixotrophy or heterotrophy, they will be competing with the chemotrophic bacteria for organic resources, which can impact system performance (Grover, 2000; Thingstad et al., 1998). Predators and grazers commonly have much more detrimental impacts on algal performance, which is primarily a result of algal biomass being directly reduced (Baird et al., 2003). Depending on the conditions, predators can result in frequent turnover of algal communities, preventing the system from treating the water as anticipated (Sutherland et al., 2017). Whether including bacteria, predators, or both, considering other organisms can quickly increase the complexity of a model due to the need to account for all the parameters the organism can affect (e.g., oxygen, CO2, organic carbon, pH, light penetration, etc. (Solimeno et al., 2017; Zambrano et al., 2016)). However, if phytoplankton are in a closed system (e.g., a closed photobioreactor), invasive organisms can be assumed to be negligible if not present in the influent.

Physico-chemical processes

Gas-liquid mass transfer

Gas-liquid mass transfer was included in 17% of reviewed articles. The most important dissolved gases in phytoplankton processes are oxygen (O2), carbon dioxide (CO2), and ammonia (NH3); depending on their concentration, they can either promote or inhibit growth. Generally speaking, algae require 1.8–2.4 g CO2 for each gram of biomass grown, which results in a N demand of 0.02–0.25 g per gram of algae (given that the C:N ratio can vary from

![Table 4](image)

Models of temperature impacts on growth with associated equations and a list of parameters.

<table>
<thead>
<tr>
<th>Model</th>
<th>Equation</th>
<th>Parameters</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arrhenius</td>
<td>( \mu_{\text{max}}(T) = \mu_{\text{max}*} \frac{E_a}{RT} )</td>
<td>( E_a ) = activation energy [m^2 kg s^{-2}] ( k ) = Boltzmann constant [m^2 kg s^{-2} K^{-1}] ( T ) = absolute temperature [K]</td>
<td>(Arrhenius, 1889)</td>
</tr>
<tr>
<td>CTMI</td>
<td>( \mu_{\text{max}}(T) = \mu_{\text{max}^*} \frac{(T - T_{\text{max}})}{(T_{\text{opt}} - T_{\text{max}})} ) ( = \frac{1}{(T - T_{\text{min}})^2} ) | ( \frac{(T_{\text{opt}} - T_{\text{max}})(T_{\text{opt}} - T_{\text{min}})}{(T_{\text{opt}} - T_{\text{min}} - 2T)} )</td>
<td>( T_{\text{opt}} ) = optimum temperature [(^\circ)C] ( T_{\text{max}} ) = maximum temperature [(^\circ)C]</td>
<td>Rosso et al., (1993)</td>
</tr>
</tbody>
</table>

![Fig. 7](image)

Conceptual representation of model complexity by type. Empirical models convert substrate to biomass through the use of yield coefficients. Lumped pathway models use simplified metabolic reactions to simulate growth. Metabolic flux models track individual metabolites as they are consumed and converted in the cell. Numbers in parentheses are manuscripts that included this process. Specific citations can be found in Listing S1 of the SI.
roughly 2.6–32 g C g N⁻¹ (Geider and La Roche, 2002; Leow et al., 2015; Li et al., 2017)). Microalgae also produce approximately 1.5−1.92 g O₂ g biomass⁻¹ (Grobbleaer et al., 1988; Muñoz and Guieysse, 2006). Moreover, CO₂ and NH₃ can significantly impact pH (see below). CO₂ and NH₃ are also involved in bacterial processes and consequently play an important role in algal-bacterial interactions, necessitating their accurate simulation in WRRF process models (Solimeno et al., 2017, 2015).

In the case of gas-liquid system with a relatively dilute liquid phase, Henry’s law is generally used to describe the equilibrium relationship (Eq. 1 in Table 5). Due to the continuous production and/or consumption of gaseous components by biological processes, however, gas-liquid mass transfers occur continuously. The driving force of gas-liquid mass transfer is therefore the difference between the saturation concentration and the real concentration in the liquid phase (Eq. 2 in Table 5). The local mass transfer coefficient represents the resistance of the interface to transfer. For gases with low solubility (e.g., O₂ and CO₂), dissolution into the liquid phase is more challenging, while more soluble gases (e.g., NH₃) are more difficult to remove from solution. The mass transfer rate depends on gas and liquid physico-chemical properties, temperature, and turbulence of the medium.

Several models exist to describe this mass transfer rate, including Higbie’s penetration theory (Higbie, 1935) or the double layer model (Lewis and Whitman, 1924). However, both the local mass-transfer coefficient and the interfacial area require calibration, and determining these parameters is challenging. Therefore, the combination of both parameters (i.e., the local mass transfer coefficient and the interfacial area) is often considered; this combined coefficient is typically referred to as k_La,i. The prevalence of aeration in the activated sludge process has resulted in extensive in situ determination of the oxygen-specific k_La,i value (i.e., k_LaO₂) (Amaral et al., 2017; Kayser, 1979) as well as models to describe it (Gillot et al., 2005). Overall mass transfer coefficients for other gases with low solubility (e.g., CO₂) are usually calculated from the oxygen transfer rate using Higbie’s penetration theory (Eq. 3 in Table 5). In order to determine the effect that these gaseous compounds have on gas-liquid mass transfer should be considered when constructing a WRRF model.

**pH/acid-base equilibrium**

Acid-base equilibrium considers the change in concentration of one or several compounds of interest (e.g., CO₂,aq, HCO₃⁻, and CO₃²⁻) in aqueous phase and the subsequent effect it has on pH and therefore microalgal processes (Brönsted, 1923; Lowry, 1923; 13% of models). Similar to compounds involved in liquid-gas transfer, these compounds are continuously produced or consumed by biological processes. Contrary to liquid-gas transfer, however, acid-base equilibrium occurs rapidly compared to biological processes. As a result, equilibrium is often assumed to be established instantaneously (i.e., on the order of 10⁴ and 10⁵ d⁻¹ (Jupsin et al., 2003; Reichert et al., 2001; Solimeno et al., 2015)). The direct impact of pH on algal growth (included in 16% of models) is rather limited when pH is between 5–8.5, depending on the algal species and other operational parameters such as temperature (Benemann et al., 1987; Berge et al., 2012; Mayo, 1997). Algal growth can affect pH by consuming CO₂ during photosynthesis, resulting in a net increase in pH as protons are emitted to maintain chemical equilibrium (Berenguel et al., 2004). In addition to the bicarbonate system, the nitrogen species present in solution have an impact on pH. If ammonium is applied as the nitrogen source, pH drops due to the release of protons during assimilation; conversely, pH can rise when nitrate is used due to the consumption of protons needed to assimilate nitrate (Nguyen and Rittmann, 2015). Most phytoplankton models for municipal wastewater treatment exclude pH dependence for algal growth (Broekhuizen et al., 2012; James et al., 2013; Mayo, 1997), but pH indirectly affects algal growth by changing the relative concentrations of substrates such as CO₂ (Decoste et al., 2016; Wolf et al., 2007). Additionally, high pH (i.e., above 9) may induce inhibition of algal growth by free ammonia (Abelivich and Azov, 1976; Konig et al., 1987).

Some models simulate pH as a function of the bicarbonate system (Decoste et al., 2016; Fernández et al., 2010; Liehr et al., 1988), but the simplicity of these models results in inaccuracy when other pH-altering components are present. The River Water Quality Model 1 (RWQM1 (Reichert et al., 2001)) introduced a set of differential equations to model the acid-base equilibria for inorganic carbon, ammonia, and phosphate, as well as the water acid-base equilibrium. This approach was later modified to a system of differential (for biological process rates) and algebraic (for fast chemical reactions) equations (DAE), where pH is numerically estimated by closing the charge balance (i.e., mono-dimensional numerical methods) in the system (Broekhuizen et al., 2012; Gehring et al., 2010; Wolf et al., 2007). The application of DAE systems to predict pH is commonly used to model other wastewater systems, as it reduces the stiffness of the model compared to differential equation systems (Batstone et al., 2012, 2002; Hellenga et al., 1999). This approach has been improved by (Vangsgaard et al., 2013) by formulating a set of equations describing the mass balances and equilibrium equations of weak acids and bases and a charge balance, which are solved using a multidimensional Newton-Raphson method. During the last few years, the acid-base system has been better described by including the effect of ionic strength, ion pairing, and developing improved numeric solvers to ensure robustness in the calculations (Flores-Alsina et al., 2015; Lizarralde et al., 2015; Solon et al., 2015). Other common ways to estimate pH changes in water systems are by closing proton or alkalinity balances (Luff et al., 2001; Serralta et al., 2004).

### Table 5

<table>
<thead>
<tr>
<th>Equation Parameters</th>
<th>Eq. No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{hi} = \frac{C_{li,i}}{\rho_i}$</td>
<td>1</td>
</tr>
<tr>
<td>$p_i = k_{li,i} \cdot a \cdot (C_{li,i} - C_i)$</td>
<td>2</td>
</tr>
<tr>
<td>$k_{li,O_2} = \frac{D_{O_2}}{\alpha_i}$</td>
<td>3</td>
</tr>
</tbody>
</table>
Calibration/validation and uncertainty/sensitivity analyses

Whether constructing a new model or tailoring an existing one to a specific scenario, a model should always be calibrated and validated (to the degree possible) prior to use (69% of models were calibrated, but only 36% were validated). Calibration involves adjusting parameter values to fit the model to collected and reconciled data (i.e., data that have been cleaned and ensured are accurate (Hauduc et al., 2010)); validation ensures that the model also matches data from a different set of conditions (distinct from the calibration dataset) by comparing the calibrated model to another set of data (Riegert al., 2012). This step in building a model is important because it ensures that model outputs are both reliable and reproducible; in the context of wastewater treatment, a model that can accurately predict performance can mean the difference between meeting discharge requirements or not. Calibration of WRRF models is often performed manually, which frequently results in unreliable output values (Sin et al., 2005). Though calibration protocols have been proposed (e.g., BIOMATH (Vanrolleghem et al., 2003), STOWA (Hulsbeek et al., 2002), and the IWA Unified Protocol (Riegert et al., 2012)), there is no consensus on calibration methodology, making the comparison and evaluation of calibration results of models incredibly challenging. While calibrating all parameters is extremely difficult and prone to errors—particularly for complex models with many parameters (e.g., RWQM1 and ADM1 with 24 and 45 parameters, respectively)—most of the variability in a model is often due to a subset of parameters. This subset can be determined through a sensitivity analysis (Sin et al., 2005). Ensuring that these parameters are calibrated carefully is essential; other parameters do not need to be calibrated as carefully or standard values can be sourced from previously published articles. Model prediction quality can be evaluated in relation to quality of fit (e.g., through the use of Akaike or Bayesian Information Criterion (Grimaud et al., 2017)) or by examining the percentage of experimental data falling within the confidence interval of the model output (Ramin et al., 2017; Wagner et al., 2018).

Uncertainty and sensitivity analyses can be used in tandem to determine how parameter variations affect output variability as well as which parameters affect model outputs the most. Broadly, uncertainty analyses in WRRF modeling pass randomly chosen input values through the model and record the outputs to determine either the range of potential outputs or estimate the likelihood of meeting a target value (e.g., discharge requirements (Loucks et al., 2005)). Uncertainty analyses can be conducted with Monte Carlo simulation, wherein probability density functions are assigned to inputs and are randomly sampled tens of thousands of times (Saltelli et al., 2006; Sin et al., 2009). To reduce the computational burden of characterizing uncertainty around WRRF process model results, Latin hypercube sampling (LHS) can be used to reduce the required number of simulations in a Monte Carlo analysis by discretizing probability density functions into equal probability portions (Helton and Davis, 2003; McKay et al., 1979; Shoener et al., 2016).

Following the uncertainty analysis, model inputs and outputs can be compared to determine parameter sensitivity. Similar to calibration procedures, there is no universal sensitivity analysis procedure, though sensitivity analyses can be broadly categorized as local or global. Morris one-factor-at-a-time (Morris, 1991) and Markov-Chain Monte Carlo (Shari et al., 2014) are examples of global sensitivity analyses, whereas differential analysis (Brun et al., 2002) is local. Sensitivity metrics can be used to concisely compare sensitivities of model outputs to individual inputs (e.g., Spearman’s rank correlation coefficient (Iman and Conover, 1982; Marino et al., 2008; Wang et al., 2015)). When evaluating a model, the analysis chosen should depend on the core objective (e.g., reducing model structural uncertainty or input uncertainty) which should be explicitly stated when describing the model (only 30% of models utilized sensitivity analyses, 26% of which did not describe the procedure used). Rigorously performing uncertainty and sensitivity analyses will not only simplify model calibration and validation, but will also help to inform future experimentation by identifying critical information needs for model application and development. For instance, if a particular model output is very sensitive to a highly uncertain parameter, extensive experimentation could be conducted to ensure the parameter is accurately calibrated and the model can be validated.

Integration with existing WRRF models

Standard wastewater modeling nomenclature

As we seek to develop phytoplankton-based processes for wastewater treatment, the likelihood of broad adoption of developed models will be directly influenced by their alignment with existing WRRF modeling structure and nomenclature. To the degree possible, any phytoplankton models developed should be able to be integrated with the IWA’s Activated Sludge Models (e.g., ASM2d and ADM3) and Anaerobic Digestion Model 1 (ADM1). There may be circumstances or applications of phytoplankton modeling that necessitate or warrant alternative model structures, but it should be recognized that any deviations will reduce transparency and the likelihood of adoption by industry.

The original IWA ASM1 publication popularized a standard approach to naming the state variables upon which pseudo-mechanistic model structures are built (Henze et al., 1987). Each state variable defines one fundamental component around which a dynamic mass balance is determined. The ASM1 state variable structure allows for grouping of different sets of variables by their fundamental composition (e.g., chemical oxygen demand, nitrogen, phosphorus), and by the processes which act upon them (e.g., biomass state variables undergo growth and decay) through rate equations. In addition to establishing a straightforward nomenclature (discussed below), the ASM1 also used a Petersen matrix (Petersen, 1965), which concisely displays an entire model (both kinetics and stoichiometry) and helps ensure the maintenance of mass balances.

The original ASM1 (Henze et al., 1987) sets out a nomenclature that specifies attributes of the state variables, and allows straightforward grouping of the state variables based on inherent characteristics and the processes acting on them (46% of articles focusing solely on wastewater; 5.0% of all articles). A capital letter specifies the component category (“S” for soluble and “X” for slowly biodegradable or particulate) and subscripts are then used to uniquely identify the component. Subscripts may specify if a component is readily biodegradable (e.g., S3) or inert (e.g., X4) or can identify the short-form chemical composition (e.g., , for soluble nitrate and nitrite). Biomass components typically use the nomenclature X9 followed by a second subscript to distinguish between biomass types (e.g., X9H for heterotrophs, X9A for autotrophs). The subsequent IWA models ASM2 (Henze et al., 1995), ASM2d (Henze et al., 1999), ASM3 (Gujer et al., 1999), and ADM1 (Batstone et al., 2002) maintained this nomenclature structure, adhering to the S and X prefixes for soluble and particulate components. The ASM2 model report also explicitly states that the X components must be electrically neutral (no ionic charges) whereas the soluble components may carry charges.

Further model development by academic researchers and commercial simulation companies has generally followed the established IWA nomenclature, but diversity in model complexity and approach to naming state variables has led to some confusion and inconsistency across the wastewater modeling industry.
(Corominas et al., 2010) summarized many of the issues faced with conflicting variable names, including consistency of variable names and meanings (e.g., “b” has been used to stand for “biomass” or “biodegradable”), usage of upper or lower cases, and different units associated with the same variable (e.g., biomass has been represented by g COD m⁻³, g-C m⁻³, or mol-C m⁻³). Additionally, state variable definitions should incorporate the processes which act upon them; this aspect is often disregarded when identifying variables. (Corominas et al., 2010) proposed a structured methodology for setting names for state variables and model parameters, using a series of subscripts to represent biodegradability, organic/inorganic nature, organism name (if relevant) and any other distinguishing characteristics. This standardized nomenclature has not been broadly adopted, however. To the degree possible, developers of new models should also use existing ASM and ADM variables (e.g., XLI for stored lipids, XCH for stored carbohydrates (Guest et al., 2013)) and only focus on adapting variables that are not already represented (e.g., XA, for algae biomass). Ultimately, nomenclature congruent with IWA naming conventions must be adhered to if phytoplankton pseudo-mechanistic models are to be usable by the wastewater industry.

Integration approaches

In order to develop plant-wide models that include phytoplankton-based processes alongside activated sludge or anaerobic digestion processes, existing integration techniques for ASMs and ADM1 can be leveraged. At present, there are three main approaches to plant-wide dynamic WRRF modeling: (i) the interfaces approach (Nopens et al., 2009; Vanrolleghem et al., 2005; Volcke et al., 2006; Zaher et al., 2007); (ii) the standard supermodel approach (Jones et al., 2007; Seco et al., 2004); and (iii) the tailored supermodel approach (Grau et al., 2007). A summary of these approaches can be found in (Grau et al., 2009). A number of software packages are commercially available for whole-plant simulations (e.g., GPS-X™, BioWin, WEST™, and Sumo®), taking advantage of both the interface and supermodel (e.g., GPS-X™, BioWin) approaches for whole-plant modeling. Although the interface approach can use state variable transformers to add new unit processes to models, developing these transformers is typically difficult (Grau et al., 2009). The supermodel approach requires the development of new state variables and processes, but allows for greater flexibility in terms of model development and is therefore more frequently used in practice (e.g., used in GPS-X™, BioWin, and Sumo®). Additionally, when integrating models, a continuity check should be performed to avoid inconsistencies in mass balances (see (Rieger et al., 2012) for details).

Path forward

Model development

Algae modeling needs in wastewater treatment

Accurately modeling full-scale phytoplankton treatment systems — accounting for reactor design, influent composition, algal-bacterial interactions, and environmental conditions — is essential. Models should be able to accommodate different reactor types (e.g., HRAPs, raceway ponds, PBRs, etc. (Shoener et al., 2014)) as well as influents (e.g., raw sewage, primary or secondary effluent, etc.). In order for a model to be useful, an objective should be established prior to model construction (e.g., prediction, control, monitoring, etc.) which will help practitioners navigate tradeoffs between complexity and accuracy (Bernard et al., 2015). A model should also be as simple as possible while maintaining the ability to assist with decision-making (Daigger, 2011).

Key factors in model development

The number of disparate models uncovered during this review and the absence of critical model components (e.g., calibration and validation methodology) in many papers underscore the need for more rigorous phytoplankton model formulation procedures. Before constructing a model, researchers must first consider environmental and reactor conditions to determine which processes to include and which equations can best simulate that process (Table 6). This consideration is of particular importance given that several influential components are frequently omitted from published models (e.g., carbon, phosphorus, and temperature were only included in 27%, 26%, and 24% of reviewed articles, respectively). When considering growth, Monod kinetics are adequate in circumstances with stable nutrient conditions (e.g., chemostats with fixed environmental conditions) and substrate concentrations below inhibitory values. In configurations or process designs in which algae are exposed to fluctuating nutrient concentrations, Droop model formulation should be used to decouple nutrient uptake and growth. Additionally, high ammonia or substrate (e.g., acetate) concentrations that could be inhibitory may be modeled via Haldane/Andrews (Andrews, 1968; Haldane, 1930); similarly, inhibition from high irradiance can be modeled via the Eilers and Peeters or Steele expressions (Eilers and Peeters, 1988; Steele, 1962). Given that the choice of a PI sub-model will only have an impact at high light intensities (where photoinhibition and pho- toacclimation may occur (Eilers and Peeters, 1988)), if only non-photoinhibitory situations are being considered, the simplest model that can still accurately simulate the process may be chosen. For substrate uptake and utilization, given the dynamicity of WRRFs, organisms must be modeled with their history in mind. For example, if the organism was previously in N-replete conditions but then experiences N-deplete conditions, it will continue growing off internal reserves before forming storage compounds. Additionally, the impacts of changing environmental conditions on phytoplankton processes should be assessed (e.g., temperature changes across seasons); excluding environmental history or dynamic conditions will limit the accuracy of a model.

In addition to selecting model components carefully, the model foundation should be built upon a mechanistic understanding of phenomena. This will not only improve the generalizability and accuracy of the model, but may also reduce the number of calibrated parameters, provide a theoretically grounded range for biochemical model parameters, and increase model transparency by reducing the number of empirical parameters. This comprehensive understanding will also enable the integration of separate, but related models (e.g., hydrodynamics), which will be critical to translate laboratory-scale data into predictions of full-scale performance. Also, when gathering experimental data, it should be noted that diel cycles will often necessitate frequent sampling (multiple times per day) to enable model validation. High-resolution models require multiple measurements each day (or even each hour) in order to capture diel variability; this is crucial during model calibration and validation. More broadly, the lack of clearly defined calibration and validation procedures is a serious shortcoming in many of the papers reviewed (17% of papers that calibrated their models did not specify a procedure). Calibration and validation need to be conducted to assess the accuracy and precision of a model appropriately and ensure the fidelity of a model is upheld. If a model cannot be calibrated and validated reliably, it cannot be utilized for a full-scale reactor, let alone as a component of a WRRF model.

Broader context

WRRF practitioners are tasked with protecting the health of the
public and the aquatic environment, and an inaccurate model could have serious deleterious effects on WRRF investment, design, and performance. When proposing a model, the phytoplankton process should be considered as part of an entire WRRF, not just an independent unit. Promising model structures should be able to handle the frequent changes that occur at WRRFs. In order to ensure models are accessible to the broadest possible audience and can be compared easily, IWA nomenclature should be used more frequently to increase WRRF model inter-compatibility. The ASMs proposed the use of a stoichiometric matrix to quickly and efficiently display the model structure. However, only 6.2% of papers reviewed used a stoichiometric matrix. This approach should be used to make models simpler to interpret and to implement in simulation software, thus widening the user base of the models. Phycocyanin and cyanobacterial technologies have the potential to achieve effluent nutrient concentrations below existing biological nutrient removal systems, but there is no established modeling framework for engineered phytoplankton treatment systems.

In the context of wastewater treatment, a Droop formulation for nutrient uptake and an Eilers and Peeters formulation for irradiance response can accurately simulate external and internal conditions.

- The effects of temperature fluctuations on phytoplankton growth rates should be assessed with either an Arrhenius or CTMI formulation.
- Lumped pathway metabolic models rooted in a mechanistic understanding of biochemical processes can simplify model structure and reduce parameter uncertainty.
- Calibration and validation should be rigorous and clearly defined.

**Acknowledgements**

This work supported by the National Science Foundation Graduate Research Fellowship Program (GRFP) for B.S. and S.S., and the Paul L. Busch Award from the Water Environment & Reuse Foundation (now the Water Research Foundation) for J.G. The CONICYT doctoral grant supported C.M. The IPL Algae in Silico project supported C.M. and O.B. The Phycover Research Project (ANR-14-CE04-0011) funded by the French National Research Agency (ANR) and the French clusters Trimatec, Mer Bretagne Atlantique, and Mer Méditerranée supported O.B. and J.-P.S. The authors would also like to thank Elena Ficara and Francesca Casagl for their thoughtful comments during the drafting of this manuscript.

**Appendix A. Supplementary data**

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jwroa.2018.100024.

**References**


