Biotechnology Advances 85 (2025) 108709

Contents lists available at ScienceDirect

Biotechnology Advances

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

Modelling challenges to unlock the power of phototrophic systems for wastewater valorization

Francesca Casagli [a], Andrea Turolla [b,*], Damien J. Batstone [c], Gabriel Capson-Tojo [d], Elena Ficara [b], Joan García [e], Eva Gonzalez-Flo [f], Julien Laurent [g], Tatjana Lorenz [h], Michaël Pierrelée [i], Benedek Gy. Plósz [j,n], Gustavo Henrique Ribero Da Silva [k], Ángel Robles [l], Simone Rossi [b,o], Estel Rueda [e], Lars Stegemüller [m], Jean-Philippe Steyer [d], Olivier Bernard [a], Borja Valverde-Pérez [i]

[a] Centre INRIA d'Université Côte d'Azur, GreenOwl team, Sophia-Antipolis 06902, France
[b] Department of Civil and Environmental Engineering, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy
[c] Australian Centre for Water and Environmental Biotechnology, The University of Queensland, Brisbane 4072, QLD, Australia
[d] INRAE, Univ Montpellier, LBE, 102 Avenue des Étangs, Narbonne 11100, France
[e] GEMMA - Group of Environmental Engineering and Microbiology, Department of Civil and Environmental Engineering, Universitat Politècnica de Catalunya BarcelonaTech, c/ Jordi Girona 1-3, Building D1, E-08034 Barcelona, Spain
[f] GEMMA - Group of Environmental Engineering and Microbiology, Department of Civil and Environmental Engineering, Escola d'Enginyeria de Barcelona Est (EEBE), Universitat Politècnica de Catalunya·BarcelonaTech, Av. Eduard Maristany 16, Building C5.1, E-08019 Barcelona, Spain
[g] Université de Strasbourg, ENGEES, CNRS, ICube UMR 7357, F-67000 Strasbourg, France
[h] Department of Architecture and Civil Engineering, ZEuUS - Competence Centre for Sustainable Engineering and Environmental Systems, THM - University of Applied Sciences, Wiesenstr. 14, 35390 Giessen, Germany
[i] Department of Environmental and Resource Engineering, Technical University of Denmark, Bygningstorvet, Bygning 115, 2800 Kgs, Lyngby, Denmark
[j] Department of Chemical Engineering, University of Bath, Claverton Down, Bath BA2 7AY, UK
[k] Department of Civil and Environmental Engineering, School of Engineering, São Paulo State University (UNESP), Bauru, SP, Brazil
[l] Departament d'Enginyeria Química, ETSE-UV, Universitat de València, Avinguda de la Universitat s/n, Valencia, Burjassot 46100, Spain
[m] Department of Chemical Engineering, Technical University of Denmark, DTU, Søltofts Plads 228A, Lyngby 2800, Denmark
[n] SWING – Department of Built Environment, Oslo Metropolitan University, Postboks 4 St Olavs Plass, 0130 Oslo, Norway
[o] Energy Technologies and Processes Research Group, Generation Technologies and Materials Department, Ricerca sul Sistema Energetico - RSE S.p.A., v. Rubattino 54, 20134 Milano, Italy

ARTICLE INFO

Keywords:
Microalgae
Cyanobacteria
Purple phototrophic bacteria
Mechanistic modelling
Data-driven modelling
Digital twins
Wastewater treatment
Circular economy

ABSTRACT

Phototrophic microorganisms are gaining prominence for their dual role in wastewater treatment and resource recovery, converting wastewater into valuable bioproducts. However, their effective deployment needs robust modelling frameworks capable of predicting performance across complex, real-world scenarios. Despite significant advances, key challenges hinder the development and application of such models:

- Biological complexity: phototrophic systems involve intricate processes (e.g., photosynthesis, nutrient uptake, microbial interactions, and predation) that are difficult to represent accurately due to their dynamic interdependencies.
- Environmental variability: permanent fluctuations in light, temperature, pH, and toxic compounds in outdoor reactors require high-resolution dynamic data for reliable model calibration and prediction.
- Data limitations: lack of comprehensive, high-quality datasets (e.g., biological, environmental, and operational conditions) constrains model development, particularly for data-driven approaches.
- Multi-scale integration: bridging molecular, cellular, and ecosystem-level processes into a unified modelling framework, including physics, remains a significant hurdle.

* Corresponding author.
E-mail address: andrea.turolla@polimi.it (A. Turolla).

https://doi.org/10.1016/j.biotechadv.2025.108709
Received 3 April 2025; Received in revised form 3 September 2025; Accepted 5 September 2025
Available online 8 September 2025

- Parameter and uncertainty management: models often suffer from non-identifiable parameters, sensitivity to approximations, and insufficient validation against long-term experimental data.
- Balancing complexity and applicability: selecting the appropriate level of ecological and mathematical details, tailored to specific applications (e.g., biomass production and nutrient removal) and data availability is critical yet challenging.
- Computational and interdisciplinary barriers: high computational costs, especially for hybrid and data-driven models, alongside the need for cross-disciplinary collaboration, further complicate model development.
- To overcome these barriers, this work argues for standardized protocols in model design, calibration and validation, alongside enhanced data collection and reconciliation efforts. Integrating innovative approaches, such as metabolic modelling, machine learning and hybrid modelling into digital twins, will be essential to unlock the full potential of phototrophic systems, bridging the gap between theoretical models and industrial implementation.

1. Introduction: Why modelling complex phototrophic systems?

Phototrophic systems represent a sustainable biotechnological solution that make use of solar energy to remove contaminants from wastewaters while simultaneously recovering resources (Capson-Tojo et al., 2020; Oviedo et al., 2022; Mutale-Joan et al., 2023). These systems can convert inorganic and organic C, as well as nutrients (e.g., N and P), into valuable bio-products, including biofuels (e.g., methane, biodiesel), biofertilizers, microbial feeds or foods, or bioplastic precursors.

While all phototrophic microorganisms use light as an energy source, three main groups can be defined according to their physiological and metabolic traits: microalgae, cyanobacteria, and purple phototrophic bacteria (PPB). Microalgae (eukaryotes) and cyanobacteria (prokaryotes) perform oxygenic photosynthesis, using water as electron donor. In contrast, PPB (prokaryotes) carry out anoxygenic photosynthesis using diverse electrons donors (e.g., sulphide, simple organic compounds, or molecular hydrogen). These phototrophic microorganisms are mainly autotrophic, harvesting energy from sunlight and carbon from CO_2. Most of them can switch to a mixotrophic metabolism, using both light and simple organic compounds to get energy and carbon, and thus enabling simultaneous assimilation of organic pollutants and nutrients. PPB can use light as the energy source and organic compounds as both C and electron sources, with biomass yields up to the

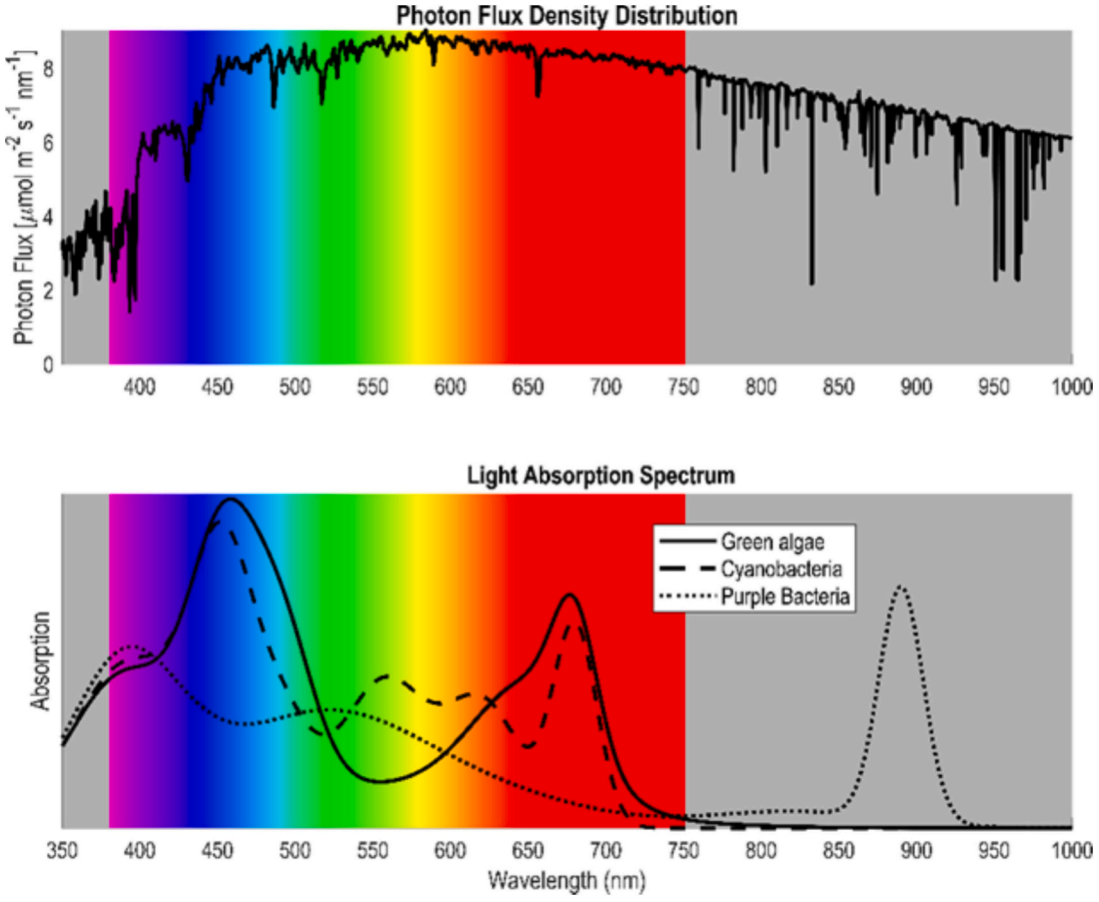

Fig. 1. Photon flux density distribution (under the atmosphere) within the 350–1000 nm range, along with the light absorption spectra of green algae (solid line) cyanobacteria (dashed line) and purple bacteria (dotted line) over the same wavelength range. The shaded grey area between 350 and 380 nm indicates the ultraviolet range, while the shaded grey region between 750 and 1000 nm represents the infrared range. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

maximum possible value during heterotrophy, of 1 g $COD_{biomass}$·$gCOD^{-1}_{removed}$ (COD being the chemical oxygen demand), when growing photoheterotrophically (Capson-Tojo et al., 2020). As illustrated in Fig. 1, microalgae and cyanobacteria primarily absorb visible light (400–700 nm; photosynthetically active radiation (PAR)) using chlorophyll *a* and b as main pigments. PPB instead rely mostly on bacteriochlorophylls (mainly a and b) to harvest near infra-red (NIR) light (850–1100 nm).

The microbial and metabolic diversity of phototrophic systems enhances their resilience and broadens their applicability but also makes them extremely complex to study and to optimize. In wastewater treatment systems, phototrophic microorganisms participate in intricate ecological interactions with diverse microbial communities, spanning mutualistic to parasitic relationships, which critically shape their collective physiology and metabolism (Amin et al., 2012). Phototrophic systems are built around the pivotal role of photosynthetic growth, but the metabolic versatility of these organisms must also be considered (e.g., mixotrophic growth or non-photosynthetic activity). Equally important are the interactions between photosynthetic microorganisms and other key microbes. Bacteria can stimulate growth of phototrophs by releasing CO_2, supplying essential nutrients (e.g., ammonium, vitamins, growth factors), regulating N fluxes, breaking down complex organic molecules, and producing extracellular polymeric substances that promote flocculation and reduce predation (Ramanan et al., 2016; Vulsteke et al., 2017). In systems based on oxygenic phototrophs (e.g., microalgae and cyanobacteria), a symbiotic relationship forms with nitrifying or heterotrophic bacteria, enabling effective wastewater treatment. Phototrophs produce O_2, which aerobic bacteria use as an electron acceptor to degrade organic matter. In return, bacteria produce the CO_2 that photoautotrophs need for growth (Oviedo et al., 2022). Anoxygenic PPB interact primarily with anaerobic bacteria, which hydrolyse and ferment complex organic compounds into volatile fatty acids (VFA), serving as substrates for PPB (Capson-Tojo et al., 2020). These interactions also include the exchange of other key metabolites and the mitigation of inhibitory conditions (Hays et al., 2017; Natrah et al., 2014; Palacios et al., 2022). These interactions are pivotal for effective wastewater treatment and bio-production, and must be properly understood and controlled.

Abiotic factors also play a crucial role in phototrophic systems. Sunlight-driven processes are highly sensitive to fluctuations in light and temperature, affecting microbial populations and metabolism, and system performance. This dependency on environmental parameters, combined with metabolic complexity and microbial interactions, makes phototrophic systems extremely challenging to control and optimize. Here, digital twins, defined as numerical tools based on integrated mathematical models (i.e., considering physical, chemical and biological dimensions) which are dynamically updated through automated links to the physical entity (Torfs et al., 2022), can play a transformative role, enabling the full exploitation of these processes via in-silico assessment, design, control, and optimization. However, digital twins face challenges. First, they depend on continuous, automated data flows from the physical system, making data availability critical for their successful application. Second, they rely on high-fidelity mathematical models to represent the underlying processes. These models are their core.

Models synthesize scientific knowledge to enhance understanding and enable predictive capabilities. They support supervision and optimization strategies for balancing production, environmental, and economic objectives, along with scenario analysis to bolster the competitiveness of phototrophic processes. Furthermore, modelling is a powerful tool to navigate the inherent complexity and variability of phototrophic systems, addressing key challenges in their scalability and operational stability:

- **Understanding the complexity of biological processes**: Phototrophic systems involve several intertwined biological processes like photosynthesis, nutrient and carbon storage, growth, and respiration, alongside interactions between phototrophic and non-phototrophic microorganisms. Models help to understand these mutual interactions.
- **Accounting for variable environmental conditions**: Light, pH, temperature, or toxic compounds vary widely across time and space, especially in outdoor reactors fed with wastewater. Models can capture this variability and its effects on phototrophic growth, which is essential for evaluating process performance.
- **Simultaneously managing the different time scales**: Processes involving phototrophs or mixed microbial communities span multiple temporal scales. Rapid dynamics occur over milliseconds (light harvesting) or seconds (e.g., chemical reactions) to days for microbial growth or decay. Critically, acclimation to environmental parameters (e.g., light intensity, temperature) operates on even longer timescales of several days. This inherent variability necessitates tailored modelling approaches to capture system behaviour accurately.
- **Simulating implementation scenarios and process optimization**: Validated models are invaluable tools for technology assessment and optimization, allowing for fast and cost-effective screening of multiple environmental and operational conditions, which would be prohibitive experimentally. Phototrophic technologies, still under development, would benefit greatly from this.
- **Advanced process monitoring and control:** Models support the design and development of software sensors (observers), closed-loop control strategies, fault detection systems, and diagnosis schemes. Despite being crucial for implementation, these aspects remain understudied in phototrophic systems while their interests have been demonstrated in many other applications involving biological processes (e.g., in pharmaceutical or in wastewater treatment fields).
- **System scaling-up**: Models enable simplified design procedures for large-scale systems, supporting the implementation and optimisation of complex biological processes.

To fully exploit the potential that digital twins can provide, key challenges must be overcome:

- **Increasing data availability and quality**: Acquiring high-quality, reliable environmental and operational data through both online (e.g., incident PAR/NIR irradiance, temperature profiles, dissolved oxygen, pH, nutrient levels) and offline measurements (e.g., COD concentrations, microbial diversity indices, photosynthetic activity, suspended solids) is critical for constructing robust models, particularly in data-driven approaches. These datasets must capture the biology of phototrophic organisms, their environmental conditions, and their dynamic interactions. These factors are often challenging to measure at spatiotemporal scales sufficient for predictive extrapolation. Furthermore, systematic data reconciliation remains a critical yet unresolved objective in modelling phototrophic systems, limiting their predictive accuracy and broader applicability.
- **Multiple scale integration**: Phototrophic systems can be modelled across multiple scales, ranging from molecular and cellular levels to full ecosystems, each offering unique insights. However, integrating these multi-scale models into a cohesive framework remains a major challenge, requiring innovative approaches to bridge disparities in resolution, dynamics, and mechanistic representation.
- **Uncertainty and sensitivity analysis**: The accuracy of model predictions can be degraded due to the approximations and assumptions supporting the modelling. Quantifying and managing uncertainty, and assessing parameter sensitivity, are essential but difficult.
- **Parameter estimation and identifiability**: Models often require many parameters, which must be estimated from experimental data. Accurately estimating these parameters can be difficult, and small errors can lead to significant deviations in model predictions.

- **Model validation**: Validating models against experimental data is essential to ensure their accuracy and reliability. However, validation is challenging due to the complexity of the systems and the lack of suitable and long-term validation data.
- **Balancing complexity and realism**: Selecting the optimal level of mathematical complexity, i.e. the appropriate representation of biological species and processes, depends on both contextual factors (e.g., climatology, wastewater composition, operational parameters) and modelling objectives (e.g., biomass yield, removal efficiency, emission assessment). Divergent modelling assumptions not only complicate the evaluation and adaptation of existing models for specific applications but also hinder the development of unified frameworks capable of performing robustly across diverse scenarios.
- **Computational demands**: Complex models, especially those that operate at fine temporal and spatial scales or integrate across scales, can be computationally intensive. Furthermore, the increasing application of data-driven or hybrid models, which involve artificial intelligence algorithms, face high computational cost for training on big data and for implementing backpropagation algorithms. Balancing accuracy with computational feasibility is a critical challenge.
- **Interdisciplinary integration**: Phototrophic modelling involves multiple disciplines, from biology, chemistry, physics, to ecology. Coordinating such interdisciplinary collaborations is a last but not least challenging aspect.

This article explores these different modelling challenges in detail and proposes guidelines to address them. It critically reviews current modelling approaches, contextualizing them according to modelling objectives. Research needs are discussed, highlighting the need for improved data collection and integration methods, and the importance of developing standardized computational methods for model development and application. Future directions are proposed for the development of accurate, scalable digital twins for phototrophic systems. The establishment of good modelling practices and standardized methodological recommendations represents a crucial but largely unexplored area in the context of phototrophic systems.

2. What are the key elements needed for modelling phototropic systems?

2.1. Focusing on system specifics and its function

Given the central role of photosynthesis in these systems, the effects of abiotic factors on the kinetics of photo-induced metabolic processes and microbial interactions must be understood. Light dependency is obviously essential, but temperature and gas-liquid mass transfer also matter, as they influence kinetics and CO_2 fixation and O_2 concentrations (Béchet et al., 2013; Batstone and Flores-Alsina, 2022). The impact of these environmental factors on process performance demands the consideration of reactor design and geometry, technological/engineering parameters (e.g., mixing), and system scale (Vulsteke et al., 2017; Capson-Tojo et al., 2022). Geometry affects the light path (and thus its attenuation) and gas transfer through surface-to-volume ratios. Effective mixing is critical to ensure homogeneous conditions, prevent localized substrate limitations, and avoid the accumulation of inhibitory compounds such as oxygen, which under high light intensity can trigger the formation of damaging reactive oxygen species (ROS). Scaling up introduces challenges, such as heterogeneities due to mixing difficulties, biomass aggregation or fouling, which may not appear in lab-scale setups.

Compared to conventional processes (i.e., activated sludge systems) and their respective models, phototrophic systems require different modelling choices and hypotheses. While standard models of the International Water Association (IWA), such as the Activated Sludge Models (ASMs; Henze et al., 2006) or the Anaerobic Digestion Model 1 (ADM1; Batstone et al., 2002) have been instrumental for bioprocess modelling, the unique characteristics of phototrophic systems demand distinct modelling strategies, which are described below.

The modelling strategy should be determined by the system's functional objectives. In wastewater treatment applications, model fidelity must prioritise accurate prediction of effluent quality descriptors. Conversely, for resource recovery and value-added product generation, the focus shifts towards biomass productivity and biochemical composition, necessitating enhanced metabolic detail to reliably forecast product yields and economic viability. Often, there are several objectives to be considered at the same time, leading to multi-objective optimisation. Modelling can facilitate multiple approaches instrumental to on-line process optimisation and operation, including predictive control systems, decision-support frameworks, operator training simulations, and real-time process monitoring solutions. The model must be carefully tailored to its intended purpose, accounting for the availability and temporal resolution of data, whilst remaining rigorously aligned with the system's core functionality.

The claimed advantages of photosynthetic systems for WWT compared to conventional systems include: i) reduced aeration and energy demand by oxygenation through photosynthesis (or fully for anaerobic phototrophs); ii) decreased C footprint; iii) cost-efficient treatment; and iv) enhanced circularity through biomass valorisation (Maurya et al., 2022). In practice, the last point is not only an advantage but a requirement, as these processes must produce valuable outputs to remain competitive. Common products that can be generated using phototrophs include pigments (e.g., astaxanthin), lipids (for biodiesel or as essential oils), carbohydrates, proteins, or polyhydroxyalkanoates (PHAs; plastic precursors) (Maurya et al., 2022; Capson-Tojo et al., 2020). The value of the biomass depends on its composition and, importantly, its applicability, which links directly to resource recovery potential within a circular economy. Microbial protein as feed is a perfect case in point: its use depends not only on quality but also on local/national legislations. Depending on the latter, feed use may be allowed only if proteins are derived from food-grade by-products, like cheese whey or slaughterhouse wastewater. Biomass derived from wastewaters that limit high-value applications can be valorized as products with lower value, such as fertilizers and soil amenders (Coppens et al., 2016). Therefore, to account for circular economy strategies generating value from waste, models must predict biomass composition and potential market value, while also assessing environmental impacts, making multi-objective optimization an essential tool.

It is unlikely that a single model can simultaneously and accurately represent all the factors and outputs in such complex systems. Model complexity must be therefore adapted to the amount of available information for calibration and validation. A model will be more efficient when it has the minimal complexity to solve most straightforwardly the problem for which it has been designed. In the following sub-sections, the most relevant factors that can (or cannot) be included in models for phototrophic systems are discussed depending on its goal, either wastewater remediation or resource recovery.

2.2. Predicting effluent quality for wastewater treatment

Effluent quality is a critical aspect of any mainstream WWT process, which must ensure that environmental standards for water discharge and/or reuse are met. Phototrophs play a significant role in improving effluent quality by removing main pollutants such as N and P, but also other contaminants such as pharmaceuticals and heavy metals (Richmond, 2004). For most phototrophs, ammoniacal nitrogen is the preferred N source (Perez-Garcia et al., 2011). pH controls the equilibrium between ammonium (NH_4^+) and free ammonia (NH_3), the latter being an important inhibitor of photosynthesis above certain concentrations. At pH values beyond 8.5 (resulting in more than 5 % of the total ammonia-N as NH_3), modelling NH_3 enhances model accuracy and improves bioreactor design (Rossi et al., 2020). P is an essential element for

the synthesis of nucleic acids, phospholipids and adenosine triphosphate (ATP; Singh et al., 2018). Some phototrophs might take up more P than needed by their metabolism, enhancing its removal via luxury uptake (Singh et al., 2018). The chemistry of P is very complex, and its availability can decrease due to precipitation of several complexes at high pH values (usually above 8). Some phototrophs store C, N and P beyond their metabolic needs (Droop, 1983), so their growth can still occur even after the complete depletion of one of these elements from the culture medium. In autotrophic cultures, C is often provided as CO_2 or from inorganic carbon in the water stream. Otherwise, many phototrophic microorganisms can also grow heterotrophically and mixotrophically, consuming simple organic carbon sources (acetate, glucose or glycerol). In addition, heterotrophic bacteria may play a key role hydrolysing complex organics (Baroukh et al., 2015; Mohamadnia et al., 2023; Capson-Tojo et al., 2020). Modelling C, N, and P flows is thus key to predict effluent quality.

2.3. Light

Light is a primary factor differentiating phototrophic systems from non-phototrophic processes. Outdoor reactors are the main growth strategy for phototrophs due to the high costs of artificial illumination. In these systems, light permanently fluctuates at a scale varying from seconds (clouds), hours (day-night cycles), and seasons. Light intensity directly influences the rates of photosynthetic processes and, consequently, the growth of phototrophs and their interactions with other organisms. Phototrophic wastewater treatment reactors are designed with shallow depths to ensure light penetration despite medium turbidity. The light intensity and spectral quality experienced by individual cells vary spatially within the reactor, depending on their position. Consequently, reactor geometry and hydrodynamics critically determine the dynamic light regime encountered by cells as they circulate through these light gradients. Additionally, for reactors inside greenhouses, 10 % of the light energy is lost by the greenhouse transparent film (Pessi et al., 2022). Accounting for light attenuation throughout the culture is even more crucial in PPB reactors, as NIR light is more significantly attenuated than visible wavelengths (i.e., PAR) (Capson-Tojo et al., 2022).

Reactions related to light absorption occur in very short time frames (less than milliseconds) and are classified as the light-dependent reactions of photosynthesis, while the light-independent enzymatic reactions, which are driving Calvin's cycle consuming CO_2, occur over much longer time scales (up to hours) (Williams and Laurens, 2010). Over billions of years, photosynthetic organisms have evolved sophisticated adaptive strategies to cope with solar flux variations spanning from growth-limiting to photoinhibitory intensities. These include protective mechanisms such as the synthesis of photoprotective pigments (e.g., carotenoids) and energy dissipation pathways to safely neutralize excess absorbed light energy. Another strategy for the cell to harvest light optimally is to adapt the pigment contents (and proportions) to light intensity (and spectrum). This adaptive capacity, termed photoacclimation, accounts for observed variations in cellular pigment concentrations spanning nearly an order of magnitude (MacIntyre et al., 2002). Under high irradiance, cells minimize their photosynthetic pigment content to decrease energy absorption, while maximizing pigment production under light-limited conditions. This acclimation process typically occurs over timescales of several days (Ross and Geider, 2009), reflecting the physiological trade-off between photoprotection and light-harvesting efficiency.

2.4. Temperature

Temperature is also a crucial factor for outdoor phototrophic cultures because its fluctuations directly impact both chemical and metabolic reaction rates. Specifically, temperature dynamics result from fluctuations of the solar flux in outdoor reactors. Working with low reactor depths (therefore with lower thermal inertia) and being exposed to variable solar fluxes, imply that temperature fluctuates at a much larger amplitude when compared to standard WWT processes. To illustrate, a standard 56 m^2 raceway pond in southern France exhibits diurnal temperature variations with amplitudes exceeding 15 °C (Casagli and Bernard, 2022). Closed photobioreactors face even greater thermal oscillations, due to amplified greenhouse effects and limited evaporative cooling (Endres et al., 2016). Flat-panel systems in Brisbane, Australia, reached diurnal variation amplitudes of up to 30 °C during summer (Hülsen et al., 2022). High temperatures can destroy the cellular structures, potentially leading to protein denaturation and cell mortality (Béchet et al., 2013; Serra-Maia et al., 2016). Mortality must be integrated into mathematical models especially when temperature goes beyond the optimal values (typically around 35 °C for microalgae, Ras et al., 2013).

2.5. Gas transfer

In phototrophic systems, continuous gas-liquid exchange occurs through stripping or dissolution, driven by gas production or consumption from biochemical and chemical processes. The mass transfer rate is governed by the physicochemical properties of the gases and liquid phase, temperature, and medium turbulence. Consequently, accounting for gas-liquid exchange dynamics is essential, as it directly influences remediation efficiency, biomass productivity, and gaseous emission profiles.

In microalgae-based systems, the dynamics of dissolved O_2 and CO_2 availability govern the synergistic interactions between algal and bacterial communities (Casagli et al., 2023). Accurate representation of these gas dynamics is essential for modelling both phototrophic (CO_2-dependent) and heterotrophic (O_2-dependent) growth processes. Diurnal solar cycles impose periodic fluctuations in O_2 production and CO_2 consumption rates, resulting from the balance between biological activity and atmospheric exchange.

Controlled injection of air and/or CO_2 is typically implemented to:

i) mitigate O_2 oversaturation, and.
ii) maintain pH within an optimal range (typically 7.5–8.5) where CO_2 is not limiting.

Notably, predicting CO_2 availability requires explicit pH modelling to determine the speciation of inorganic carbon, particularly the bioavailable CO_2 fraction. Monitoring dissolved O_2 concentrations is also key in open outdoor cultivation systems for PPB, as their phototrophic metabolism is inhibited by O_2. PPB then lose their main competitive advantage resulting in their out competition by heterotrophic aerobes (Capson-Tojo et al., 2021).

Other gases such as free ammonia (NH_3), molecular hydrogen (H_2) or nitrous oxide (N_2O) could also be relevant in WWT applications due to their significant environmental and operational impacts. N_2O is a powerful greenhouse gas, with a global warming potential approximately 300 times greater than that of CO_2. It can be produced during biological N removal through conventional nitrification and denitrification, or in some microalgal metabolic pathways (Domingo-Félez and Smets, 2019; Zhang et al., 2023). In phototrophic systems, inorganic C limitation by nitrifiers can trigger N_2O production (Peng et al., 2015; Mellbye et al., 2016; Casagli et al., 2021a, 2021b).

2.6. Resource recovery: Predicting biomass production and value

Predicting the growth of microorganisms and their productivity is crucial in phototrophic systems. The produced biomass offers a wide range of valorisation options, including the production of energy, biofertilizers, bioplastics, and other valuable bioproducts. Generating value is essential to make these systems competitive against traditional wastewater treatment processes. Biomass harvesting is a crucial step for

cost-effective resource recovery (Fang et al., 2016) but modelling of separation technologies is beyond the scope of this review.

2.6.1. Pigments

Phototrophs produce a wide range of pigments, including chlorophylls, carotenoids, and phycobilins, with potential applications in food, feed, cosmetics, and pharmaceuticals. Predicting pigment dynamics and productivity is however challenging. Cell pigments are strongly affected by light intensity and temperature for adapting to optimize light harvesting and for photoprotection. Reversely, their concentration affects light distribution in photobioreactors (Wágner et al., 2018) and therefore photosynthetic growth (Martínez et al., 2018; Cerruti et al., 2022).

2.6.2. Polyhydroxyalkanoates and glycogen

PHAs are a family of polyesters synthesized naturally by many different prokaryotic microorganisms and have similar properties to common thermoplastics. How PHA is produced by phototrophs depends on the type of metabolism (i.e., autotrophic vs. heterotrophic). Autotrophic production by cyanobacteria generally requires nutrient starvation (Anderson and Dawes, 1990; Wu et al., 2001), so cells start accumulating glycogen that, during prolonged chlorosis, is catabolized into PHA (Dutt and Srivastava, 2018; Koch and Forchhammer, 2021). Thus, modelling PHA accumulation from cyanobacteria should also consider the mechanism of intracellular carbon storage (Rueda and García, 2021). PPB can produce PHA when grown on VFA and nutrient limiting conditions, producing H_2 as by-product (Capson-Tojo et al., 2020). To enhance PHA accumulation and minimize H_2 production, NH_4^+ can be provided in excess, while the other nutrients could still be limiting. The addition of different types of VFA leads to different types of polymers (Bravo-Porras et al., 2024).

2.6.3. Lipids, carbohydrates and proteins

Phototrophs can be rich in proteins or essential fatty acids, which makes them appropriate for dietary supplements for feed and food (del Rosario Rodero et al., 2024; Mutale-Joan et al., 2023). Residual byproducts rich in nutrients can support the accumulation of microbial proteins. Food-grade streams are typically employed to eliminate potential health risks for end consumers, whether human or animal (e.g., Rasouli et al., 2018; Gorzelnik et al., 2023). Alternatively, if the treated streams contain micropollutants or pathogens, proteins, together with stored P and pigments, can be applied as organic fertilizers (Coppens et al., 2016). Furthermore, hydrolysed amino acids are useful as biostimulants (Romero-García et al., 2012).

The production of lipids and carbohydrates from phototrophs has been widely explored due to their promising application for bioenergy production as sustainable alternatives to conventional fossil fuels. Biomass with lipid storage above 40 % dry weight (DW) is suitable for biodiesel production, while for lower content, biogas is more appealing (Maurya et al., 2022). Indeed, nitrogen-rich microalgal biomass serves as an excellent feedstock for co-digestion with carbon-rich feedstocks, creating a balanced nutrient profile for enhanced anaerobic digestion performance (Wágner et al., 2016b). Several reviews (e.g., Lee et al., 2015; Morales et al., 2021) have explored the role of nutrient availability, light intensity, and temperature in triggering lipid accumulation over time. Carbohydrates, aside from being valuable products, are considered in some models as crucial carbon intermediates and temporary storage forms (Guest et al., 2013; Shoener et al., 2019), which are mobilized before lipid accumulation in phototrophic microorganisms.

3. Balancing empirical and mechanical approaches for modelling phototrophic systems

Various approaches exist to model nonlinear biological dynamics in phototrophic systems, ranging from purely empirical to fully mechanistic frameworks. However, most practical models lie somewhere between these extremes, combining mechanistic structure with empirical parameterizations. Fig. 2 illustrates this duality of empirical and mechanistic approaches, highlighting their interdependence, where empirical methods (experience → knowledge) and mechanistic frameworks (knowledge → explanation) synergistically compensate for their individual limitations. For instance, macroscopic mass-balance models (Section 3.1.2) include stoichiometric constraints (mechanistic) but rely on empirical bio-kinetic expressions (e.g., Monod for nutrient limitation, CTMI for temperature dependence). Metabolic models (Section 3.2), while fundamentally mechanistic since their stoichiometry is derived from genome-scale analyses, often require empirical constrains from omics data. Data-driven models (Section 3.3.1), while fundamentally empirical, often incorporate physics-inspired architectures or are post-processed with mechanistic rules. Hybrid models (Section 3.3.2) directly merge data-driven and first-principles components to exploit the strengths of both approaches (e.g., replacing kinetic terms with neural networks, physical informed neural networks (PINNs) enforcing conservation laws).

Table 1 summarizes the key features, advantages, and challenges of these modelling paradigms, along with their applications to phototrophic systems and their interaction with chemistry and physics-based models. The following subsections detail each approach, highlighting how empirical and mechanistic components are integrated in practice.

3.1. Traditional growth models

3.1.1. Focusing on photosynthesis

This work aims to review and explore the key modelling challenges of phototrophic systems. Therefore, we first analyse how photosynthesis can be modelled. This focus is crucial because photosynthesis is the key process driving growth in such systems, and its modelling often requires specific modelling approaches that go beyond standard macroscopic balance frameworks. There exists a broad range of models representing photosynthetic rates (see Béchet et al., 2013, for a detailed review). These models include three levels of complexity. First, the modelling strategy can be based on a core photosynthesis model valid for low density cultures, where the light can be considered homogeneous. This gross photosynthetic model is generally represented by a function μ(I) (in day^{-1}), which is first linear with the light intensity (I) for small

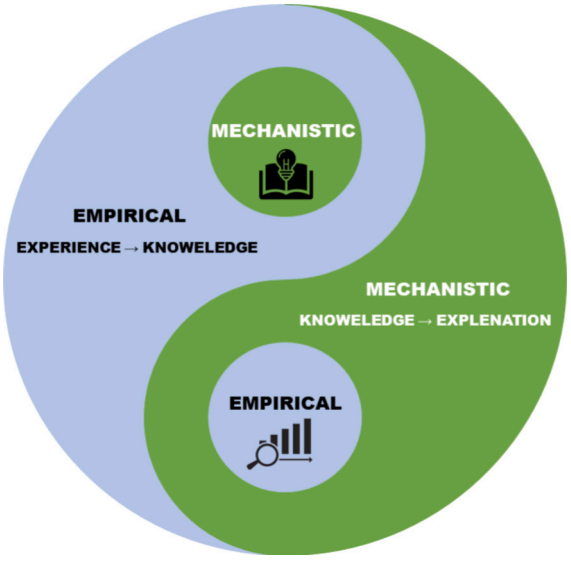

Fig. 2. The Yin-Yang duality of modelling approaches. The blue (left) and green (right) regions represent empirical (experience → knowledge) and mechanistic (knowledge → explanation) approaches respectively, with their opposing dots symbolizing how each paradigm contains elements of the other. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Main characteristics of existing modelling approaches for phototrophic systems.

Modelling structure	Key features	Advantages	Challenges	Applications to phototrophic systems	Interaction with Chemistry/Physics
Macroscopic mass balance-based models	Conservation laws; Dynamic behaviour; Simplified representation; Parameterized by experimental data	Predictive power; Scalability; Integration of experimental data; Flexibility and adaptation to various biological processes	Hypotheses and simplifications; Parameter estimation; Nonlinearity and complexity; Lack of biological knowledge	Wastewater treatment; Bioreactor design; Identification of optimal management strategies; Life cycle and techno-economic assessment	Can integrate chemical reactions, hydrodynamics, and heat transfer via conservation laws.
Metabolic models	Stoichiometric balance; Intracellular metabolic pathways; Flux balance analysis; Omics data integration	Detailed metabolic flux insights; Flexibility to different biotechnological contexts	High complexity; Parameter uncertainty; Difficulty to represent dynamics; Metabolic data requirement; No coupling with physics frameworks	Photosynthesis optimization; Bioproduct generation; Nutrient uptake; Exchange of metabolites	Link to chemistry (metabolite exchange); no direct coupling to heat/fluid dynamics.
Data-driven models	Empirical basis; Machine learning and statistical methods; Big data	No need for mechanistic knowledge; Handling complex nonlinear systems; Identifying hidden patterns and relationships	Big data dependency; No respect of physical constrains; Interpretability; Computational intensity; Overfitting; Limited generalizability when far from training data	Optimizing biomass production and nutrient removal; Real-time control and automation; Soft sensing	Can be used to simulate physics/chemistry if trained on related data, or integrated in chemical-physic frameworks.
Hybrid models	Combination of mechanistic and data-driven approaches; Modular structure	Filling the gap between lack of knowledge and data availability; Interpretability and adaptability; Physical constrains to the data-driven component	Model integration; Computational demand; Data quality; Validation and robustness	Wastewater treatment; Model predictive control of photobioreactors; Optimization	Can be integrated in larger frameworks including chemistry and physics.

irradiance levels. The slope then progressively decreases until cancelling or even becoming negative. If we consider the photosynthetic yield (ν(I), in $m^2.\mu mol^{-1}$), defined as:

$$\nu(I) = \frac{\mu(I)}{I} \quad (1)$$

The yield ν is represented as a decreasing function inversely proportional to: i) a term linear with respect to the light intensity (Michaelis-Menten type); ii) a quadratic term with light (Haldane, or Eilers and Peeters type); or iii) exponentially decreasing (Platt type, Jassby and Platt, 1976).

At high light intensities, photoinhibition takes place, rapidly reducing the photosynthetic yield. This can also be seen as a larger number of photons absorbed at higher light intensities to fix the same amount of CO_2 per biomass unit. In theory, 8 photons are needed for fixing one molecule of CO_2 in optimal conditions. Wasted photons at higher light intensities are visible when considering the quantity $\varphi(I) = \frac{I}{\mu(I)}$ (in $\mu mol.m^{-2}$) indicating how many photons are eventually used per surface unit. Fig. 3 illustrates this general pattern mathematically represented by several models. The most used are the Platt model (Platt and Jassby, 1980) and the Haldane model (Haldane, 1930). Calibrating models with inhibition is not straightforward (see Section 5), especially since not all the parameters have a clear biological meaning but are rather empirical. In particular, the optimal light intensity and the maximum photosynthetic rate are not explicitly represented, which generates correlations between parameters during model identification. The reparameterization proposed by Bernard and Rémond (2012) is recommended for a safer calibration. Since wastewater is often very turbid or coloured (e.g., digestate), thereby attenuating light rapidly, there is a debate around whether it is necessary to use a model with inhibition, as this phenomenon appears only in the first centimetres of the reactor or during start-up periods where biomass concentrations are very low. When dealing with purple bacteria since infrared are rapidly attenuated in water, photoinhibition is taking place only at the surface layer.

Dealing with light attenuation is the second specificity that models must consider. Photosynthetic organisms are permanently advected through a light gradient. The photosynthetic light harvesting systems

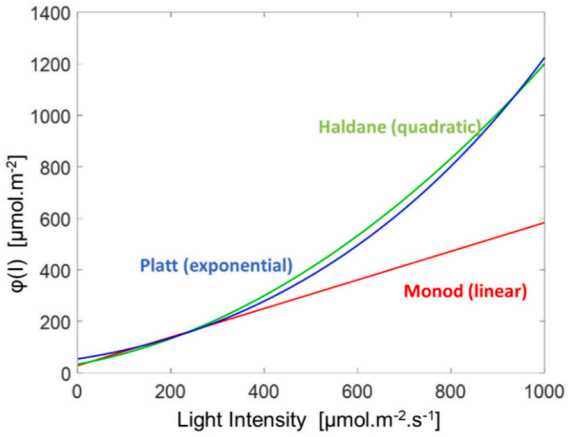

Fig. 3. Number of absorbed photons per m^2 eventually used for the fixation of one mole of CO_2 per mole of biomass $\left(\varphi(I) = \frac{1}{\nu(I)} = \frac{I}{\mu(I)}\right)$. This representation illustrates the similarities of the Platt (blue line) and Haldane model (green line) and their deviations from the Monod model (red line). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

are dynamic structures involved in photon harvesting. At this point, there is a clear crossroad and a decisive modelling choice to be made. Béchet et al. (2013) classified the models into three different categories:

- **Type I light models**: the simplest approach assumes that the average growth rate can be related to the light intensity at a specific depth of the culturing system. Most classically, the average growth rate is supposed to be represented for the light intensity at depth for which the average light $\bar{I}$ is reached. Assuming a Lambert-Beer law to represent the exponential decrease of light irradiance from the surface irradiance (I_0, in $\mu mol.m^{-2}.s^{-1}$) to the one at depth z, due to absorption and scattering phenomena, the light profile from the surface to the bottom at depth L is represented by:

$$I(z) = I_0 e^{-Kz} \tag{2}$$

where K (in m^{-1}) is the light extinction coefficient. The average light can then directly be deduced as:

$$\bar{I} = \frac{I_0}{KL}\left(1 - e^{-KL}\right) \tag{3}$$

$$\bar{\mu}(I_0) = \mu_I\left(\frac{I_0}{KL}\left(1 - e^{-KL}\right)\right) \tag{4}$$

- **Type II light models:** a simplified approach consists in using a static model representing the growth rate as a function of light, and to compute its average along the water depth L (in m):

$$\bar{\mu}(I_0) = \frac{1}{L}\int_0^L \mu(I(z))\mathrm{d}z \tag{5}$$

In both cases, resulting average growth rate $\bar{\mu}(I_0)$ can be explicitly computed for various growth models. Considering the standard mathematical formulation-based Monod kinetics (thus, neglecting photoinhibition):

$$\mu_M(I) = \mu_{max}\frac{I}{I + K_I} \tag{6}$$

Where μ_{max}(day^{-1}) is the maximum growth rate and K_I ($\mu mol.m^{-2}.s^{-1}$) the half saturation constant, leads to:

$$\bar{\mu}_M(I_0) = \frac{\mu_{max}}{KL}\ ln\frac{I_0 + K_I}{I_0 e^{-KL} + K_I} \tag{7}$$

The computations for models with inhibition are trickier, but analytical expressions can be found (Bernard, 2011; Martínez et al., 2018).

- **Type III light models:** the most complicated models are implicitly representing the cell advection in the light field through computational fluid dynamics (CFD) modelling. Then, a dynamic model, such as the Han model, is used to represent the dynamics of the photosystems to assess growth in these conditions of high frequency light change (see as examples Nikolaou et al. (2016) and Fierro et al. (2024)). However, these approaches are complex and highly computational demanding, and the benefit on accuracy is debatable (Fierro et al., 2024).

■ **Modelling photo-acclimation**

Photo-acclimation critically influences growth dynamics. The light-dependent growth rate μ(I) depends not just on instantaneous irradiance, but also on the acclimation light level that shaped the cells' pigment composition. This is not a secondary effect, since the maximal achievable growth rates can vary twofold between dark acclimated and light acclimated cells (MacIntyre et al., 2002). Photo-acclimation also depends on temperature and nutrients availability, especially nitrogen. The Geider's model (Geider et al., 1996) is a rather popular model for representing the photo-acclimation dynamics due to both light variations and nitrogen limitation. The model presented by Bernard et al. (2015) simplifies these coupled phenomena, based on an existing semi-empirical model that describes photo-production, photo--regulation, and photo-inhibition by incorporating acclimation rules for key parameters (Nikolaou et al., 2016). Photo-acclimation is therefore an important mechanism so far neglected in most models of phototrophic systems for WWT, implicitly assuming a dark acclimation of microalgae. More work must be carried out to determine how an upgrade in the representation of this mechanism could impact the growth prediction.

3.1.2. Macroscopic mass balance-based models

Mass balance-based models, often referred to as mechanistic models, are widely used to design and optimize phototrophic systems. These models are efficient for process optimization, to generate data inventories for life cycle assessment (LCA; Fang et al., 2016; Penaranda et al., 2025) and provide a robust conceptual basis for techno-economic evaluations for water treatment and resource recovery. In the domain of wastewater management, most of the modelling efforts have focused on implementing biological models sharing a common structure (Eq. 8 and 9), which is based on elemental conservation:

$$\dot{\xi} = K\rho(\xi, T, \theta, H^+) + \Delta(\xi, \xi_{in}, T, H^+) \tag{8}$$

$$\dot{V} = Q_{in} - Q_{out} + Q_{rain} - Q_{evap} \tag{9}$$

where: ξ represents the state variables vector associated with dissolved or particulate compounds; K is the stoichiometric matrix; ρ is the vector of the biological reaction rates, depending on the model states (ξ), reactor temperature (T), climatology (θ) and pH (H^+); Δ is the vector of fluxes (solid, liquid and gaseous) entering or leaving the system; V is the reactor volume depending on the inflow rate (Q_{in}), outflow rate (Q_{out}), rainfall rate (Q_{rain}), and evaporation rate (Q_{evap}). Although these comprehensive phototrophic growth models share a common structure, they differ in modelling choices, mainly associated to the specific case study (climatology and wastewater characteristics) and the target functions (bioremediation efficiency or resource recovery) they were built for. The central challenge in modelling phototrophic systems lies in achieving an optimal trade-off between mathematical tractability and biological fidelity. This requires judicious selection of: 1) the state variables (ξ), including the granularity of microbial representation (e.g., single-species microalgae vs. complex algae-bacteria consortia), and the biochemical compounds affecting their activities; 2) The process biokinetics ρ, especially the number and type of metabolic processes included (e.g., photosynthesis, nutrient uptake, respiration), and of their mathematical expression, as discussed in the following subsections; 3) Environmental dependencies, i.e., how these processes respond to external drivers (light, temperature) and nutrient availability. Striking this balance is critical, as oversimplification risks biological irrelevance, while excessive complexity can render models computationally intractable or unidentifiable.

A comparison among the main comprehensive phototrophic system models applied to wastewater remediation is provided in Table 2.

3.1.2.1. Nutrient dependence and their interplay. Various approaches capture the complex relationship between nutrient availability and microbial behaviour. These methods range from simple phenomenological models, like Michaelis-Menten kinetics, to more sophisticated models that consider the intracellular content of the cell, such as the quota models following a Droop-like approach. More complex approaches must be used to deal with detailed descriptions of the metabolism, with accumulation and reuse in the cell (Baroukh et al., 2015). Each method provides unique insights, depending on the objective, available data, and the specific biological questions being addressed.

- **Extracellular dependence**

Macroscopic models in biological processes provide a broad and simplified view of how nutrient availability affects growth or metabolic activity. These models often describe the relationship between nutrient concentration and the rate of a biological process, such as microbial growth or enzyme activity, without exploiting the molecular mechanisms.

Monod model: as Michaelis-Menten enzyme kinetics, the model assumes that the growth rate, denoted as μ, is dependent on nutrient concentration, increasing with nutrient availability and eventually reaching a plateau as the system becomes saturated:

Table 2
Structure and characteristics of the main mechanistic phototrophic growth models applied to wastewater remediation.

		RWQM1	PHOBIA	Modified RWQM1	Modified ASM3	Bioalgae2	ALBA	integrated MPBR	ABACO2	ADAB	PAnM	ePAnM	PBM
	Reference	Reichert et al., 2001	Wolf et al., 2007	Broekhuizen et al., 2012	Arashiro et al., 2017	Solimeno et al., 2019	**Casagli et al., 2021a, 2021b**	**Aparicio et al., 2023**	Nordio et al., 2024	Carecci et al., 2024	Puyol et al., 2017	Capson-Tojo et al., 2023	Alloul et al., 2023
	Phototrophic microorganisms	Suspended biomass (microalgae)	Biofilm (microalgae/ cyanobacteria)	Suspended biomass (microalgae)	Suspended biomass (microalgae)	Suspended biomass (microalgae)	Suspended biomass (microalgae)	Suspended biomass (microalgae)	Suspended biomass (microalgae)	Suspended biomass (microalgae)	Suspended biomass (PPB)	Suspended biomass (PPB and microalgae)	Suspended biomass (PPB)
	Other microbial populations	Heterotrophic bacteria, AOB, NOB, Consumers	Heterotrophic and nitrifying bacteria	Heterotrophic bacteria, AOB, NOB, Phagotrophic organisms	Heterotrophic bacteria, AOB, NOB	Heterotrophic bacteria, AOB, NOB	Heterotrophic bacteria, AOB, NOB	Heterotrophic bacteria, AOB, NOB	Heterotrophic and nitrifying bacteria	Heterotrophic bacteria (aerobic/ anaerobic), AOB, NOB	**No**	Heterotrophic bacteria (aerobic/ anaerobic), sulphate reducing bacteria	Heterotrophic bacteria (aerobic/ anaerobic)
	State variable (n°)	24	16	24	16	19	19		8		**10**	**21**	**15**
	Biological processes (n°)	22	13	22	21	18	19		7		**6**	**30**	**11**
	Parameters (n°)	120	75	138	47	108	72		79		**53**	**108**	**62**
	Growth kinetic type	Multiplicative	Minimum	Multiplicative	Multiplicative	Multiplicative	Minimum/ Multiplicative	Multiplicative	Multiplicative	Minimum/ Multiplicative	Multiplicative	Multiplicative	Multiplicative
	Considered C-forms	C_{ORG}	C_{ORG}, CO_2, HCO_3	C_{ORG}, CO_2, HCO_3, CO_3^{2-}	C_{ORG}	C_{ORG}, CO_2, HCO_3	C_{ORG}, CO_2, HCO_3, CO_3^{2-}		C_{ORG}		C_{VFA}, C_{AC}, CO_2, HCO_3	C_{VFA}, C_{AC}, C_{ORG}, CO_2, HCO_3	C_{VFA}, C_{ORG}, CO_2, HCO_3
	Considered N-forms	NH_3, NH_4^+, NO_3^-, NO_2^-	NH_3^+, NO_3^-	NH_3, NH_4^+, NO_3^-, NO_2^-, N_2	NH_4^+, NO_3^-, NO_2^-	NH_3, NH_4^+, NO_3^-, NO_2^-	N_{org}, NH_3, NH_4^+, NO_3^-, NO_2^- HNO_2, HNO_3, N_2	N_{org}, NH_3, NH_4^+, NO_3^-, NO_2^- HNO_2, HNO_3, N_2	NH_4^+, NO_3^-	N_{org}, NH_3, NH_4^+, NO_3^-, NO_2^- HNO_2, HNO_3, N_2	NH_3, NH_4^+, N_{org}	NH_3, NH_4^+, N_{org}	NH_3, NH_4^+, N_{org}
Model structure	Considered P-forms	$H_2PO_4^-$, HPO_4^{2-}	–	$H_2PO_4^-$, HPO_4^{2-}	–	SPO4 (****)	H_3PO_4, $H_2PO_4^-$, HPO_4^{2-}, PO_4^{3-}	H_3PO_4, $H_2PO_4^-$, HPO_4^{2-}, PO_4^{3-}	PO_4^{3-}	H_3PO_4, $H_2PO_4^-$, HPO_4^{2-}, PO_4^{3-}	$H_2PO_4^-$, HPO_4^{2-}	$H_2PO_4^-$, HPO_4^{2-}	$H_2PO_4^-$, HPO_4^{2-}
	Continuity check (mass conservation)	C, O, N, P	(n.s.)	C, O, N, P	(COD, N, P.)	(n.s.)	C, H, O, N, P, COD				**COD, C, N, P**	**COD, C, N, P**	
	PAR model	Steele	Eilers & Peters	Smith	Poisson	Eilers & Peters	Bernard & Remond	Steele	Molina	Bernard & Remond	**Monod**	Steele	**Monod**
	Chemical model	NH_4^+, NH_3, CO_2, HCO_3 CO_3^{2-}, $H_2PO_4^-$, HPO_4^{2-}, Ca^{2+}, H^+, OH^-	NH_4^+, NH_3, CO_2, HCO_3 CO_3^{2-}, H^+, OH^-, $\Delta_{CAT,AN}$	NH_4^+, NH_3, CO_2, HCO_3 CO_3^{2-}, $H_2PO_4^-$, HPO_4^{2-}, Ca^{2+} H^+, OH^-	–	NH_4^+, NH_3, CO_2, HCO_3 CO_3^{2-}, H^+ OH^-	NH_4^+, NH_3, CO_2, HCO_3 CO_3^{2-}, H_3PO_4, $H_2PO_4^-$, HPO_4^{2-}, PO_4^{3-}, NO_2^-, HNO_2, NO_3^-, HNO_3, H^+, OH^-, $\Delta_{CAT,AN}$, TA	H_3PO4, H_2PO4^-, $HPO4^{2-}$, PO_4^{3-}; NH_4^+, NH_3; NO_2, HNO_2; NO_3, HNO_3; CO_2, HCO_3^-, CO_3^{2-}; H^+, OH^-; Mg; K; Ca; Fe; Al	–	H_3PO4, H_2PO4^-, $HPO4^{2-}$, PO_4^{3-}; NH_4^+, NH_3; NO_2, HNO_2; NO_3, HNO_3; CO_2, HCO_3^-, CO_3^{2-}; H^+, OH^-; Mg; K; Ca; Fe; Al	Ac, Ac-, VFA, VFA-NH_4^+, NH_3, HCO_3 CO_2, $H_2PO_4^-$, HPO_4^{2-}, H^+, OH^-	Ac, Ac-, VFA, VFA-NH_4^+, NH_3, HCO_3 CO_2, $H_2PO_4^-$, HPO_4^{2-}, H^+ OH^- HS^-, H_2S	–
	Molar conc./activity	Molar conc.	Molar conc.	Molar conc.	–	Molar conc.	Molar conc.	Activity	–	Activity	Molar conc.	Activity **(for NH_3)**	Molar conc.
	Thermal model	–	–	–	–	–	✓	–	–	✓	–	–	–
	pH growth dependence	–	–	Gaussian law	–	CPMI	CPM	Monod	CPM	CPM	–	**Empirical lower inhibition function**	–

(*continued on next page*)

Table 2 (*continued*)

		RWQM1	PHOBIA	Modified RWQM1	Modified ASM3	Bioalgae2	ALBA	integrated MPBR	ABACO2	ADAB	PAnM	ePAnM	PBM
	Temperature growth dependence	Arrhenius	–	Arrhenius	–	CTMI	CTMI, Arrhenius	Ratkowsky, Arrhenius	CTMI	CTMI, Arrhenius	–	CTMI	–
	Ammonification	–	–	–	–	–	✓		–		✓	✓	✓
	Oxygen inhibition	–	–	–	–	✓	✓		✓		–	✓	✓
	Gas-liquid mass transfer	O_2	–	O_2	–	O_2, CO_2, NH_3	O_2, CO_2, NH_3, N_2, evaporation	O_2, CO_2, NH_3, N_2	O_2	O_2, CO_2, NH_3, N_2, evaporation	–	O_2, CO_2	O_2, CO_2, NH_3, H_2
	Reactor type	River environment	Laboratory incubator	Raceway	Cylindrical photobioreactor	Cylindrical photobioreactor	Raceway	MPBR	Raceway	Raceway	**Schott flasks**	**Flat plate photobioreactor**	**Raceway**
Experimental setup	Reactor installation/ Volume	Outdoor	Indoor (lab), 3 L	Outdoor, 8 m^3	Indoor (lab), 2 L	Indoor (lab), 4 L	Outdoor, 17 m^3–1 m^3	Outdoor, 0.24 m^3	Outdoor, 12m^3	Outdoor (greenhouse), 1530 m^3 (hypothesized case study)	**Indoor (lab), 100 mL**	**Outdoor, 900 L**	**Indoor, 100 L**
	Influent	Wastewater discharge	MM	MWW	DSC	MWW	SWW, DSC	Centrate	MWW	Centrate	**MM**	**Poultry-processing WW**	SWW
	Calibration dataset	–	–	365 d	24 h	8 d	30 d	–	41 d	–	**< 30 d**	**1–4 d**	**40 d**
	Short-term dynamics	–	–		24 h	n.s.	3, 14 d	–		–	–	4 d	–
Calibration - Validation	Long-term dynamics	–	–	330 d	–	–	413 d, 189 d	91 d	206 d	–	–	33 d	–
	Sensitivity analysis	✓	✓	–	✓	–	✓	–	✓	–	–	–	✓
	Parameter Uncertainty	✓	–	–	✓	–	✓	–	✓	✓	✓	✓	–

Notes: DSC: Diluted Swine Centrate; MM: Mineral Medium; MWW: Municipal Wastewater; SWW: Synthetic Municipal Wastewater; TA: Total Alkalinity; CTMI: Cardinal Temperature Model with Inflection; CPMI: Cardinal pH Model with Inflection; CPM: Cardinal pH Model; MPBR: membrane photobioreactor

$$\mu(S) = \mu_{max} \frac{S}{K_S + S} \tag{10}$$

where μ_{max} is the maximum specific growth rate (day^{-1}), S is nutrient concentration (e.g., $mol.L^{-1}$) and K_S (e.g., $mol.L^{-1}$) is the half-saturation constant, the substrate concentration at which the growth rate is half of μ_{max} (Monod, 2012).

Haldane model: extends the Monod framework by accounting for substrate inhibition. This is relevant in bioprocesses where high substrate concentrations can be toxic (e.g., ammonia), inhibiting microbial growth or enzymatic activity:

$$\mu(S) = \overline{\mu} \frac{S}{K_S + S + \frac{S^2}{K_i}} \tag{11}$$

where K_i is the inhibition constant (e.g., $mol.L^{-1}$), representing the nutrient concentration at which inhibition becomes significant (the meaning of $\overline{\mu}$ (day^{-1}) and K_I (e.g., $mol.L^{-1}$) a delicate point for this model, will be discussed in Section 5). Other macroscopic models describing nutrient dependence in biological processes can be found in literature (see for example Bastin and Dochain, 1990).

- **Intracellular dependence**

Intracellular dependence focuses on how the internal storage in cells influences growth and metabolism. Unlike macroscopic models that rely on external substrate concentrations, intracellular models account for the uptake, storage, and exploitation of substrates within the cell. These models are particularly significant for organisms like microalgae, which can store nutrients internally and whose growth may be limited by intracellular rather than external nutrient levels. By modelling the internal nutrient pools, these approaches offer a mechanistic understanding of nutrient limitation and cellular adaptation.

Droop model: the standard mathematical representation for intracellular nutrient dependence is the Droop model (Droop, 1983; Bernard, 2011), which describes microbial growth in terms of the internal nutrient quota (μ(Q)), whereby growth rate is modulated by the internal quota (Q) of the limiting element (e.g., $gN.gDW^{-1}$) as follows:

$$\mu(Q) = \mu_{max} \frac{1 - \frac{Q_{min}}{Q}}{1 - \frac{Q_{min}}{Q_{max}}} \tag{12}$$

Note that here the model has been reparametrized so that the maximum growth rate (μ_{max}) is reached when the internal quota is at its maximum value Q_{max} (e.g. $gN.gDW^{-1}$), while Q_{min} (e.g. $gN.gDW^{-1}$) is the minimum intracellular nutrient quota required to sustain growth.

In addition, this model is associated to a nutrient uptake model. The difference in uptake rate and dilution due to cell growth exactly explains the quota dynamics. The uptake model is generally taken as a Monod type model with respect to the consumed nutrient. More complex models exist to account for the complex uptake regulations appearing when several nutrients can become limiting at the same time (Bougaran et al., 2010).

This model structure is especially valuable for applications targeting microbial biomass-based resource recovery, as it captures not only biomass productivity but also quality attributes such as composition (e.g., protein content; Wágner et al., 2021) and carbon storage potential (e.g., for biogas or biodiesel production; Guest et al., 2013).

- **Multiplicative and minimum law**

Different approaches can describe how multiple nutrients interact to influence microbial growth and providing different perspectives on how nutrient limitations affect biological processes. Two common frameworks are the multiplicative law and Liebig's minimum law.

Multiplicative law: this approach assumes that the availability of several nutrients contributes to the microbial growth synergistically,

with each nutrient acting independently, but collectively determining the growth rate. A common formulation of this approach is reported in Eq.13, where the growth rate is proportional to the product of individual nutrient availability terms.

$$\mu(S) = \mu_{max} \prod_{i=1}^{n} \frac{S_i}{K_i + S_i} \tag{13}$$

where n is the number of limiting nutrients.

Liebig's minimum law: this approach is based on the idea that microbial growth is driven by the most limiting nutrient, regardless of the availability of other nutrients. As for the multiplicative law, no further growth can occur once the most limiting nutrient is exhausted, even if other nutrients are present in excess.

$$\mu(S) = \mu_{max} min\left(\frac{S_1}{K_1 + S_1}, \frac{S_2}{K_2 + S_2}, \ldots, \frac{S_n}{K_n + S_n}\right) \tag{14}$$

Even if both approaches have practical applications in several modelling works, the Liebig's minimum law provides a more realistic representation of nutrient limitation in biological systems, as it captures the critical role of the most limiting nutrient in constraining growth. This approach aligns with empirical observations in various ecosystems, where the availability of a single nutrient drives the overall grow rate, regardless of the surplus of other resources. Note that it can be extended to a Droop-type model (Bougaran et al., 2010).

3.1.2.2. Temperature and pH dependence. There exists a variety of models representing the influence of temperature on growth (Grimaud et al., 2017). The Cardinal Temperature Model with Inflection (CTMI model, Rosso et al., 1995) is the most used, and it can be adapted to photosynthetic organisms (Bernard and Rémond, 2012). It represents growth within the range of temperatures in which the microorganism can grow. There is a minimum temperature (T_{min}) above which growth becomes possible. Then, by increasing temperature, growth (modulated by the function $\varphi_T(T) \in [0,1]$), increases until it reaches a maximum at optimal temperature (T_{opt}). Then it drops due to mortality losses for excessive temperatures, until growth cancels at maximum temperature (T_{max}). The CTMI model is given by the following function for $T \in [T_{min}, T_{max}]$:

$$\varphi_T(T) = \frac{(T - T_{max}) \cdot (T - T_{min})^2}{(T_{opt} - T_{min}) \cdot \left((T_{opt} - T_{min}) \cdot (T - T_{opt}) - (T_{OPT} - T_{max}) \cdot (T_{opt} + T_{min} - 2 \cdot T)\right)} \tag{15a}$$

and $\varphi_T(T) = 0$ for $T \notin [T_{min}, T_{max}]$. Note that temperature acclimation is a dynamic process, like photo-acclimation to light, taking place at daily scale. The last part of the temperature response, with the drop in the growth rate at temperatures higher than the optimum, distinguishes the CTMI model from the more standard Arrhenius law (commonly applied in IWA and other biological wastewater treatment models). This response at high temperatures is crucial for processes whose temperature can become dramatically high in summer. It is even likely that a mortality term at higher temperature (or high temperature in conjunction with high oxygen levels; López Muñoz and Bernard, 2021) is also important for catching the effect of overwarming on the microbial ecosystem.

Similarly, pH dependence can be described using the Cardinal pH Model (CPM) (Rosso et al., 1995), which requires three parameters, defining the effective growth range for each microbial population. Within this range, growth (modulated by the function $\varphi_{pH}(pH) \in [0,1]$) increases from 0 at the minimum pH value (pH_{MIN}) to a maximum at optimal pH value (pH_{OPT}), then cancelling at maximum pH value (pH_{MAX}), and $\varphi_{pH}(pH) = 0$ for $pH \notin [pH_{min}, pH_{max}]$.

$$\varphi_{pH}(pH) = \frac{(pH - pH_{MIN})(pH - pH_{MAX})}{(pH - pH_{MIN})(pH - pH_{MAX}) - (pH - pH_{OPT})^2} \tag{15b}$$

The CPM captures both the inhibitory effects at low and high pH values and the maximum growth rate achieved at the optimum. When not controlled, pH can indeed exhibit strong daily variations in phototrophic systems, and its influence on microbial activity is essential for accurate process modelling. This model has been extensively validated for bacteria. For photosynthetic organisms, it must be used with care, since pH has a double effect. Beyond the direct effect on internal regulation of protons, it also triggers the availability of CO_2, the substrate of Rubisco in the Cavin cycle. It therefore also indirectly acts on the photosynthesis rate.

3.2. Metabolic models

Metabolic modelling, coupled with flux balance analysis, harnesses -omics data to construct comprehensive metabolic networks, which are then transformed into mathematical models for computational analysis and validation against metabolic and phenotypic measurements (Steuer et al., 2012; Baroukh et al., 2015). These models are crucial for understanding and predicting the behaviour of biological systems, optimizing metabolic pathways, and designing experiments in biotechnology and synthetic biology. Furthermore, metabolic models can help to understand the interactions within microbiomes (e.g., syntrophy or competition), enabling more in-depth tools for optimization (Pessi et al., 2023).

Advances in genome exploration have led to more detailed metabolic networks, which have been proposed for several species of microalgae, and more and more at genome scale. Based on these networks, a steady state model can be built based on algebraic equations, which is often solved through flux balance analysis (FBA), whereby metabolic fluxes are optimized based on an objective function (e.g., microbial growth; Orth et al., 2010).

Several approaches have been proposed to extend the hypothesis of static environment to more realistic dynamic frameworks (Baroukh et al., 2015; Baroukh et al., 2017). Dynamic FBA is rarely applied to environmental biotechnologies, due to its inherent complexity and the underlying optimization criterion which must be implemented. To solve this, the main metabolism should be embedded into a meta-model where the main fluxes can be represented. This simplified framework is still too far from the reality of the outdoor photosynthetic process. Advances such as the DRUM (Baroukh et al., 2014 and Baroukh et al., 2017) approach have been decisive in integrating dynamics into metabolic models, with application in WWT (Baroukh et al., 2017; Assis Pessi et al., 2023). The advantage of this approach is to drastically reduce the number of parameters to be included in the model by considering subnetworks and the associated Elementary Flux Modes (Baroukh et al., 2016). Yet, these developments often focus on a single species. Modelling an ecosystem as a meta-organism (also known as a microbiota) is emerging in other fields (Dittami et al., 2021). This approach has the advantage of bypassing the notion of species and reconstructing the metabolism of the microbial ecosystem by bringing together the metabolism of the individuals that make it up. The main challenge is to develop metabolic modelling for a set of interacting microorganisms, accounting for their diverse metabolic networks, varying functions, and species succession over time. It also leads to a gigantic metabolism of

tens of thousands of metabolic reactions. This is even more complicated when dealing with an environment submitted to periodic fluctuations of light and temperature, and a large amount of work is still necessary to be able to setup and use such metabolic models in the wastewater field.

3.3. Emerging growth modelling approaches

3.3.1. Data-driven models

Recent advancements in computational power have facilitated the increased application of machine learning (ML) techniques to biological models for wastewater bioremediation, helping to improve and optimize the process efficiency (Otálora et al., 2023; Bahramian et al., 2023; Caparroz et al., 2024). The integration of artificial intelligence (AI) and big data within the Dynamic Data-Driven Application Systems (DDDAS) paradigm enables the deployment of real-time digital twins in complex systems, following the example of oil reservoir management (Parashar et al., 2023). Real-time assimilation of field data via IoT (Internet of Things) sensors, combined with AI-based learning models, can contribute to adapt simulations to the process state. Edge computing plays a crucial role in this context by allowing local data processing and lightweight model inference at the source, reducing both latency and communication overhead. For instance, the use of lightweight surrogate models, trained to approximate high-fidelity CFD simulations, allows for rapid evaluation of production strategies directly on edge devices. This hybrid architecture ensures both scalability and responsiveness, enabling closed-loop control and optimization in dynamic and uncertain environments characteristic of wastewater treatment processes. Furthermore, DDDAS is likely to integrate AI-driven inference, distributed computing, and large-scale data orchestration to support autonomous, real-time decision-making in algae-bacteria processes. These methods require advanced tools for managing large, complex datasets and are highly applicable to wastewater treatment systems to improve process control and overall performance. The integration of AI, IoT, and edge computing plays a transformative role in enabling smart monitoring, advanced modelling, and real-time control in WWT processes. While these technologies offer clear benefits in terms of efficiency and sustainability, their effective deployment still requires addressing practical challenges related to integration, cost and scalability (Cairone et al., 2024).

Unlike physics-based models, data-driven approaches derive their structure directly from large datasets, exploiting their ability to recognize complex patterns without relying on explicit physical laws or first-principles equations. Data-driven models can be classified according to their training method.

In WWT applications, supervised learning is the most common approach, where models in the training phase learn relationships between input features and labelled target outputs (experimental measurements) by adjusting internal parameters in order to minimize a cost function (often called loss function). In the field of WWT, ML is mainly used for: i) non-linear regressions, for catching non-linear relationships in biological processes (e.g. Symbolic Regression); ii) classification models as support vector machine (SVM), random forests or convolutional neural networks (CNN), frequently employed for image classification in phototrophic systems (Chong et al., 2023). These models are mostly applied as soft sensors for monitoring the microbial communities in mixed phototrophic systems (Otálora et al., 2021; Gincley et al., 2025), pigmentation changes or life cycle of specific algal species (Stegemüller et al., 2025); iii) clustering techniques, as k-means clustering and hierarchical clustering, for helping in understanding patterns, data segmentation and analysing the microbial communities structure in WWT systems; iv) artificial neural networks (ANNs), which are particularly effective for modelling complex non-linear relationships in biological processes and for predicting the WWT process outcomes (Rodríguez-Rángel et al., 2022); v) recurrent neural networks (RNNs) and LSTM (Long Short-Term Memory) have proven to be particularly effective for dynamic bio-process modelling, due to their ability to handle sequential data (Pisa et al., 2019; Wongburi and Park, 2023).

Among these techniques, ANNs are the most widely used models for phototrophic processes, and more in general for complex biological system models (Del Rio-Chanona et al., 2019). Their general structure is described in Eq. 16, while Eq. 17 defines the cost function minimization from where the parameters Ω of the network are derived:

$$\dot{\xi} = A_{NN}(\xi(t), u(t), \Omega^*) \tag{16}$$

$$\Omega^* = argmin_{\Omega} \left| \xi(t_j, \Omega) - \xi^{meas}(t_j) \right| \tag{17}$$

where ξ(t) is the model state vector, u(t) is the input data vector, Ω* is the parameters vector minimizing the cost function, $\xi^{meas}(t_j)$ is the vector of experimental data at the N time instants t_j. There are bottlenecks when using purely data-driven models:

- Big data are required for properly training the model, avoiding overfitting problems and unbalances between available data and model parameters. The amount of available data in WWT is often too small, especially when studying new and emerging biotechnologies, since experimental data mainly come from pilot scale applications, where long-term records at high frequency are lacking.
- There is no integrated physical knowledge, therefore predictions can become negative or not respecting elemental mass balances.
- The computational and memory cost required for model calibration can be very high when manipulating big data and time-integrated backpropagation algorithms.

3.3.2. Hybrid models

Hybrid models, coupling mechanistic and data-driven components, represent the last frontier among modelling techniques. These models found a great consensus in biological models applied to WWT, since they can cover the physical, chemical, and biological knowledge, and embed the missing complexities hidden in real data (de Wolff et al., 2022; Mahanty, 2023). Hybrid models are often classified according to their structure, i.e., parallel or series configurations (Schneider et al., 2022; Syed et al., 2024). In general, in the parallel structure (Fig. 4A) the data-driven component (e.g., a neural network) acts on the output of the mechanistic model (e.g., a mass balance-based system), providing a correction term (Eq. 18a). In the series structure (Fig. 4B and C), the output of one component is used as input to the other, with the most common configuration being a data-driven component (e.g., a neural network) estimating the most uncertain part (e.g., bio-kinetic rates, specific growth/decay parameters) that are then used as input to the mechanistic model (e.g., a mass balance-based system) (Eq. 19a).

When dealing with dynamic systems, this classification in series or parallel (Fig. 4) loses its meaning. It becomes more subtle when considering how the predictions of the data-driven component propagate through the state of the system, which in turn implicitly depends on the parameters of the neural network. Since the data driven output is embedded in the system's time evolution (e.g., via neural ODEs), the interaction between components becomes iterative and co-dependent, losing the clear difference between "input-output" or "parallel correction" structure. In this last case, it can be more appropriate to describe hybrid models depending on i) where the mechanistic knowledge is located (in the dynamics or in the loss function for the PINN's) ii) how the loss function is computed, especially if the computed difference includes directly the output of the neural network, if the difference is indirect through the state variables, and if penalty terms are added to force the hybrid system to satisfy some constraints. This leads to three main categories of approaches:

i) *augmented models*, when the dynamic data-driven component estimates uncertain mechanistic parameters over time. As an example, a neural network predicting time-varying kinetic rates, integrated into ODEs governing mass balance

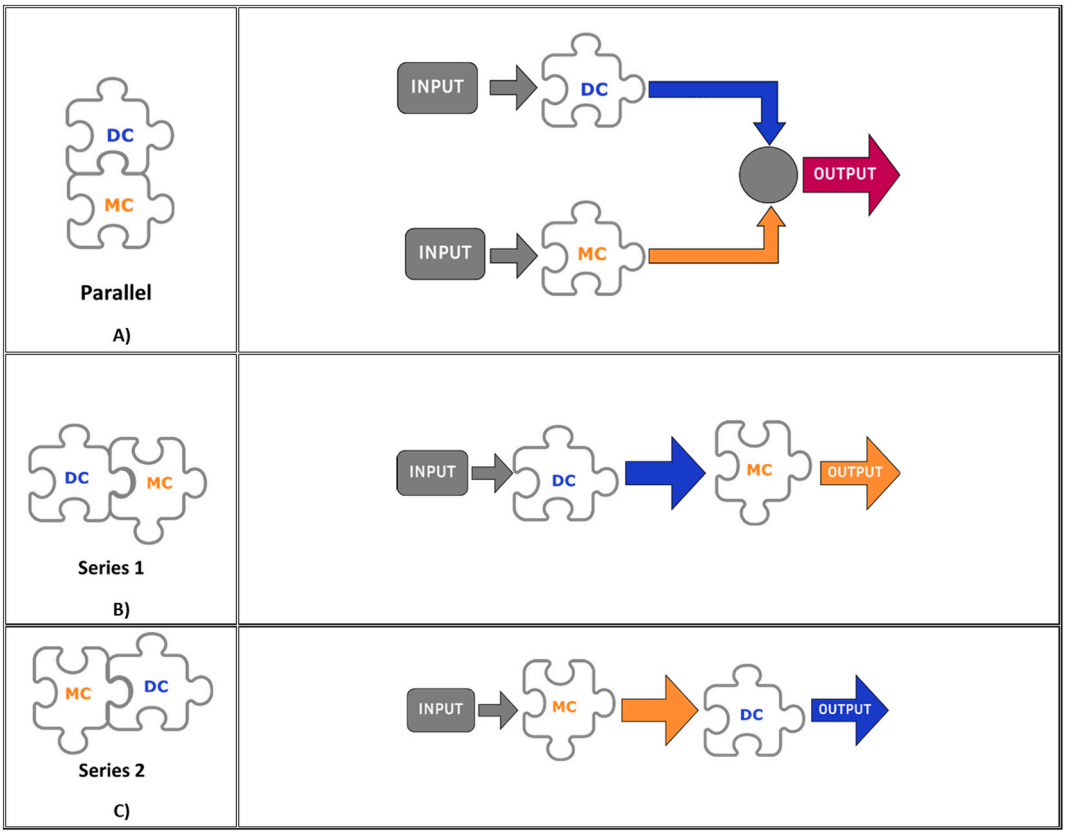

Fig. 4. General scheme for hybrid model configuration: parallel (A) and series (B and C), as combination of mechanistic component (MC) and data-driven component (DC).

$$\dot{\xi}(t) = K\rho_{NN}(\xi(t), u(t), \Omega^*) + D(\xi_{in} - \xi(t)) - Q(\xi(t)) \tag{18a}$$

Where K is the stoichiometric matrix, and ρ_{NN} is the ANN predicting the reaction rates. The computation of the optimal parameters of the network, Ω^*, results from the minimisation of a loss function. In the simplest case, it is static ($\mathscr{L}_{1,static}$), i.e. direct minimisation of the network output error when it is computed as the difference of the prediction of the ANN, ρ_j^{NN}, and its measurements ρ_j^{meas} at time t_k, as follows:

$$\mathscr{L}_{1,static}(\Omega) = \sum_j \left\| \rho_j^{NN}(t_k, \Omega) - \rho_j^{meas}(t_k) \right\|^2 \tag{18b}$$

where Ω represents the parameters of the ANN. This however requires a measurement (or an estimation) of the reaction rates. The computation of Ω^* can be dynamic ($\mathscr{L}_{2,dynamic}$), by minimizing the state predictions computed through the neural ODEs. The typical loss function is then:

$$\mathscr{L}_{2,dynamic}(\Omega) = \int_{t_0}^{t_f} \| \xi(t, \Omega) - \xi^{meas}(t) \|^2 \, dt \tag{18c}$$

The loss function is no more explicitly referring to the output of the data-driven component, but to a quantity (typical the state variable) which is computed in a dynamic system involving the data-driven component. Here, the backpropagation during the training phase involves to solve sensitivity equations solution of an ODE (e.g., adjoint methods), and the loss depends on the integrated trajectory through the state variable $\xi(t)$.

ii) *corrected mechanistic models*, when the data-driven term is a time-dependent network $R_{NN}(\xi(t), u(t), \Omega)$, correcting the mechanistic model's state (e.g., a LSTM network learning residual errors in a physical simulation, where corrections are applied at each integration step). The dynamics of the system writes then:

$$\dot{\xi}(t) = K\rho(\xi(t)) + D(\xi_{in} - \xi(t)) - Q(\xi(t)) + R_{NN}(\xi(t), u(t), \Omega^*) \tag{19a}$$

Here again, the available measurements, $\xi^{meas}(t)$, are not directly predicted by the output of the ANN, and the loss function is finally defined as reported in Eq. 18c.

Another approach consists in correcting the predictions of the mechanistic model (ξ_j^{mec}) by an independent network, which will not influence the trajectories. In this case, the ANN is modelling the prediction error by a term $A_{NN}\left(\xi_j, u_j \Omega\right)$. The loss function to be implemented is then:

$$\mathscr{L}_{3,static}(\Omega) = \sum_j \left\| \xi_j^{mec} + A_{NN}(\xi_j, u_j \Omega) - \xi_j^{meas} \right\|^2 \tag{19b}$$

iii) *PINNs: Constrained hybrid models*. When hybrid models incorporate physical equations (e.g., mass and energy balances) as explicit penalty terms in the loss function, they fall under the specific subclass of Physics-Informed Neural Networks (PINNs). A general structure for a dynamic PINN, where the dynamics is given by a neural network $R_{NN}(\xi(t), u(t), \Omega^*)$ is as follows:

$$\dot{\xi}(t) = R_{NN}(\xi(t), u(t), \Omega^*) \tag{20a}$$

$$\mathscr{L}_{PINN}(\Omega) = \frac{1}{N} \sum_k \| \xi(t_k, \Omega) - \xi^{meas}(t_k) \|^2 + \lambda \int_{t_0}^{t_f} \left\| R_{phy}(t, \xi, u(t)) \right\|^2 dt \tag{20b}$$

The constraint imposed by the physical knowledge is given by the term $R_{phy}(t, \xi, u(t))$ and λ is the weight of the physical residual. This penalty term represents physical or biological constrains, such as mass conservation, incompressibility, positivity, etc. The more this constraint

is violated, the higher this penalty term is. For constraining the hybrid model to respect a mass balance model structure, the following penalty can be considered:

$$R_{phy}(t) = \dot{\xi}(t) - [K\rho(\xi(t), u(t)) + D(\xi_{in} - \xi(t)) - Q(\xi(t))] \quad (20c)$$

This classification includes most of the current hybrid structures. But often, these different approaches are mixed for more efficiency, as illustrated in Fig. 4. Including physical constraints (such as non-negativity of states) can be achieved through either: i) penalty terms in the loss function or ii) architectural enforcement, which is classified as general hybrid systems with soft physical regularization (de Wolff et al., 2022). This distinction is crucial, since only hybrid models enforcing physics through the loss function are qualified as PINNs.

The main advantages in using hybrid structures are: i) the data-driven component potential can be efficiently exploited, constraining it with respect to some physical properties in the cost function or directly in the model structure, avoiding weird predictions and conserving elemental mass; ii) the calibration phase can become more straightforward, allowing for faster adaptation of the model to very different case studies; iii) they can be applied for online simulations and optimization (Zhang et al., 2019). These hybrid model structures have been applied to bioprocesses (Meenatchisundaram et al., 2024). The most common approach is coupling mass balance-based models with ANNs (Chen et al., 2000; Karama et al., 2010; de Wolff et al., 2022). One of the main challenges when hybridizing mass balance-based models and ANNs is to adapt the neural network structure (e.g, number of layers, neurons per layer) to the available experimental data, in order to avoid overfitting. The computational cost can also be a strong limitation when using neural ODEs, reason why it is recommended to use programming languages that support automatic differentiation (e.g., Julia, Phyton-JAX).

3.4. Chemical modelling

In phototrophic systems, microbial processes affect the concentration of compounds that drive the evolution of pH and undergo acid-base speciation. In turn, pH directly affects microbial growth or indirectly by affecting the concentration of the real substrates or inhibitors through speciation or affecting gas transfer. Therefore, it is fundamental to accurately represent the pH and dissociation states of the key chemical species (e.g., carbonate system). However, biological models' structure is still rarely associated to a chemical sub-model (Shoener et al., 2019). In fact, several phototrophic models are implementing the effect of the pH dynamics on the biokinetic rates using pH measurements as a model input. This choice simplifies the overall model implementation, but it does not allow exploring scenarios under different operational and environmental conditions (Casagli et al., 2021a, 2021b and Casagli et al., 2023), limiting the detection of sub-optimal conditions and the identification of optimal operational strategies for improving the system performances.

The main pH models proposed in literature are based on dissociation equilibria, mass balances of acids and bases and on charge balance, through which the concentration of protons is computed. These sub-models mainly differ on the following aspects:

- **Ionic species implementation:** the choice of the ionic species to be included in the chemical model is mainly driven by the wastewater composition and characteristics (e.g., municipal wastewater vs. digestates), and from the choice of bioprocesses and metabolites dynamics included in the biological model.
- **Numerical resolution method:** formulating the chemical model resolution as a differential algebraic equation system (DAE) is the most widely used approach for complex systems, where the slow dynamics of biokinetic rates are implemented as differential equations, while faster chemical equilibria are implemented as algebraic equations. Another approach is to use a differential equation system (DE) for solving acid-base equilibria as fast differential equations (Casagli and Bernard, 2022). This last approach is mainly recommended for limiting the computational cost. In both cases, a proper initialization of the chemical model is required.
- **Representation in terms of concentration or activity:** the choice is related to a crucial hypothesis that must be stated. If the system can be considered enough diluted in terms of solutes concentration, therefore closer to ideal conditions (i.e., infinite dilution), the chemical equilibria can be expressed in molar concentrations. However, if the solutes concentrations are high (a characteristic typically defined by the ionic strength of the system), then the chemical model must be expressed in terms of activities, connecting the molar concentration of the ionic species to their activity coefficients (Batstone and Flores-Alsina, 2022; Solon et al., 2015). This last case is typical of systems dealing with waters characterized by a high salinity (e.g., marine water or high-strength digestates, Aparicio et al., 2023; Carecci et al., 2024). Neglecting the implementation in terms of activity for wastewater characterized by high ionic strengths would challenge accurate pH predictions, since the ion pairing equilibria (i.e., solute species associating to others, forming ion pairs) can have a significant impact and bias the model predictions (Millero and Schreiber, 1982).

If pH conditions induce chemical precipitation, it should be included in the modelling framework to properly model substrate availability and effluent quality (Mbamba et al., 2015). Moreover, chemical modelling enables the simulation of system alkalinity, a key parameter influencing biological process stability and buffering capacity (Wolf-Gladrow et al., 2007). Additionally, for saline or high-strength waters, it allows the estimation of electrical conductivity, which provides also valuable information (e.g., on ionic strength, potential osmotic stress).

3.5. Physical modelling

3.5.1. Hydrodynamics

Small-scale reactors can often be considered as perfectly mixed, but at industrial scale this assumption is no longer valid. Hydrodynamics influence key processes such as light exposure, nutrient transport, gas-liquid mass transfer and solids transport (e.g., settling). Various modelling approaches with different complexity levels are available depending on system objectives and data availability (Jourdan et al., 2019). Balancing model complexity and the expected increment of prediction capability is essential, since numerical methods implementing the Navier-Stokes equations are computationally demanding (Laurent et al., 2022).

To simulate hydrodynamic behaviour of photobioreactors, one can consider using the systemic modelling approach: this simple approach is a black-box model that considers only inlet and outlet data, based on the residence time distribution (RTD), firstly introduced by Danckwerts in 1953. By modelling a network of ideal reactors (e.g., stirred tank, plug-flow) connected under various configurations (e.g., in series, parallel, with recirculation), it provides global-scale insights with lower computational effort (Levenspiel, 1998). However, this approach does not capture the local phenomena which are of particular importance for phototrophic systems (e.g., variability of gas-liquid coefficients, light penetration, cells trajectories).

Alternatively, computational fluid dynamics (CFD) have become a state-of-the-art tool in the WWT industry (Laurent et al., 2022) and for photobioreactors design (Gernigon et al., 2019; Ranganathan et al., 2022). It solves the Reynolds averaged Navier-Stokes (RANS) equations by numerical schemes (e.g., finite elements or finite volume), to predict how geometry influences process outputs through 2D or 3D velocity fields. Simulating phototrophic systems with CFD requires key considerations (Wicklein et al., 2015):

- Selection of turbulence model and boundary conditions.

- Definition of system geometry and a space discretization (meshing).
- Selection of a numerical scheme, either through Eulerian approaches, or through Lagrangian particle tracking for cell trajectories where light variations are reconstructed with a radiative transfer sub-model (Gernigon et al., 2019).
- Accurately modelling rotating elements (propellers, paddlewheels, etc.) can be computationally demanding (Amini et al., 2018). Simplified methods involve introducing a momentum source term in the governing equation (Liffman et al., 2013) or setting a constant inlet velocity on a virtual inlet boundary (Hadiyanto et al., 2013), useful for design studies (e.g., impact of baffles, deflectors). More detailed approaches, such as dynamic or sliding mesh techniques, simulate the momentum transferred to the fluid through the rotating mesh, but require significant computational effort (Hreiz et al., 2014; Huang et al., 2015; Yang et al., 2016), particularly when the model is integrated with biokinetics, gas transfer, or light models. If pH control involves gas injection, its impact on flow should also be considered.

Once the velocity field and Lagrangian particle tracking are implemented, sub-models for light, species transport, thermal effects, chemical equilibria and biological reactions can be integrated. This coupling can be one-way (e.g., hydrodynamics influencing cells transport and biokinetics) or two-ways (e.g., light affecting microalgae growth, which in-turns impacts light diffusion; Gao et al., 2017).

Currently few studies integrate hydrodynamics, light irradiance, mass transfer and biokinetics, mainly focusing on small-scale photobioreactors for biotechnological applications (Gao et al., 2017; Fierro et al., 2025; Pozzobon and Perré, 2020). The high computational cost limits the ability to perform long-term simulations in wastewater applications, especially those with dynamic boundary conditions (e.g., varying flowrates, pollutants load, or meteorological conditions) or in digital twins.

Due to the computational load induced by CFD modelling, model complexity reduction can be useful for evaluating dynamic PBR operations:

- Compartmental modelling offers a more computationally efficient alternative to CFD, capturing local-scale phenomena without requiring the same computational effort (Jourdan et al., 2019). It uses functional zones (compartments) to represent system geometry and spatial distributions, with data on compartments and flux exchanges obtained from local measurements, expert knowledge, or preliminary CFD simulations. Currently this approach has not been applied to phototrophic systems in wastewater applications. Implementing (bio)chemical sub-models is straightforward in this approach, incorporating sink/source terms in ordinary or partial differential equations for species transport.
- Degeneration of CFD to zero- or one-dimensional simulation models can also be considered. Methods for complexity reduction (e.g., Islam and Lye, 2009), include statistical meta-models that bypass micro-scale fluid dynamics (Guyonvarch et al., 2015). For example, complex hydrodynamic effects on mixing and solids transport can be represented using a one-dimensional advection-dispersion equation. The pseudo-dispersion coefficient is used there to account, e.g., for the effects of multi-dimensional motion on solids transport. This coefficient is predicted using polynomial regression meta-models, relying on statistical inferences derived from Latin hypercube or factor screening sampling methods, considering design, operational conditions and process parameters. Such meta-models can also be used to simulate gas-liquid mass-transfer (Qiu et al., 2024b). However, they require numerous preliminary CFD simulations for their calibration, demanding significant time and effort.

3.5.2. Heat transfer model

Most of the current mathematical models for outdoor biotechnological processes strictly rely on measuring the medium temperature or the temperature of the soil below the reactor. Consequently, it is not possible to predict future process dynamics or to perform scenario analyses under different climatic conditions. Thus, it is useful to integrate the biological model in a wider framework, where a physical model predicting the reactor temperature is also implemented. Thermal models of different complexities can be implemented, according to the reactor geometry and configuration, as well as the target application of the model (e.g., automatic control, monitoring, scenario analysis).

- **Regression models:** simple approach whereby reactor temperature is predicted by a simple auto-regression model, using available measurements from the local climatology, such as air temperature and incident irradiance. Eqs. (21) and (22) describe two regression models for simulating the reactor temperature at time t in an outdoor raceway, denoted as $T_{R1}(t)$ and $T_{R2}(t)$ (temperatures are in Kelvin) respectively, as proposed by Casagli et al. (2021b) and Pessi et al. (2022).

$$T_{R_1}(t) = \alpha_1 T_{air}(t) + \alpha_2 T_{air}(t - t^*) + b \tag{21}$$

$$T_{R_2}(t) = c_1 T_{air}\left(t - t_1^*\right) + c_2 T_{air}(t) + c_3 I(t) + c_4 T_{R_2}\left(t - t_2^*\right) \tag{22}$$

Where α_1, α_2, c_1, c_2, c_3 and c_4 are the regression coefficients; $T_{air}(T)$ is the air temperature measured at time t, respectively; $T_{air}(t\text{-}t^*)$ and $T_{air}(t\text{-}t_1^*)$ and $T_{air}(t\text{-}t_2^*)$ is the air temperature measured with a delay time t^*, $t_1^*$and t_2^*, respectively; I(T) is the incident irradiance ($W.m^{-2}$) measured at time t.

The main bottleneck of this approach is that the model has no generalization capability to other case studies, for which a re-calibration of regression coefficients and the delay in time (t^*) are necessary.

- **Full mechanistic heat transfer models:** implementing a complex physical model considering all the heat transfer rate terms acting on the reactor is a more robust approach (Béchet et al., 2011). This model considers the geometry of the reactor, its location in space (i.e., whether it is lying on or raised off the ground), and the thermal characteristics of the construction materials (Casagli and Bernard, 2022; Carecci et al., 2024). Additionally, local climatology experimental data (air temperature, incident irradiance, wind speed, relative humidity) are required as model inputs. In this framework, the dynamics of water temperature and level in the reactor are governed by energy and mass balance equations, as reported in Eq. (23) and (24), which refer to the case of an outdoor raceway raised above the ground (the heat balance of the reactor material is not reported here):

$$\begin{aligned} S\rho_w C_w \frac{\mathrm{d}(h_L T_R)}{\mathrm{d}t} = {} & Q_{ra,r}(T_R) + Q_{ra,s}(I_0) + Q_{ra,a}(T_a) + Q_{h,evap}(\eta, T_a, T_R, \nu_0) \\ & + Q_{conv}(T_a, T_R, \nu_0) + Q_{h,in}(T_{in}, T_R) + Q_{h,out} + Q_{h,rain}(T_a, T_R) \\ & - Q_{conv,cond_w,m}(T_R, T_m) \end{aligned} \tag{23}$$

$$S\frac{\mathrm{d}h_L}{\mathrm{d}t} = -S\frac{m_e}{\rho_w} + q_{rain} + q_{in} - q_{out} \tag{24}$$

Where: S is the reactor surface (m^2); T_R is the reactor temperature (K); T_a is the air temperature (K); I_0 is the incident irradiance ($W.m^{-2}$); η is the relative humidity (−); ν_0 is the wind speed ($m.s^{-1}$); h_L is the water level (m); V is the reactor volume (m^3); $Q_{ra,r}$ is the radiation from the reactor surface (W); $Q_{ra,s}$ is the total (direct + diffuse) solar radiation (W); $Q_{ra,a}$ is the radiation from the air (W); $Q_{h,evap}$ is the evaporation flux (W), Q_{conv} is the convective flux at the pond surface (W); $Q_{h,in}$ and $Q_{h,out}$ is the heat flux associated with the influent and effluent water, respectively (W); $Q_{h,rain}$ is the heat flux related to rain (W); $Q_{conv,condw,m}$ is the conductive/convective term related to the heat exchange between the water in the reactor and the reactor material; m_e is the evaporation rate

($Kg.s^{-1}.m^{-2}$); ρ_w is the water density ($Kg.m^{-3}$); C_w is the water specific heat capacity ($J.Kg^{-1}\cdot K^{-1}$); q_{rain}, q_{in} and q_{out} are the rain, inflow and outflow rates, respectively ($m^3\cdot s^{-1}$). These models guarantee a better generalization when applied to different case studies, without requiring a re-calibration, but involve higher development and implementation complexities.

- **Self-adaptive reduced models:** this approach represents a trade-off between the above-mentioned types of heat transfer models. In this case, a simplified version of the full mechanistic heat transfer model is presented. After a sensitivity analysis, only the most relevant heat transfer components are considered and a simplified generic model keeping the main mathematical structure is extracted (Gharib et al., 2024; Gharib et al., 2025). The number of parameters and the overall model complexity are significantly reduced, so that just few days of experimental data are required for an efficient fine-tuning calibration.

$$\frac{dT_R}{dt} = \left(\theta_1 A(T_R) + \theta_2 B\left(T_R^4\right) + \theta_3 C(T_a) + \theta_4 D\left(T_a^4\right) + \theta_5 E(I_0) + \theta_6 F(T_R, T_a, \nu_0, \eta)\right)\left(h_L^{-1}\right) + Q_{flow}(T_R q_{in})(h_L S)^{-1} \quad (25)$$

$$\frac{dh_L}{dt} = \theta_6\theta_7 F(T_R, T_a, \nu_0, \eta) + q_{in}S^{-1} - q_{out}S^{-1} + q_r \quad (26)$$

Where: A, B, C, D, E are linear functions with respect to T_R, T_R^4, T_a, T_a^4, and I_0, respectively; θ_1 to θ_7 are adaptation coefficients; F is a function proportional to the evaporation flux; Q_{flow} represents the difference between the influent and effluent heat flux ($m^3\cdot K.s^{-1}$).

This approach offers a trade-off for simulating the temperature in reactors with different geometries and affected by different climatology, eventually guaranteeing a powerful generalization potential and a faster recalibration (Gharib et al., 2025). Given its reduced parametrization, this approach is also suitable for automatic control applications.

3.5.3. Gas-liquid exchange

The most common way of simulating gas-liquid exchange is by implementing Fick's law, as reported in Eq. 27 and 28 for O_2 and CO_2, respectively:

$$Q_{O2} = k_L a_{O2}(T)[H_{O2}(T)p_{O2} - S_{O2}] \quad (27)$$

$$Q_{CO2} = k_L a_{O2}(T)\left(\frac{D_{CO2}}{D_{O2}}\right)^{0.5}[H_{CO2}(T)p_{CO2} - S_{CO2}] \quad (28)$$

where Q_{O2} and Q_{CO2} are the mass transfer rates of O_2 and CO_2, respectively (e.g., $g.m^{-3}\cdot d^{-1}$); $k_L a_{O2}(T)$ (e.g., in day^{-1}) is the global oxygen mass transfer coefficient, depending on temperature following the Arrhenius exponential law; $H_{O2}(T)$ and $H_{CO2}(T)$ are the Henry constants with temperature dependence for O_2 and CO_2, respectively (e.g., $g.m^{-3}\cdot atm^{-1}$); p_{O2} and p_{CO2} are the gaseous O_2 and CO_2 partial pressures at the interface (e.g., atm); S_{O2} and S_{CO2} are the dissolved concentrations of O_2 and CO_2, respectively (e.g., $g.m^{-3}$); D_{O2} and D_{CO2} are the diffusivity coefficients for O_2 and CO_2, respectively ($m^2\cdot s^{-1}$). This approach (Eq. 28) can be extended to other components that can be transferred to the gaseous phase (e.g., NH_3, H_2 and N_2O). It is of outmost importance to reliably estimate the mass transfer coefficient, as it drives the overall stripping/dissolution rates. Different approaches can be carried out to evaluate the mass transfer coefficients: i) with dedicated experiments, measuring $k_L a_{O2}$ and follow the mass transfer model from Eq. 27 for the other gasses; ii) with dedicated experiments, determining all mass transfer coefficients; iii) from online recorded data, calibrate the mass transfer coefficient to match model predictions and dissolved gases (or pH) measurements. For this, online measurements at high frequency of the specific dissolved component must be available. Note that identifiability and correlation problems can occur when many biological parameters must be simultaneously calibrated. Therefore, an experimental determination is strongly recommended (see also Section 4). This is even more critical when air injection and pH control systems are active. In these cases, $k_L a$ values, driven by the reactor specific gas sparging system, are typically much higher compared to the ones describing the natural exchange with the atmosphere, which is mainly influenced by the reactor mixing system.

Mass transfer of CO_2, similarly to O_2, is primarily diffusion-driven, allowing the use of the diffusion coefficient ratio to adjust $k_L a_{O2}$ in relation to CO_2. On the contrary, both NH_3 and N_2O present specific challenges due to their specific behaviour in aqueous environments. The stripping of NH_3 is considerably more complex because its speciation between ammonia and ammonium is extremely sensitive to pH. Thus, the mass transfer of NH_3 is driven by both diffusion and the chemical equilibrium between the gaseous and ionic forms. Furthermore, NH_3 is much more volatile than CO_2, so its mass transfer is not limited by diffusion in the same way and stripping rates that are not adequately captured by a simple diffusion-based correction.

For reactors relying on blowers, estimating the $k_L a$ of N_2O for representing the exchange with atmosphere based on Eq. 27 is relevant for modelling N_2O gas-liquid exchange processes, especially considering that phototrophic bioreactors are relatively shallow (Chrysochoidis et al., 2024). However, factors such as interactions with other species in solution or reactor geometry can affect model predictions. Even if the prediction uncertainty of the flow seems limited, the impact (e.g., on climate change effects) may become large. Thus, alternative stripping models may be needed (Chrysochoidis et al., 2024). For open pond reactors, where mixing relies on paddlewheels, $k_L a$ N_2O shows strong variability as function of the distance to the mixing area. Thus, compartmental models or CFD models become instrumental to properly model emissions (see Section 3.5.1). To keep simplicity, complex hydrodynamic models can be used to infer meta-models that can predict overall reactor emissions in the open pond (Qiu et al., 2024). Given the complexity of modelling microbial N_2O production, approaches employing machine learning techniques are emerging, which lump both microbial and physic-chemical aspects (Hwangbo et al., 2021; Li et al., 2022).

3.6. Integration and selection of modelling frameworks

The comparison among different phototrophic modelling frameworks is a very useful practice and must be encouraged (Nordio et al., 2025), since it helps in evaluating the most suitable model according to a specific case (climatology, type of reactor and operational strategies applied), and to the target function to be optimized (e.g., maximizing the biomass productivity, the bioremediation efficiency, minimizing greenhouse gas emissions, or optimal control studies). Since all mathematical models are a simplified representation of the reality, there are no models better than others in absolute terms. To guide developers in selecting model structures and complexity levels, Fig. 5 provides a practical workflow based on:

- modelling objectives (process control, system optimization, mechanistic prediction)
- data requirements (online sensors, environmental inputs, omics data)
- integration with chemistry/physics (CFD, heat and/or mass transfer, chemical speciation)
- validation steps (sensitivity analysis, calibration, testing and uncertainty analysis)

In the perspective for scaling up the system, it is important that the model can be interfaced and integrated within available plant-wide modelling frameworks (Carecci et al., 2024). This is important because photobioreactors, whether they are the main biological treatment step or integrated in complex wastewater resource recovery facilities, are always combined with other unit operations needed for

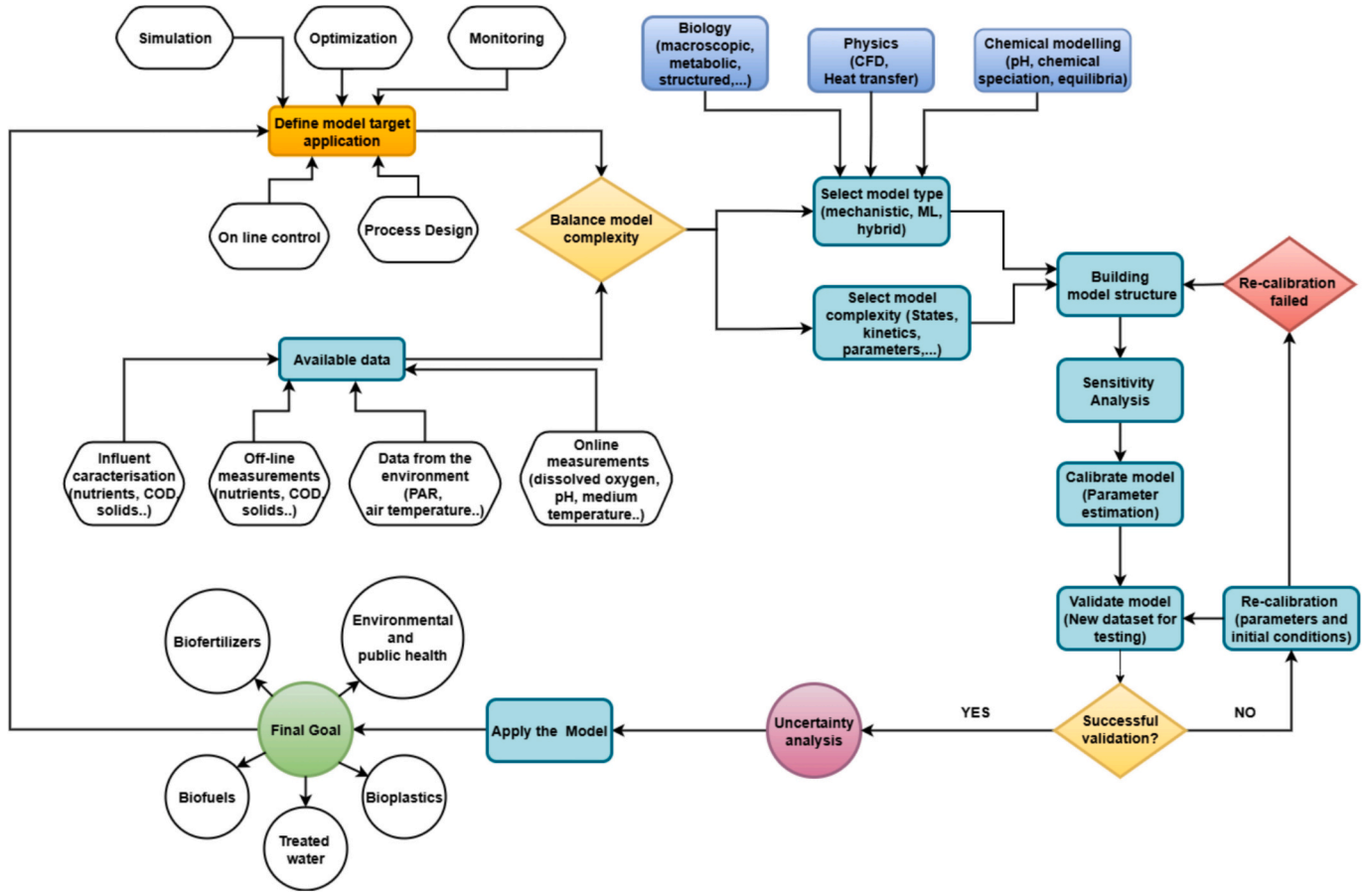

Fig. 5. Practical workflow for phototrophic model selection. The diagram outlines key decision steps from problem definition to model application, covering: i) objective identification (monitoring/optimization/simulation), ii) data and system characterization, iii) model type selection (mechanistic/data-driven/hybrid), iv) integration with physical/chemical frameworks, and v) validation procedures.

wastewater pre-treatment and biomass harvesting.

4. The importance of data collection for mathematical models

4.1. Experimental data across multiple timescales

Accurate and useful mathematical models cannot be developed without high-quality data. However, monitoring campaigns are often not designed with the specific goal of building and validating a model, resulting in missing or insufficient essential measurements. This section highlights how a monitoring campaign should be strategically organized to ensure the acquisition of relevant and comprehensive data, supporting the development of a robust mathematical model that can effectively capture process dynamics and support the system optimization.

Processes involving phototrophs or mixed microbial communities exhibit a wide range of dynamics: fast of the order of seconds (chemical reactions), minutes or hours (e.g., photosynthetic responses to light changes, gas/liquid transfer, evaporation), and slower of the order of days (microbial growth and decay). Therefore, an effective monitoring should integrate online sensors for fast-changing variables (e.g., temperature, dissolved oxygen, pH, light intensity) and offline methods for slowly changing variables (as biomass and nutrient concentrations).

A data acquisition campaign for model development must consider environmental and process constraints (such as environmental conditions, influent composition, reactor configuration, and operational mode) and suspension characteristics (e.g., microbial biomasses and substrate/product concentrations), including factors that influence microbial growth over time (e.g., reactor temperature, dissolved oxygen, and inhibitor concentrations).

4.1.1. Reactor characterization

Reactor fluid dynamics affect process efficiency, with poor mixing leading to strong spatial heterogeneity, particularly for large facilities. For instance, full-scale raceway ponds exhibit long mixing times (hours), poor vertical mixing, and plug-flow behaviour in long channels (Mendoza et al., 2013). To accurately identify these dynamics, tracer tests should be performed before the reactor start-up to characterize the RTD (Levenspiel, 1998). Gas transfer coefficients (k_La) are estimated through re-aeration tests, often carried out in clean water, where dissolved oxygen is removed using reducing chemicals (e.g., sodium sulphite; Casagli et al., 2021a, 2021b). However, mass transfer is affected by the wastewater matrix and k_La from clean water tests needs correction via the alpha factor (Chrysochoidis et al., 2024). Ideally, re-aeration tests are conducted in active reactors in dark conditions (e.g., night) where photosynthetic activity is negligible, following, e.g., the "gas out – gas in" method (Tribe et al., 1995). When possible, this test should be carried out for each gas of interest, instead of correcting the k_La based on oxygen (see Section 3.4.3) (Barceló-Villalobos et al., 2018). However, these approaches can be challenging at industrial scale. Probe location in non-homogenous reactors can impact k_La estimation (Mendoza et al., 2013). Thermal models also require characterization of reactor geometry and materials properties (e.g., specific heat, thermal conductivity).

4.1.2. Monitoring influent and effluent compositions

Accurate characterization of influent and effluent streams is essential to ensure reliable mass balances. It is the cornerstone for verifying the continuity of key elemental components (C, H, O, N, P) and COD. It is also determinant for efficiently training data-driven models, and constrain metabolic networks in metabolic modelling. However,

Table 3
Summary table for a comprehensive data collection plan for a continuously operated reactor.

Class	Measurement type	Frequency
Preliminary tests/data collection	Tracer test k_La assessment Geometry and materials	Before reactor start-up and each time relevant changes are made to the reactor configuration
Influent characterization	Liquid inflow: Total/soluble COD, total/soluble BOD_{20}, TKN, total P, oligo-elements and heavy metals (when relevant), TSS, VSS, VS, TS, pH value, total concentration of relevant acids and bases (TIC, N-NH_4, N-NO_2, N-NO_3, P-PO_4, VFA), alkalinity, salinity, ions involved in precipitation; optical density, temperature. Flow rate. Gaseous inflow: CO_2, O_2, other gases partial pressure and flow rate (when relevant). Inoculation rate and concentration (when relevant)	Depending on the influent variability
Effluent characteristics	Liquid flows: soluble COD, soluble BOD_{20}, TIC, N-NH_4, N-NO_2, N-NO_3, P-PO_4, VFA, alkalinity, salinity, ions involved in precipitation, OD_{680}, OD_{750} (for indirect biomass concentration assessment), flow rate. Gaseous flows: CO_2, O_2, (NH_3, N_2O, N_2 if relevant) partial pressure and flow rate.	Typically, 2–3 times per week, or 3–4 measurements distributed over the hydraulic retention time
In-reactor measurements	Total COD, total BOD_{20}, TKN, total P, TSS, VSS, VS, TS, cell counts, biomasses concentration (cell counts, cytometric, fluorimetric measurements), biomass elemental composition	Typically, once per week, or 1–2 measurements distributed over the hydraulic retention time
	DO, incident light intensity, temperature, pH value, turbidity.	At least 1 per hour for outdoor reactors
	Microbial biomass composition and diversity	Depending on influent variability and solid retention time
Other boundary conditions	For indoor reactors: lighting strategy, temperature regulation	To catch variability and specific conditions
	For outdoor reactors: air temperature, solar radiation at the ground level, wind speed, and humidity, rain events	Hourly

influent properties are often inadequately characterized compared to effluents, despite their critical influence on process dynamics. Influent and effluent characteristics must include: i) concentrations of substrates for biological processes, organic matter fractions and related biodegradability, inorganic nutrients and carbon, specific inhibitors (when known and relevant), volatile and total and suspended solids; ii) data for physical-chemical models, including pH, alkalinity, the total concentrations of all species included in the chemical model, salinity, and ions potentially involved in precipitation processes; iii) physical characteristics, such as temperature and optical properties. Ideally, this information should be available in the reactor suspension, in the influent and effluent streams. For resource recovery, macro-molecular or elemental characterization is often needed, although its frequency for monitoring is usually limited by cost. Similarly, microbial diversity monitoring may be needed for accurate prediction of the system dynamics, including multi-omics for metabolic models. When using a solid/liquid separation system to decouple the hydraulic and biomass retention times, both clarified effluent and biomass wastage flows must be monitored (e.g., Robles et al., 2020; Crouchett-Catalán et al., 2025).

In the worst case, when this information cannot be directly measured, specific datasets on the chemical composition of common algal strains are available (e.g., the Phyllis database). Open-access data on wastewater composition can also be found in the literature (e.g., Tchobanoglous et al., 2014; Henze et al., 2008) and in standardized wastewater treatment plant models such as the Benchmark Simulation Model (Copp, 2002). However, these databases are not sufficient in both quantity and quality for being used as reliable model inputs. Indeed, the mathematical characterization of the influent wastewater and the algal and bacterial biomass constitutes a major challenge for modelers, as these data are often missing or insufficiently collected during experimental campaigns. Influent composition is of fundamental importance and must be characterized for each specific case study, as relying on literature data referring to averaged values or other case studies leads to parameter identifiability issues and non-reliable models. This is particularly true when, in addition to the biological model, a chemical submodel is also included, where highly specific input and initialization data, such as system alkalinity, must be provided. A summary of data to be collected with suggested frequency for photobioreactors is reported in Table 3.

4.1.3. Recording environmental conditions

Worldwide meteorological data, essential for modelling outdoor photobioreactor, are accessible through several platforms, such as the Photovoltaic Geographical Information System (PVGIS) developed by the EU Science Hub, or the Prediction Of Worldwide Energy Resources (POWER) developed by NASA. However, meteorological data are often recomputed by a model or recorded by weather stations located away from the site of the reactor and may not accurately represent the local weather conditions. The accuracy of the energy flux on the process is fundamental to predict its fate, especially through appropriate characterization of light and temperature. Other critical factors like relative humidity and wind speed can also have deep impact on evaporation, temperature dynamics and gas exchange. The spatial discrepancy between the location of the weather station and the photobioreactor can introduce uncertainties in the model, affecting significantly the accuracy of simulations and the overall performance assessment of the system. In addition, meteorological data can be used to model water evaporation which can have a non-negligible impact on open processes like raceways (Béchet et al., 2018). Therefore, dedicated meteorological measurements are highly recommended to be accounted for during the development phase of the model. The databases can moreover be used for scenario predictions, such as simulating process upscaling in different locations or for exploring optimization strategies.

4.2. Methods for monitoring phototrophic systems

4.2.1. Conventional parameters in wastewater systems

Monitoring phototrophic systems shares similar analytical procedures with activated sludge systems for both offline and online parameters, particularly referring to solids, off-gases and water quality monitoring, most of them being well documented (Silkina et al., 2015; Rieger et al., 2013; van Loosdrecht et al., 2016). Some data can be monitored using online probes, providing valuable insights on fast and highly dynamic conditions, supporting advanced model calibration. It is even crucial when employing machine learning techniques, to reach the very large amount of data required by advanced artificial intelligence approaches. The most widespread on-line sensors measurements include (Robles et al., 2020):

- dissolved oxygen concentration, influenced by photosynthesis, chemotroph and phototroph respiration, and gas/liquid transfer. This highly informative online measurement captures both fast dynamics (e.g., high frequency deriving from environmental solar light variation, e.g. clouds, and long-term trends, e.g., diurnal/seasonal cycles);
- pH, affected by ionic strength, buffering compounds, gas/liquid mass transfer (CO_2, NH_3) and microbial activity (e.g., ammonium, nitrate,

phosphate, organic acids, and inorganic carbon dynamics). Changes in pH affect salt precipitation, biomass growth kinetics, and pH-based CO_2 bubbling control, making pH variation and control dynamics highly informative and essential to record. In case pH is regulated, the quantity of injected CO_2 (or added acid) must be recorded.

- turbidity and/or optical density (at various wavelengths in the visible, as 420, 668, 680 and 750 nm, or in the NIR range) used as proxy of total or microbial biomass. Commercially available online sensors are designed for bioprocess monitoring and are particularly useful in batch growth tests or highly dynamic feeding conditions. While useful, application at different wavelengths for biomass or pigment monitoring may be challenged at high turbidity from wastewater. These parameters also help estimating dynamic light attenuation and penetration.
- substrates, with online sensors commercially available for ammonia, nitrate, nitrite, dissolved CO_2, or dissolved organic carbon.
- off-gasses, with floating chamber sensors, suited to monitor gas flow and composition, including CO_2, N_2O and NH_3 (Marques et al., 2016).
- Medium temperature is an important parameter, and it is often included in combination with other parameters in commercially available probes.

When operating indoor reactors, heating and lighting (at specific wavelengths) conditions must be known. For outdoor reactors, meteorological data (air temperature, solar radiation, wind speed, weather events, and humidity) should be measured at least hourly to capture daily dynamics.

4.2.2. Light monitoring

The solar flux is crucial for photosynthetic organisms since it fuels a large fraction of the microbial biomass in the process. It is therefore particularly important to measure the PAR, together with the flux of photons in the NIR spectrum for purple bacteria. The development of radiative transfer models requires determining incident radiative fields and suspension optical properties. Suspension optical properties, essential for modelling the light distribution, are often simplified using a single extinction parameter via spectrophotometry (Beer-Lambert law). For a more in-depth description of the light distribution, inherent optical parameters must be measured (e.g., absorption and scattering coefficients, phase function; Wágner et al., 2018; Amini et al., 2025). The spectral information of how light is distributed in the reactor is rarely exploited. However, it illustrates the strong consumption of red and blue photons and indirectly traces the concentration of the various pigments. Spectro radiometric measurements taken at different locations in photobioreactors, either on external surfaces or inside the reactor provide indeed a rich information. While external surface measurements are generally used for radiation transfer models, mapping light spectrum inside the reactor is very useful to calibrate and validate models predicting light distribution and relating it to growth (de Mooij et al., 2016).

4.2.3. Monitoring the microbial diversity and biomass composition

While total biomass can be approximated with optical sensors (typical by measuring absorbance at 750 nm), or suspended solids, both total and volatile measurements, microbial diversity is more challenging to track. Approximate estimations of various microbial species can be achieved using molecular biology tools via amplicon sequencing (16S rRNA genes for bacteria and 18S for microalgae and protozoa), DNA microarray, qPCR, and metagenomics (Bradley et al., 2016; Garlapati et al., 2019). Targeting each function (e.g., photosynthesis, nitrification, denitrification) is challenging, but would allow to validate the microbial dynamics, which are the core of the process.

However, there are alternative methods easier to automate based on colorimetric methods, fluorometry or spectral methods. Absorbance or optical density have the advantage of being rapid and non-destructive. Chlorophyll-specific wavelengths are used for microalgae and cyanobacteria (660–690 nm) and picked according to the strain-specific absorption spectrum (Santos-Ballardo et al., 2015). Exciting chlorophyll in the blue (around 420 nm) and measuring fluorescence in the red (680 nm) is another approach, but it can be rapidly perturbed by interferences with organic matter. Those measurements can be collected online via commercial or home-made sensors (e.g. Nguyen and Rittman, 2018; Barbosa et al., 2020). Nevertheless, in case the wastewater matrix has high turbidity or colour, accurate pigmentation monitoring is challenging or even unfeasible.

Microscopy has been a popular method, being light microscopy most commonly used, since it offers a low-cost and non-invasive method to enumerate microbial cells (Wágner et al., 2021; Chong et al., 2023). Besides offline analyses of samples at a stationary microscope, online, inline or at-line methods are possible through the implementation of bypass systems or by direct integration of a camera (Havlik et al., 2013). Cell counts are largely used, based on optical/manual procedures, automated cell-counters or image recognition tools (Takahashi, 2018). However, cell biovolume can change within the same strain according to the growth phase and environmental conditions. Microscopic observations can inform on morphological changes and quantify cell metabolites such as pigments, lipids or starch (Guo et al., 2017; Ota and Kawano, 2019). Technological advances in camera hardware and digital image processing are the main drivers of innovation in this field (Jeckel and Drescher, 2021). There exist databases with images (e.g., ImageNet) which can be used to pre-train classification models before training them on specific images of the system of interest, for which one may have limited amount of good quality data (referred to as transfer learning; Kornblith et al., 2019; Deng et al., 2009).

Pulse-amplified modulated (PAM) fluorescence measurements (Mayer et al., 1997; Gregor and Maršálek, 2005) yield valuable information about the fate of the energy of light absorbed by the light harvesting antenna, reflecting its photochemical utilization and thermal dissipation. These characteristics make fluorescence analysis attractive for online monitoring of the physiological condition of cultivated microalgae (Havlik et al., 2013; Havlik et al., 2022). Biomass composition can be targeted by applying techniques, such as nuclear magnetic resonance spectroscopy (Bisht et al., 2021), near infrared reflectance spectroscopy (Morgado et al., 2024), Fourier transformed infrared spectroscopy (Wagner et al., 2010) or electrochemical impedance spectroscopy (Al Ahmad et al., 2020). These methods can also be developed into online monitoring systems. For more insights about online and soft sensing to characterize biomass composition in phototrophic systems, the reader is referred to specialized reviews (Podevin et al., 2017; Havlik et al., 2022).

4.2.4. Integrating multi-omics data into metabolic models

The progress in molecular tools and bioinformatics has been instrumental in characterizing microbial communities and better understanding the role of different microbes in engineered systems. Omics analyses (e.g., metagenomics or transcriptomics) offer a snapshot of a microbiome composition and potential activity, useful for metabolic modelling tools. In particular for metabolic models, metagenome assemble genomes (MAG) or genomes available from databases are data needed for model formulation using available software (e.g., CarveMe, Machado et al., 2018). Generated models need to undergo different sanity checks based on different tests and toolboxes (e.g., COBRA toolbox, Heirendt et al., 2019) to check their basic properties before being used. Other tools, such as transcriptomics, fluxomics, or metabolomics, can be applied as model constraints (i.e., signalling metabolic routes that are active) when solving the optimization problem posed through FBA.

4.3. Handling data uncertainties

Measurements noise can deeply impact the parameter identification step, and model accuracy. Beyond uncertainties derived from the instruments and their calibration protocols or ranges of operability, data sets often present outliers or gaps that need to be treated carefully. There is a diverse set of statistical methods used for error detection and reduction of input uncertainty widely tested for bacterial based systems (Belia et al., 2021). Statistical process control techniques involve monitoring process variables over time using statistical control charts to detect abnormal operational periods, either because of process or instrument malfunction, that should be avoided for model calibration. The variables of interest are visualized over time and compared to control limits to determine if the process is within control (i.e., its correlation structure is unchanged). However, due to phenomena such as autocorrelation, seasonality, and non-constant variance typical of water treatment systems, including phototrophic systems, it can be difficult to apply traditional control charts. Alternatively, it is recommended to fit time-series models (e.g., auto-regressive integrated moving average or exponentially weighted moving average) to normal operating data and using it as charting tool. Since this is a univariate method, it fails to consider correlation between process variables. Thus, multivariate statistical techniques, such as principal component analysis or partial least squares, are better suited to monitor process data (Mašić et al., 2015). The generated models are more efficient detecting outliers deviating from normal operation.

Data reconciliation adjusts process measurements so that they are consistent with known conservation laws and other process constraints. It allows closing the mass balances for the system considered, identifying measurements error, supporting estimation of missing data and/or identification of significant biological/chemical/physical reactions in certain units. Yet, in most cases these techniques have been limited to flow rate and traditional pollutants, such as COD or nutrients. This method is particularly useful in a plantwide context, helping to close the balances across the treatment train. Main steps during data reconciliation are detailed on state-of-the-art studies on modelling of conventional wastewater treatment processes (Monje et al., 2022; Le et al., 2022).

Finally, the quantity of data needed for model calibration depends on the quality of the data (Dochain and Vanrolleghem, 2001). Therefore, extending sampling campaigns do not necessarily reduce the uncertainty of parameter estimates. There are a variety of methods for optimal experimental design and data series selection to increase parameter identifiability.

4.4. Sharing data for model benchmarking: Balancing open resources and industrial constraints

The availability of data is of crucial importance for benchmarking different models on the same process (Nordio et al., 2025) to support process scaling up from lab or pilot scale to commercial units, using records from other plants together with their environmental conditions (Benner et al., 2022).

Despite the availability of open-source data, there is limited access to data from industrial-scale applications. In fact, proprietary constraints imposed by private companies often limit the availability of key operational data, creating a major barrier to the development of accurate and scalable models.

Several studies have already emphasized the urgent need for stronger collaboration between industry and academia to facilitate data sharing (Muñoz and Guieysse, 2006; Hülsen et al., 2014). The integration of industrial data into academic research frameworks could significantly enhance model reliability and applicability, ensuring that models of phototrophic systems are representative of real-world operational conditions. However, proprietary restrictions often prevent the disclosure of crucial process parameters, creating a barrier to innovation in WWT technologies.

To overcome these limitations, it is imperative to establish structured mechanisms for data exchange, while ensuring the protection of intellectual property rights. Encouraging partnerships between industry and research institutions could enable the development of more robust and validated models, facilitating the scaling up of innovative phototrophic based bioprocesses for WWT.

5. Parameter estimation and identifiability

Model calibration, also called parameter estimation, is the cornerstone for guarantying an accurate model prediction capability, where the model outputs match the experimental records. A first difficulty is that not all the state variables are recorded (e.g., the microbial diversity is rarely measured) and some quantities are related through complex nonlinear relationships to the state (e.g., pH). Furthermore, for mechanistic models, the parameters generally have a physical or biological meaning, which can provide interesting information. For example, a high value of the parameter representing the mortality rate of the microalgae population can reveal a viral issue or grazing by zooplankton. However, not all models can have their parameters unambiguously determined from the data, either because of the model structure or because of data availability (both quality and quantity; Wieland et al., 2021). Sometimes different parameter values can lead to the same predictions, making it impossible to derive a value which can be physically or biologically interpreted. This issue is known as the identifiability problem and can be particularly important in the context of complex systems where models may have correlated parameters and nonlinear dynamics (Gábor et al., 2017; Walter and Pronzato, 1997).

The calibration process consists in manipulating the model parameter values for minimizing the discrepancy between data and predictions. The next stage is the overall modelling validation, by comparing the model outcomes with an independent dataset. These steps are detailed hereafter, and we invite the reader to refer to Yahia et al. (2015) or Mairet and Bernard (2019) for more support on bioprocess modelling.

5.1. Theoretical parameter identifiability

A way to mathematically prove identifiability is to show how the parameters can be uniquely computed from the set of measurements and their time derivatives up to any order (Raue et al., 2014). This involves complex mathematical analyses which can be very challenging, especially for models of high state and parameter dimensions, motivating the need to keep parametrization as simple as possible. We refer to Raue et al. (2014) for constructive methods and software to test identifiability. Identifiability problems can appear for many possible reasons. Often, this is due to parameter redundancy or an over-parametrization involving artificially too many parameters. For example, it will happen when two parameters α and β always appear in a product term (α•β). Any of their combination with the same product will eventually provide the same results. This case is rather simple to detect. There are more complex cases for which a more careful analysis is necessary. We give here two examples which are emblematic of photosynthetic organisms' models. The first example is a kinetic model often appearing in photosynthetic models to represent growth with inhibition at high substrate (S) concentration, including light:

$$\mu(S) = \overline{\mu}\frac{S}{S+K_S}\frac{K_I}{S+K_I} \tag{29}$$

where the constant $\overline{\mu}$, K_S and K_I are often (wrongly, as it is explained in Section 6.2) interpreted as maximum growth rate, half saturation constant and inhibition constant, respectively. The point is that two triple of parameters can provide the same results. As a matter of illustration, the reader can evaluate Eq. 29 for the two triples ($\overline{\mu}$, K_S, K_I) valuing either (2,2,1) or (1,1,2).

Here, we detail another case that appears in many photosynthetic models, resulting from the combination of several parameters which have similar effects. Let us assume that light intensity in the medium is strongly attenuated (as often in turbid wastewater). In such a case, it is crucial to accurately represent growth at low light. A linear model can do this job:

$$\mu(I) = a\,I \tag{30}$$

where a represents the photosynthetic efficiency ($m^2.\mu mol^{-1}$) and I is the light intensity (in $\mu mol.m^{-2}.s^{-1}$). The classical models (Monod, Haldane or Steele; Béchet et al., 2013) are all equivalent to this simple model for low light intensity. Often, since the average light is lower than 15 % of the incident light (Bernard et al., 2015) the linear model will provide similar results to more complex models (Fig. 3).

Most of the models assume an exponential light decrease (with an extinction coefficient K), due to absorption and scattering, from the surface exposed to a light intensity I_0 to the depth z (Eq. 3). It results that a simple way to assess the average microalgal growth rate $\overline{\mu}$ is to assume proportionality with the average light intensity in the medium $\overline{I}$:

$$\overline{\mu}(I) = a\,\overline{I} \tag{31}$$

The average light intensity can be computed from the surface light I_0, the culture depth L and the extinction coefficient (Eq. 4; Bernard, 2011). So that, combining Eq. 31 and Eq. 4:

$$\overline{\mu}(I_0) = a\,\frac{1-e^{-KL}}{KL} I_0 \tag{32}$$

This computation (Table 4) illustrates the fact that, for a linear growth model, various combinations of K, L and a can provide the same average growth rate. This is structural for a linear model but stays valid for the classical models representing photosynthetic rates (see section 3) provided that the turbidity is high, so that the average light stays low. We refer the reader to Mairet and Bernard (2019) for other simple examples illustrating identifiability issues.

Structural identifiability is a fundamental concept in systems modelling that assesses whether the model parameters can be uniquely identified from input-output data, considering an idealised scenario without noise. A non-identifiable model will lead to convergence troubles for many automatic identification algorithms with the optimizer jumping between the different equivalent parameter sets.

5.2. Parameter meaning

While it may appear straightforward, the naming of parameters in biotechnological models can often be ambiguous. In some cases, the terminology used for certain parameters might lead to confusion, potentially resulting in calibration errors. We provide three examples illustrating this point.

Let us come back to the simple model of growth with inhibition given by Eq. 11. In such model, the parameter K_S is often presented as the "half saturation constant" (see as an example Rossi et al., 2025). Yet, for $S = K_S$ the growth rate is not half of the maximum value. Additionally, the maximum value of this function is reached for $S^* = \sqrt{K_S K_I}$, yielding:

$$\mu_{max} = \overline{\mu}\frac{\sqrt{K_I}}{\sqrt{K_I} + 2\sqrt{K_S}} \tag{33}$$

Thus, $\overline{\mu}$ should be derived from Eq. 33 and not considered the maximum observed growth rate. It is also interesting to note that the model with parameters ($\overline{\mu}, K_S, K_I$) and the model with parameters ($\overline{\mu}\sqrt{\frac{K_I}{4\,K_S}}$, $K_I/4, 4K_S$) have very similar behaviours.

Another example is the meaning of the Droop model representing growth rate as a function of an internal quota Q:

$$\mu(Q) = \overline{\mu}\left(1 - \frac{Q_{min}}{Q}\right) \tag{34}$$

where Q_{min} is the minimal cell quota of the limiting element. On the contrary to what is often stated (Rossi et al., 2025), the maximum growth rate is not $\overline{\mu}$, but computed from the maximum cell quota Q_{max}:

$$\mu_{max} = \overline{\mu}\left(1 - \frac{Q_{min}}{Q_{max}}\right) \tag{35}$$

Eq. 12 is the original and most adapted way of writing the Droop parameters.

Overall, it is important to state that the meaning of a parameter results from the model itself, and not from the a priori denomination of a parameter. This shall not affect the parameter identifiability, but will result in a wrong parameter value in the model.

5.3. Experimental data to support parameter estimation

Once identifiability of the model parameters has been demonstrated, calibrating them is the second step. Parameter identification combines steps of very different natures to calibrate the parameter values of the model. The simplest approach consists in finding parameter values from literature studies. Yet, the experimental conditions can be significantly different from the ones for which the model is developed, compromising model accuracy. There is a large range of situations from controlled laboratory experiments with constant light to outdoor conditions where light and temperature fluctuate. Hereafter follows a ranking of experimental conditions which can be used to calibrate some model parameters from simplified and constant conditions to realistic ones:

i. lab experiments in steady conditions with organisms similar to the ones of interest;
ii. lab experiments in periodic conditions with organisms similar to the ones of interest;
iii. lab experiments with samples from the outdoor reactor and in realistic conditions;
iv. calibration experiments directly run in the outdoor reactor with an experimental protocol designed to improve and facilitate parameter identification;
v. experimental campaigns in the outdoor reactor;
vi. other experimental campaigns, obtained in different reactors and for other environmental conditions.

For all these experiments, we strongly recommend that the same model is used across stages. The parameters are easier to identify from lab experiments, where operational conditions can be modified one at a time, but they will be more trustable if they have been derived from conditions closer to the ones for which the model is developed. As example, working in the lab with continuous light will give very different results than with populations synchronized by time and temperature (Bonnefond et al., 2016). The parameters obtained at level *iv* and after will ensure a model generalization to several conditions and processes, often at expenses of less accurate predictions for the different considered cases (Nordio et al., 2025). As complex conditions tend to be noisy, more data are required to extract significant effects. Before

Table 4
Possible set of parameters which produce the same resulting average growth rate with a linear model ($\mu(I) = a\,I$).

a	2.373	1.8199	3.4696	2.8962	2.3730	1.9061
K	10	4	20	15	5	10
L	0.1	0.1	0.1	0.1	0.2	0.05
$\overline{\mu}(I_0) = a\,\frac{1-e^{-KL}}{KL} I_0$	1.5 I_0	1.5 I_0	1.5 I_0	1.5 I_0	1.5 I_0	1.5 I_0

collecting these data, pilot or in-silico studies can be carried out to estimate their required amount or precision.

5.4. Estimation procedure

5.4.1. Parameter subset selection

In practice, especially for more complex models, model calibration uses a combination of various types of datasets (Casagli et al., 2021a; Casagli et al., 2021b). In general, it is not possible to minimize the cost function by modifying all the parameters and a subset of them must be selected, for example using global sensitivity analysis. The parameters with the lowest sensitivity will be taken from literature. The remaining ones can be directly inferred from experiments or estimated through model calibration (see Section 5.4.2).

5.4.2. Calibration algorithms

The simplest way to estimate model parameters entails the manual trial and error, whereby one starts with an arbitrary parameter value, and then makes incremental changes thereof in an iterative way until the model reasonably fits the experimental observational data. Despite being a subjective method, given its simplicity, this technique is still used in industry and research.

There are numerous formal statistical methods, based on the minimisation of a cost function, to estimate parameters (Wenk et al., 2019), e.g., frequentist framework (FF), maximum likelihood methods (MLM), least-square methods (LSM), covariance matrix method for stochastic models (CMS) or Bayesian methods (BM). For complex cases, with multiple local optima, heuristic algorithms such as genetic or particle swarm algorithms enable to explore more widely the parameter space (e.g., Hacizade, 2019). Particle swarms are versatile and easy to use, even if fine-tuning is required to avoid premature convergence (e.g., Freitas et al., 2020). The least-squared error (Hauduc et al., 2015) is the more widespread cost criterion. In this case, the algorithm finds the value of the parameter set θ_i minimizing the root mean square normalized error (RMSNE), i.e., the distance between the simulations of variable j at time t_i, denoted $y_j(\theta, t_i)$ and the measured value $y_{m,j}(t_i)$.

$$RMSNE(\theta) = \sqrt{\frac{1}{n}\sum_{i,j=1,1}^{n,q}\left(\frac{y_j(\theta,t_i)-y_{m,j}(t_i)}{\sigma_j}\right)^2} \tag{36}$$

When several measured quantities (e.g., pH, COD or chlorophyll) are used to setup the cost function, it is recommended that the error for each quantity is weighted by the typical standard deviation σ_j associated with the measured protocol.

5.4.3. Determining parameter uncertainty and covariance

While structural identifiability is a purely mathematical property which considers ideal measurements without noise, practical identifiability assesses how much the model outputs are depending on the parameter values and if the information contained in the data is rich enough to accurately estimate the parameters values. When a parameter does not affect much the model predictions, we cannot estimate its value. Reciprocally, even with a very inaccurate value, the estimated parameter has no impact. Sensitivity functions are involved in practical identifiability together with data accuracy and richness (Wieland et al., 2021). However, as sensitivity is computed parameter by parameter, it is not possible to detect a structural identifiability problem from the sensitivity analysis.

Measurement noise introduces multiple challenges in parameter calibration. While noise does not inherently alter the convexity of the optimization problem (e.g., linear regression remains convex even with noisy data), a key concern arises when the cost function becomes non-convex due to system nonlinearities or structural identifiability issues. In such cases, the optimization landscape may contain numerous local minima, potentially trapping algorithms in biologically implausible solutions. Combined with measurement noise, it induces a slow training convergence or even prevent it to reach a reasonable parameter set. Parameter initialization is determinant to an efficient convergence of the minimisation algorithm. It should be initialised with literature values, or, better, from values derived from lab experiments in simplified conditions. By running the minimisation algorithms many times, from an identified pre-selection of initial parameter values, local minima can be identified and disregarded. Latin hypercube sampling (LHS) can be used to create combinations of parameter sets within pre-defined parameter ranges (Helton and Davis, 2003). Cost function, e.g., RMSNE, thresholds can be arbitrarily used to select the parameter subsets (local minima) prior to normality check (Wágner et al., 2016). The practical identifiability can be established based on the normality and shape of the histograms derived from parameter estimates associated to local minima with comparable RMSNE. That is, quasi-Gaussian and narrow histograms (i.e., low parameter uncertainty) indicate more effective practical identifiability (Van Daele et al., 2015). Additionally, parameter correlation should be non-significant to practically identify parameters. The matrix of covariance indices ($\sigma_{(\theta_A\theta_B)}$), calculated for every parameter combination (θ_A and θ_B), where $\mu_i$denotes the mean estimated posterior θ_i, indicate to what extent two model parameters decrease or increase in tandem.

$$\sigma_{(\theta_A\theta_B)} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\theta_A-\mu_{\theta_A}\right)\left(\theta_B-\mu_{\theta_B}\right) \tag{37}$$

Additionally, correlation can indicate both the strength and the direction (negative/positive) of the impact. The correlation indices are calculated as

$$\rho_{(\theta_A\theta_B)} = \frac{\sum_{i=1}^{n}\left(\theta_A-\mu_{\theta_A}\right)\left(\theta_B-\mu_{\theta_B}\right)}{\sqrt{\sum_{i=1}^{n}\left(\theta_A-\mu_{\theta_A}\right)^2\left(\theta_B-\mu_{\theta_B}\right)^2}} \tag{38}$$

High correlation index ($\rho_{(\theta_A\theta_B)} > 0.5$) between parameters requires further analysis. If the variability of each parameter estimate does not lead to big impacts on the model output, the correlation between parameters does not compromise their identifiability. However, if the variation of one parameter within its confidence interval compromises the model accuracy, so the correlated parameter needs to be modified to compensate, the parameters are not identifiable. Practical identifiability is still a delicate issue for which tools are lacking (Wieland et al., 2021). Simplifying the model or focusing on a smaller parameter subset is the solution to reduce this risk.

5.4.4. Initialization of state variables

For the parameter identification approaches relying on cost function minimization, accurate initial conditions of state variables are essential. However, complex systems involving unmeasured quantities (particularly microbial species) often lack experimentally verifiable initial states, turning these into hidden model parameters that require careful estimation. Several strategies can mitigate this issue when experimental data are unavailable:

- Parameterization approach: treat initial conditions as additional unknown parameters in the calibration process. While mathematically rigorous, this increases computational load and parameter space dimensionality, potentially degrading identifiability conditions.
- Spin-off simulations: perform extended dynamic simulations (typically weeks/months, and more than five times the hydraulic retention time) prior to the study period. The unmeasured states which are simulated then implicitly depends on the model parameters.
- Literature values (not recommended): while theoretically possible, borrowing values from similar studies often introduces significant bias and should generally be avoided.

In any case, none of these methods can replace direct measurements.

5.5. Model validation

Model validation consists in comparing model predictions and measurements for datasets which were not involved in the parameter estimation process. It assesses the prediction capability of a model. Unfortunately, the real conditions for the validation phase are rarely clear. In particular, to evaluate the distance from an industrial case study, the time validation window (days, weeks, seasons or years) is a fundamental indicator of the model quality, together with the reactor size (from <1 L to 100 m^3), and how much it is affected by climate perturbations (light and temperature controlled, temperature controlled or free evolution of light and temperature). For such system submitted to permanent light and temperature fluctuations, a winter validation is more challenging. The number of variables correctly predicted is another strong index, which should be considered in relation to model complexity (e.g., a model considering five different microbial populations cannot be validated only against suspended solids).

5.6. Is identifiability an absolutely required condition?

Identifiability issues (practical or theoretical) mean that some of the model parameters cannot be uniquely identified, so their estimated value is related to other parameters. Although this has several inconveniences, if the associated parameters can be directly measured, this becomes eventually relatively negligible. To illustrate this property, let us come back to Eqs. 5, 6, or 8. In these equations, it is worth remarking depth L and the light extinction coefficient K appear as the product $K \bullet L$. Since these constants are accurately known, there is no problem if the model is not identifiable. In the end, what counts is an accurate determination of $K \bullet L$. In the same manner, in the previous proposed examples, there is always a way, either to identify the combination of parameters, or to separately estimate one of the parameters so that the identification problem considering the remaining parameters becomes well posed. The risk related to identifiability is to lose the meaning of the parameters, or to lose convexity of the loss function, preventing the convergence of any algorithm. These two problems can be treated once the identifiability problem has been identified and solved.

There are cases when unidentifiable models are very useful. For such models where parameters do not have any meaning, swapping several parameters can provide the same model outputs. This case is not marginal, since it appears in any standard artificial neural network, where the weights and biases between two neurons of the same layer can be swapped without any consequence (Bona-Pellissier et al., 2023). In general, the parameters of ANN are not identifiable, but ANN are very powerful. The backpropagation algorithm (Li et al., 2012) is the breakthrough that provided their efficiency. By efficiently minimizing the prediction error, this powerful strategy turns out to ensure a local convergence even with these issues. Neural networks can however be improved and made identifiable by imposing a rank in the biases, leading to enhanced performances (Yang et al., 2008).

5.7. Beyond model parametrization

While we can find many good examples of models calibrated and validated in literature, guidelines about how to use those models for decision making for technology selection, process design and optimization or evaluation of control strategies remain scarce. Upon robust calibration and validation, models can be applied to simulate process performance with different wastewater compositions, climatic conditions or operational strategies. When doing so, it is important to account with parameter uncertainty (as defined in Section 5.4.3) and propagate it towards model outputs (Belia et al., 2009), so eventual decisions with regards to design or optimization are robust against variability on microbial processes or environmental conditions. Uncertainty analyses assess parameter values randomly chosen within the uncertainty ranges previously established (either based on literature or derived from parameter estimation) as inputs to the model and record the simulation to determine the range of potential outputs. Uncertainty assessment methodologies addressing parameter uncertainty are well developed for conventional water treatment (Refsgaard et al., 2007; Sin et al., 2009) and can be applied to phototrophic processes (Wágner et al., 2016). In most cases, uncertainty is evaluated by propagating parameter uncertainty through the system of differential equations, after integrating the sensitivity equations, combined with scenario analysis, addressing uncertainties about the system description (e.g., expected loads). Uncertainties associated to model structure and input uncertainty are considered through parameter estimates, so they can be accounted for through conventional uncertainty analysis. Yet, more advanced tools to assess and reduce these uncertainty sources prior to model calibration are available from the ASM family (Belia et al., 2021). Should the uncertainty be too large, uncertainty assessment methods can be easily coupled to sensitivity analysis, thereby guiding new calibration efforts for new scenarios. Uncertainty assessment can go beyond process models, combined with life cycle assessment frameworks for sustainability evaluation (Bisinella et al., 2021), although systematic coupling of process modelling and mass flow analysis needed for LCA is still limited in wastewater applications (Fang et al., 2016; Varling et al., 2025; Penaranda et al., 2025).

6. Advancing phototrophic system modelling: current challenges and future directions

Most modelling efforts have focused on developing mathematical formulations describing our mechanistic understanding of phototrophic systems. Frequently, models available in the literature are validated with limited datasets, rarely measuring the dynamics of the microbial populations. Moreover, those datasets rarely reflect typical dynamics to which full scale open systems are subjected. Furthermore, model complexity can be considerably reduced depending on modelling objectives, and in most cases, data availability in conventional treatment trains prevents from adequately calibrating such a complex modelling framework. Thus, while models exist, we lack effective guidance on how to define a model complexity suited to our modelling needs and data availability. More recent modelling approaches, such as data-driven, hybrid or metabolic modelling, are still under development and have barely been adopted by industry. Given this, we do believe model developers should address the following aspects:

- Develop clear and good modelling practices, whereby practitioners and model users can identify which model structure and complexity should be considered given the modelling objectives and data availability. Developing a mathematical model without clearly defining its objectives can lead to inefficiencies. When the objectives are well-defined, they guide the model developer in choosing appropriate approaches and methodologies. For example, if the goal is to optimize microalgae production from wastewater, the model structure and parameters should be selected to reflect this specific aim. It is crucial for the hypotheses underlying the equations of a mathematical model to be explicit and transparent, ensuring that model developers and end-users understand the assumptions and limitations of the model. This transparency enhances trust and facilitates broader utilization of the model beyond its initial development. Those practices should include established protocols for model calibration and validation, like those developed for activated sludge modelling frameworks.
- For complex modelling tasks, such as fluid dynamics, light distribution or modelling of complex mixed communities, most facilities lack the necessary instrumentation and analytical capabilities for proper data collections. Efforts in hybrid and data driven modelling

should focus on reducing model complexity, while maintaining predictive capabilities to support decision-making (whether this is control, optimization or design). Non-mechanistic modules can be tested and trained on more detailed models before using them for full scale applications. Such an approach could be used, for example, for online prediction of microbial biomass growth rates without accounting for complex light attenuation models or modelling of individual microbial species.

- Microbial interactions in phototrophic microbiomes are complex, based on metabolites exchange and shading effects that influence light penetration. The current approach relies on combining existing activated sludge models with either microalgae or phototrophic bacteria, assuming that the main interactions are driven by nutrient competition and carbon exchange. Metabolic models relying on metagenome-assembled genomes can improve our understanding of the microbial ecology in these microbiomes, paving the way for the creation of more detailed models in the future. However, metabolic models are currently used mostly for steady-state optimization through flux balance analysis or simplified networks that can be integrated into dynamic modelling frameworks. In any case, these models are barely applied to environmental technologies and tend to focus on industrial biotechnology with pure or simple co-cultures. Even when metabolic models are extended to diverse microbiomes, they may still be too complex to describe the dynamics at full-scale. Thus, future developments should address how to transfer knowledge from detailed metabolic models to simpler mechanistic models or use them for training of data driven approaches.
- While fitting data is essential, a robust model should go beyond mere calibration. It should be capable of simulating various scenarios, such as different wastewater compositions or climatic conditions. This capability allows researchers and practitioners to explore the model's behaviour under diverse real-world conditions, thereby enhancing its practical utility. However, scenario analysis should consider parameter uncertainty, especially when extrapolating beyond the calibration range. This involves assessing how uncertainties in model parameters propagate into predictions. Understanding these uncertainties is critical for evaluating the reliability of model predictions and managing associated risks effectively. However, modelers lack standardized frameworks bridging process modelling uncertainties with those related to LCA for sustainability assessment.
- Mathematical models can play a crucial role in detecting system failures (e.g., predators, viruses, heat waves or nutrient imbalances) by simulating different scenarios. By identifying potential failure points, models enable stakeholders to develop and test mitigation strategies proactively.

In conclusion, addressing these challenges requires a collaborative and concrete effort from both academia and industry to refine modelling frameworks, improve data accessibility, measuring more quantities and develop standardized methodologies. By integrating emerging modelling approaches with robust calibration and validation protocols, we can enhance the predictive power and practical use of phototrophic system models. This will finally facilitate their adoption at large-scale, driving innovation in sustainable WWT and bioprocess optimization.

CRediT authorship contribution statement

Francesca Casagli: Project administration, Conceptualization, Supervision, Writing – original draft, Writing – review & editing. **Andrea Turolla:** Project administration, Conceptualization, Supervision, Writing – original draft, Writing – review & editing. **Damien J. Batstone:** Conceptualization, Supervision, Writing – original draft, Writing – review & editing. **Gabriel Capson-Tojo:** Conceptualization, Supervision, Writing – original draft, Writing – review & editing. **Elena Ficara:** Conceptualization, Supervision, Writing – original draft, Writing – review & editing. **Joan García:** Conceptualization, Supervision, Writing – original draft, Writing – review & editing. **Eva González-Flo:** Conceptualization, Supervision, Writing – original draft, Writing – review & editing. **Julien Laurent:** Conceptualization, Supervision, Writing – original draft, Writing – review & editing. **Tatjana Lorenz:** Conceptualization, Supervision, Writing – original draft, Writing – review & editing. **Michaël Pierrelée:** Conceptualization, Supervision, Writing – original draft, Writing – review & editing. **Benedek Gy. Plósz:** Conceptualization, Supervision, Writing – original draft, Writing – review & editing. **Gustavo Henrique Ribero Da Silva:** Conceptualization, Supervision, Writing – original draft, Writing – review & editing. **Ángel Robles:** Conceptualization, Supervision, Writing – original draft, Writing – review & editing. **Simone Rossi:** Conceptualization, Supervision, Writing – original draft, Writing – review & editing. **Estel Rueda-Hernández:** Conceptualization, Supervision, Writing – original draft, Writing – review & editing. **Lars Stegemüller:** Conceptualization, Supervision, Writing – original draft, Writing – review & editing. **Jean-Philippe Steyer:** Conceptualization, Supervision, Writing – original draft, Writing – review & editing. **Olivier Bernard:** Conceptualization, Supervision, Writing – original draft, Writing – review & editing. **Borja Valverde-Pérez:** Project administration, Conceptualization, Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Concerning the contribution of Universitat Politècnica de Catalunya · BarcelonaTech, this publication is part of the R&D&I project PID2021-126564OB-C32 (Cyan2Bio), funded by MICIU/AEI/10.13039/501100011033 and by ERDF/EU. JG would like to acknowledge the support provided by the ICREA Academia program. GR would like to acknowledge the Brasilian National Council for Scientific and Technological Development (CNPq process 308663/2021-7). EGF would like to thank the European Union-NextGenerationEU, Spanish Ministry of Science, Innovation and Universities, and Recovery, Transformation and Resilience Plan for her research grant [2021UPF-MS-12]. FC and OB are grateful to the support of the ANR Barrier project (ANR-22-CE34-0021). BVP acknowledges the financial support provided by the Danish Research Council for Independent Research under the Sapere Aude DFF Starting Grant 10.46540/2067–00029B.

Data availability

Data will be made available on request.

References

Al Ahmad, M., Raji, S., Mustafa, F., Rizvi, T.A., Al-Zuhair, S., 2020. Simultaneous and rapid quantification of microalga biomolecule content using electrochemical impedance spectroscopy. Biotechnology Progress 36 (6), e3037.

Alloul, A., Moradvandi, A., Puyol, D., Molina, R., Gardella, G., Vlaeminck, S.E., De Schutter, B., Abraham, E., Lindeboom, R.E., Weissbrodt, D.G., 2023. A novel mechanistic modelling approach for microbial selection dynamics: towards improved design and control of raceway reactors for purple bacteria. Bioresour. Technol. 390, 129844.

Amin, S.A., Parker, M.S., Armbrust, E.V., 2012. Interactions between diatoms and bacteria. Microbiol. Mol. Biol. Rev. 76 (3), 667–684.

Amini, H., Wang, L., Hashemisohi, A., Shahbazi, A., Bikdash, M., Kc, D., Yuan, W., 2018. An integrated growth kinetics and computational fluid dynamics model for the analysis of algal productivity in open raceway ponds. Comput. Electron. Agric. 145, 363–372.

Amini, A., Porciatti, E., Minotto, A., Moranchel, Á.T., Sassella, A., Canziani, R., Turolla, A., 2025. Exploring optical properties and radiation transfer in a mixed culture of purple phototrophic bacteria grown in a flat-plate photobioreactor via combined experimental and modelling approach. J. Environ. Chem. Eng. 115425.

Anderson, A.J., Dawes, E., 1990. Occurrence, metabolism, metabolic role, and industrial uses of bacterial polyhydroxyalkanoates. Microbiol. Rev. 54 (4), 450–472.
Aparicio, S., González-Camejo, J., Seco, A., Borrás, L., Robles, Á., Ferrer, J., 2023. Integrated microalgae-bacteria modelling: application to an outdoor membrane photobioreactor (MPBR). Sci. Total Environ. 884, 163669.
Arashiro, L.T., Rada-Ariza, A.M., Wang, M., van der Steen, P., Ergas, S.J., 2017. Modelling shortcut nitrogen removal from wastewater using an algal−bacterial consortium. Water Sci. Technol. 75 (4), 782–792.
Bahramian, M., Dereli, R.K., Zhao, W., Giberti, M., Casey, E., 2023. Data to intelligence: The role of data-driven models in wastewater treatment. Expert Syst. Appl. 217, 119453.
Barbosa, R.C., Soares, J., Martins, M.A., 2020. Low-cost and versatile sensor based on multi-wavelengths for real-time estimation of microalgal biomass concentration in open and closed cultivation systems. Comput. Electron. Agric. 176, 105641.
Barceló-Villalobos, M., Sánchez, J.G., Cara, I.M., Molina, J.S., Fernández, F.A., 2018. Analysis of mass transfer capacity in raceway reactors. Algal Res. 35, 91–97.
Baroukh, C., Muñoz-Tamayo, R., Bernard, O., Steyer, J.P., 2015. Mathematical modeling of unicellular microalgae and cyanobacteria metabolism for biofuel production. Curr. Opin. Biotechnol. 33, 198–205.
Baroukh, C., Muñoz-Tamayo, R., Bernard, O., Steyer, J.P., 2016. Reply to the comment on“ mathematical modeling of unicellular microalgae and cyanobacteria metabolism for biofuel production” by Baroukh et al.[Curr. Opin. Biotechnol. 2015, 33: 198-205]. Curr. Opin. Biotechnol. 38, 200–202.
Baroukh, C., Muñoz-Tamayo, R., Steyer, J.P., Bernard, O., 2014. DRUM: a new framework for metabolic modeling under non-balanced growth. Application to the carbon metabolism of unicellular microalgae. PloS one 9 (8), e104499.
Baroukh, C., Turon, V., Bernard, O., 2017. Dynamic metabolic modeling of heterotrophic and mixotrophic microalgal growth on fermentative wastes. PLoS Comput. Biol. 13 (6), e1005590.
Bastin, G., Dochain, D., 1990. On-Line Estimation and Adaptive Control of Bioreactors. Elsevier, Amsterdam (1990 (ISBN 0-444-88430-0). xiv+ 379 pp. Price US 146.25/Dfl. 285.00).
Batstone, D.J., Flores-Alsina, X., 2022. Generalised Physicochemical Model (PCM) for Wastewater Processes, Generalised Physicochemical Model (PCM) for Wastewater Processes.
Batstone, D.J., Keller, J., Angelidaki, I., Kalyuzhnyi, S.V., Pavlostathis, S.G., Rozzi, A., Sanders, W.T.M., Siegrist, H.A., Vavilin, V.A., 2002. The IWA anaerobic digestion model no 1 (ADM1). Water Sci. Technol. 45 (10), 65–73.
Béchet, Q., Shilton, A., Park, J.B., Craggs, R.J., Guieysse, B., 2011. Universal temperature model for shallow algal ponds provides improved accuracy. Environ. Sci. Technol. 45 (8), 3702–3709.
Béchet, Q., Shilton, A., Guieysse, B., 2013. Modeling the effects of light and temperature on algae growth: state of the art and critical assessment for productivity prediction during outdoor cultivation. Biotechnol. Adv. 31, 1648–1663.
Béchet, Q., Sialve, B., Steyer, J.P., Shilton, A., Guieysse, B., 2018. Comparative assessment of evaporation models in algal ponds. Algal Res. 35, 283–291.
Belia, E., Amerlinck, Y., Benedetti, L., Johnson, B., Sin, G., Vanrolleghem, P.A., Gernaey, K.V., Gillot, S., Neumann, M.B., Rieger, L., Shaw, A., Villez, K., 2009. Wastewater treatment modelling: dealing with uncertainties. Water Sci. Technol. 60 (8), 1929–1941.
Belia, E., Benedetti, L., Johnson, B., Murthy, S., Neumann, M., Vanrolleghem, P.A., Weijers, S., 2021. Uncertainty in Wastewater Treatment Design and Operation: Addressing Current Practices and Future Directions. IWA Publishing, London, UK.
Benner, P., Meier, L., Pfeffer, A., Krüger, K., Oropeza Vargas, J.E., Weuster-Botz, D., 2022. Lab-scale photobioreactor systems: principles, applications, and scalability. Bioprocess Biosyst. Eng. 45 (5), 791–813.
Bernard, O., 2011. Hurdles and challenges for modelling and control of microalgae for CO2 mitigation and biofuel production. J. Process Control 21 (10), 1378–1389.
Bernard, O., Mairet, F., Chachuat, B., 2015. Modelling of microalgae culture systems with applications to control and optimization. Microalgae Biotechnol. 59–87.
Bernard, O., Rémond, B., 2012. Validation of a simple model accounting for light and temperature effect on microalgal growth. Bioresour. Technol. 123, 520–527.
Bisht, B., Kumar, V., Gururani, P., Tomar, M.S., Nanda, M., Vlaskin, M.S., Kumar, S., Kurbatova, A., 2021. The potential of nuclear magnetic resonance (NMR) in metabolomics and lipidomics of microalgae-a review. Arch. Biochem. Biophys. 710, 108987.
Bisinella, V., Christensen, T.H., Astrup, T.F., 2021. Future scenarios and life cycle assessment: systematic review and recommendations. Int. J. Life Cycle Assess. 26, 2143–2170.
Bona-Pellissier, J., Bachoc, F., Malgouyres, F., 2023. Parameter identifiability of a deep feedforward ReLU neural network. Mach. Learn. 112 (11), 4431–4493.
Bonnefond, H., Moelants, N., Talec, A., Bernard, O., Sciandra, A., 2016. Concomitant effects of light and temperature diel variations on the growth rate and lipid production of Dunaliella salina. Algal Res. 14, 72–78.
Bougaran, G., Bernard, O., Sciandra, A., 2010. Modeling continuous cultures of microalgae colimited by nitrogen and phosphorus. J. Theor. Biol. 265 (3), 443–454.
Bradley, I.M., Pinto, A.J., Guest, J.S., 2016. Design and evaluation of illumina MiSeq-compatible, 18S rRNA gene-specific primers for improved characterization of mixed phototrophic communities. Appl. Environ. Microbiol. 82.
Bravo-Porras, G., Fernández-Güelfo, L.A., Alvarez-Gallego, C.J., Carbú, M., Sales, D., Romero-García, L.I., 2024. Influence of the total concentration and the profile of volatile fatty acids on polyhydroxyalkanoates (PHA) production by mixed microbial cultures. Biomass Convers. Biorefin. 14 (1), 239–253.
Broekhuizen, N., Park, J.B.K., McBride, G.B., Craggs, R.J., 2012. Modification, calibration and verification of the IWA River water quality model to simulate a pilot-scale high rate algal pond. Water Res. 46 (9), 2911–2926.
Cairone, S., Hasan, S.W., Choo, K.H., Lekkas, D.F., Fortunato, L., Zorpas, A.A., Korshin, G., Zarra, T., Belgiorno, V., Naddeo, V., 2024. Revolutionizing wastewater treatment toward circular economy and carbon neutrality goals: Pioneering sustainable and efficient solutions for automation and advanced process control with smart and cutting-edge technologies. J. Water Process Eng. 63, 105486.
Caparroz, M., Guzmán, J.L., Berenguel, M., Acién, F.G., 2024. A novel data-driven model for prediction and adaptive control of pH in raceway reactor for microalgae cultivation. New Biotechnol. 82, 1–13.
Capson-Tojo, G., Batstone, D.J., Grassino, M., Vlaeminck, S.E., Puyol, D., Verstraete, W., Kleerebezem, R., Oehmen, A., Ghimire, A., Pikaar, I., Lema, J.M., 2020. Purple phototrophic bacteria for resource recovery: challenges and opportunities. Biotechnol. Adv. 43, 107567.
Capson-Tojo, G., Lin, S., Batstone, D.J., Hülsen, T., 2021. Purple phototrophic bacteria are outcompeted by aerobic heterotrophs in the presence of oxygen. Water Res. 194, 116941.
Capson-Tojo, G., Batstone, D.J., Grassino, M., Hülsen, T., 2022. Light attenuation in enriched purple phototrophic bacteria cultures: implications for modelling and reactor design. Water Res. 219, 118572.
Capson-Tojo, G., Batstone, D.J., Hülsen, T., 2023. Expanding mechanistic models to represent purple phototrophic bacteria enriched cultures growing outdoors. Water Res. 229, 119401.
Carecci, D., Catenacci, A., Rossi, S., Casagli, F., Ferretti, G., Leva, A., Ficara, E., 2024. A plant-wide modelling framework to describe microalgae growth on liquid digestate in agro-zootechnical biomethane plants. Chem. Eng. J. 149981.
Casagli, F., Bernard, O., 2022. How heat transfer indirectly affects performance of algae-bacteria raceways. Microorganisms 10 (8), 1515.
Casagli, F., Zuccaro, G., Bernard, O., Steyer, J.P., Ficara, E., 2021a. ALBA: a comprehensive growth model to optimize algae-bacteria wastewater treatment in raceway ponds. Water Res. 190, 116734.
Casagli, F., Rossi, S., Steyer, J.P., Bernard, O., Ficara, E., 2021b. Balancing microalgae and nitrifiers for wastewater treatment: can inorganic carbon limitation cause an environmental threat? Environ. Sci. Technol. 55 (6), 3940–3955.
Casagli, F., Béline, F., Ficara, E., Bernard, O., 2023. Optimizing resource recovery from wastewater with algae-bacteria membrane reactors. Chem. Eng. J. 451, 138488.
Cerruti, M., Kim, J.-H., Pabst, M., Van Loosdrecht, M.C.M., Weissbrodt, D.G., 2022. Light intensity defines growth and photopigment content of a mixed culture of purple phototrophic bacteria. Front. Microbiol. 13, 1014695.
Chen, L., Bernard, O., Bastin, G., Angelov, P., 2000. Hybrid modelling of biotechnological processes using neural networks. Control. Eng. Pract. 8 (7), 821–827.
Chong, J.W.R., Khoo, K.S., Chew, K.W., Ting, H.Y., Show, P.L., 2023. Trends in digital image processing of isolated microalgae by incorporating classification algorithm. Biotechnol. Adv. 63, 108095.
Chrysochoidis, V., Andersen, M.H., Remigi, E.U., Faragó, M., Smets, B.F., Domingo-Félez, C., Valverde-Pérez, B., 2024. Critical evaluation of different mass transfer equations to model N2O emissions from water resource recovery facilities with diffuse aeration. Environ. Technol. 45 (17), 3339–3353.
Copp, J.B. (Ed.), 2002. The Cost Simulation Benchmark: Description and Simulator Manual: A Product of Cost Action 624 and Cost Action 682. EUR-OP.
Coppens, J., Grunert, O., Van Den Hende, S., Boon, N., Haesaert, G., De Gelder, L., 2016. The use of microalgae as a high-value organic slow-release fertilizer results in tomatoes with increased carotenoid and sugar levels. J. Appl. Phycol. 28, 2367–2377.
Crouchett-Catalán, F., Arango, J., Bernard, O., Martínez, C., Casagli, F., Jeison, D., 2025. M-ALBA: A modelling framework to guide the optimization of membrane-assisted algae-bacteria systems. Sci. Total Environ. 971, 179061.
de Mooij, T., de Vries, G., Latsos, C., Wijffels, R.H., Janssen, M., 2016. Impact of light color on photobioreactor productivity. Algal Res. 15, 32–42.
de Wolff, T., Lincopi, H.C., Martí, L., Sanchez-Pi, N., 2022. Mopinns: an evolutionary multi-objective approach to physics-informed neural networks. In: Proceedings of the Genetic and Evolutionary Computation Conference Companion, pp. 228–231.
Del Rio-Chanona, E.A., Cong, X., Bradford, E., Zhang, D., Jing, K., 2019. Review of advanced physical and data-driven models for dynamic bioprocess simulation: case study of algae–bacteria consortium wastewater treatment. Biotechnol. Bioeng. 116 (2), 342–353.
del Rosario Rodero, M., Magdalena, J.A., Steyer, J.P., Escudié, R., Capson-Tojo, G., 2024. Potential of enriched phototrophic purple bacteria for H2 bioconversion into single cell protein. Sci. Total Environ. 908, 168471.
Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L., 2009, June. Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. Ieee, pp. 248–255.
Dittami, S.M., et al., 2021. A community perspective on the concept of marine holobionts: current status, challenges, and future directions. PeerJ 9, e10911.
Dochain and Vanrolleghem, 2001. Dynamic Modelling and Estimation in Wastewater Treatment Processes. IWA Publishing, London, UK.
Domingo-Félez, C., Smets, B.F., 2019. Regulation of key N_2O production mechanisms during biological water treatment. Curr. Opin. Biotechnol. 57, 119–126.
Droop, M.R., 1983. 25 years of algal growth kinetics a personal view. Bot. Mar. 27, 99.
Dutt, V., Srivastava, S., 2018. Novel quantitative insights into carbon sources for synthesis of poly hydroxybutyrate in Synechocystis PCC 6803. Photosynth. Res. 136 (3), 303–314.
Endres, C.H., Roth, A., Brück, T.B., 2016. Thermal reactor model for large-scale algae cultivation in vertical flat panel photobioreactors. Environ. Sci. Technol. 50 (7), 3920–3927.

Fang, L.L., Valverde-Pérez, B., Damgaard, A., Plósz, B.Gy., Rygaard, M., 2016. Life cycle assessment as development and decision support tool for wastewater resource recovery technology. Water Res. 88, 538–549.
Fierro, I., Chachuat, B., Bernard, O., 2024. Optimal control of photobioreactor accounting for photoinhibition and photoacclimation. In: ADCHEM 2024- 12th IFAC Symposium on Advanced Control of Chemical Processes, Toronto, Canada.
Fierro, U, Lu, L.D., Bernard, O., 2025. Should hydrodynamics be taken into account when calculating the growth rate of microalgae in a photobioreactor? SIAM J. Appl. Math. 85 (4), 1906–1925.
Freitas, D., Lopes, L.G., Morgado-Dias, F., 2020. Particle swarm optimisation: a historical review up to the current developments. Entropy 22, 362.
Gábor, A., Villaverde, A.F., Banga, J.R., 2017. Parameter identifiability analysis and visualization in large-scale kinetic models of biosystems. BMC Syst. Biol. 11 (1), 54.
Gao, X., Kong, B., Dennis Vigil, R., 2017. Comprehensive computational model for combining fluid hydrodynamics, light transport and biomass growth in a Taylor vortex algal photobioreactor: Eulerian approach. Algal Res. 24, 1–8.
Garlapati, D., Charankumar, B., Ramu, K., Madeswaran, P., Ramana Murthy, M.V., 2019. A review on the applications and recent advances in environmental DNA (eDNA) metagenomics. Rev. Environ. Sci. Bio/Technol. 18 (3), 389–411.
Geider, R.J., MacIntyre, H.L., Kana, T.M., 1996. A dynamic model of photoadaptation in phytoplankton. Limnol. Oceanogr. 41 (1), 1–15.
Gernigon, V., Chekroun, M.A., Cockx, A., Guiraud, P., Morchain, J., 2019. How mixing and light heterogeneity impact the overall growth rate in Photobioreactors. Chem. Eng. Technol. 42, 1663–1669.
Gharib, A., Djema, W., Casagli, F., Bernard, O., 2024. Adaptive temperature model for microalgae cultivation systems. J. Process Control 141, 103280.
Gharib, A., Djema, W., Fernández, P.M., Chin-On, R., Janssen, M., Bernard, O., 2025. Validation of an adaptive temperature model for closed microalgae cultivation systems. Algal Res. 86, 103838.
Gincley, B., Khan, F., Alam, M.M., Hartnett, E., Kim, G.Y., Molitor, H.R., Fisher, A., Bradley, I., Guest, J., Pinto, A.J., 2025. Morphotype-resolved characterization of microalgal communities in a nutrient recovery process with ARTiMiS flow imaging microscopy. Water Res. 123801.
Gorzelnik, S.A., Zhu, X., Angelidaki, I., Koski, M., Valverde-Pérez, B., 2023. Application of Daphnia magna as biological harvesters for green microalgae grown in recycling aquaculture systems effluents. Sci. Total Environ. 873, 162247.
Gregor, J., Maršálek, B., 2005. A simple in vivo fluorescence method for the selective detection and quantification of freshwater cyanobacteria and eukaryotic algae. Acta Hydrochim. Hydrobiol. 33 (2), 142–148.
Grimaud, G.M., Mairet, F., Sciandra, A., Bernard, O., 2017. Modeling the temperature effect on the specific growth rate of phytoplankton: a review. Rev. Environ. Sci. Biotechnol. 16 (4), 625–645.
Guest, J.S., Van Loosdrecht, M.C.M., Skerlos, S.J., Love, N.G., 2013. Lumped pathway metabolic model of organic carbon accumulation and mobilization by the alga Chlamydomonas reinhardtii. Environ. Sci. Technol. 47 (7), 3258–3267.
Guo, B., Lei, C., Kobayashi, H., Ito, T., Yalikun, Y., Jiang, Y., Tanaka, Y., Ozeki, Y., Goda, K., 2017. High-throughput, label-free, single-cell, microalgal lipid screening by machine-learning-equipped optofluidic time-stretch quantitative phase microscopy. Cytometry A 91, 494–502.
Guyonvarch, E., Ramin, E., Kulahci, M., Plósz, B.G., 2015. iCFD: interpreted computational fluid dynamics - degeneration of CFD to one-dimensional advection-dispersion models using statistical experimental design - the secondary clarifier. Water Res. 83, 396–411.
Hacizade, U., 2019. Comparison of the genetic algorithm and the least squares method for determining the calibration function of a differential pressure sensor. Meas. Tech. 61, 1081–1085.
Hadiyanto, H., Elmore, S., Van Gerven, T., Stankiewicz, A., 2013. Hydrodynamic evaluations in high-rate algae pond (HRAP) design. Chem. Eng. J. 217, 231–239.
Haldane, J.B.S., 1930. Enzymes Longmans. Green and, Co, UK, p. 7.
Hauduc, H., Neumann, M.B., Muschalla, D., Gamerith, V., Gillot, S., Vanrolleghem, P.A., 2015. Efficiency criteria for environmental model quality assessment: a review and its application to wastewater treatment. Environ. Model Softw. 68, 196–204.
Havlik, I., Reardon, K.F., Ünal, M., Lindner, P., Prediger, A., Babitzky, A., Beutel, S., Scheper, T., 2013. Monitoring of microalgal cultivations with on-line, flow-through microscopy. Algal Res. 2, 253–257.
Havlik, I., Beutel, S., Scheper, T., Reardon, K.F., 2022. On-line monitoring of biological parameters in microalgal bioprocesses using optical methods. Energies 15, 875.
Hays, S.G., Yan, L.L., Silver, P.A., Ducat, D.C., 2017. Synthetic photosynthetic consortia define interactions leading to robustness and photoproduction. J. Biol. Eng. 11 (1), 4.
Heirendt, L., et al., 2019. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. Nat. Protoc. 14, 639–702.
Helton, J.C., Davis, F.J., 2003. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. Reliab. Eng. Syst. Saf. 81, 23–69.
Henze, M., Gujer, W., Mino, T., Van Loosedrecht, M., 2006. Activated Sludge Models ASM1, ASM2, ASM2d and ASM3. IWA Publishing.
Henze, M., Van Loosdrecht, M.C.M., Ekama, G.A., Brdjanovic, D., 2008. Biological Wastewater Treatment: Principles, Modelling and Design. IWA Publishing.
Hreiz, R., Sialve, B., Morchain, J.O., Escudié, R., Steyer, J.P., Guiraud, P., 2014. Experimental and numerical investigation of hydrodynamics in raceway reactors used for algaculture. Chem. Eng. J. 250, 230–239.
Huang, J., Qu, X., Wan, M., Ying, J., Li, Y., Zhu, F., Wang, J., Shen, G., Chen, J., Li, W., 2015. Investigation on the performance of raceway ponds with internal structures by the means of CFD simulations and experiments. Algal Res. 10, 64–71.
Hülsen, T., Batstone, D.J., Keller, J., 2014. Phototrophic bacteria for nutrient recovery from domestic wastewater. Water Res. 50, 18–26.
Hülsen, T., Stegman, S., Batstone, D.J., Capson-Tojo, G., 2022. Naturally illuminated photobioreactors for resource recovery from piggery and chicken-processing wastewaters utilising purple phototrophic bacteria. Water Res. 214, 118194.
Hwangbo, S., Al, R., Chen, X., Sin, G., 2021. Integrated model for understanding N2O emissions from wastewater treatment plants: a deep learning approach. Environ. Sci. Technol. 55 (3), 2143–2151.
Islam, M.F., Lye, L.M., 2009. Combined use of dimensional analysis and modern experimental design methodologies in hydrodynamics experiments. Ocean Eng. 36, 237–247.
Jassby, A.D., Platt, T., 1976. Mathematical formulation of the relationship between photosynthesis and light for phytoplankton. Limnol. Oceanogr. 21 (4), 540–547.
Jeckel, H., Drescher, K., 2021. Advances and opportunities in image analysis of bacterial cells and communities. FEMS Microbiol. Rev. 45.
Jourdan, N., Neveux, T., Potier, O., Kanniche, M., Wicks, J., Nopens, I., Rehman, U., Le Moullec, Y., 2019. Compartmental modelling in chemical engineering: a critical review. Chem. Eng. Sci. 210, 115196.
Karama, A., Bernard, O., Gouzé, J.L., 2010. Constrained hybrid neural modelling of biotechnological processes. Int. J. Chem. React. Eng. 8 (1).
Koch, M., Forchhammer, K., 2021. Polyhydroxybutyrate: A useful product of chlorotic cyanobacteria. Microb. Physiol. 31 (2), 67–77.
Kornblith, S., Shlens, J., Le, Q., 2019. Do Better ImageNet Models Transfer Better? Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2661–2671.
Laurent, J., Samstag, R., Wicks, J., Nopens, I., 2022. CFD Modelling for Wastewater Treatment Processes, IWA Publishing. ed. IWA Scientific and Technical Report, London, UK.
Le, Q.H., Verheijen, P.J.T., van Loosdrecht, M.C.M., Volcke, E.I.P., 2022. Application of data reconciliation to a dynamically operated wastewater treatment process with off-gas measurements. Environ. Sci. 8, 211.
Lee, E., Jalalizadeh, M., Zhang, Q., 2015. Growth kinetic models for microalgae cultivation: a review. Algal Res. 12, 497–512.
Levenspiel, O., 1998. Chemical Reaction Engineering, 3rd ed. Wiley.
Li, J., Cheng, J.H., Shi, J.Y., Huang, F., 2012. Brief introduction of back propagation (BP) neural network algorithm and its improvement. In: Advances in Computer Science and Information Engineering, 2. Springer Berlin Heidelberg, pp. 553–558.
Li, K., Duan, H., Liu, L., Qiu, R., van den Akker, B., Ni, B.J., Chen, T., Yin, H., Yuan, Z., Ye, L., 2022. An integrated first principal and deep learning approach for modeling nitrous oxide emissions from wastewater treatment plants. Environ. Sci. Technol. 56 (4), 2816–2826.
Liffman, K., Paterson, D.A., Liovic, P., Bandopadhayay, P., 2013. Comparing the energy efficiency of different high-rate algal raceway pond designs using computational fluid dynamics. Chem. Eng. Res. Des. 91, 221–226.
López Muñoz, I., Bernard, O., 2021. Modeling the influence of temperature, light intensity and oxygen concentration on microalgal growth rate. Processes 9 (3), 496.
Machado, D., Andrejev, S., Tramontano, M., Patil, K.R., 2018. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. Nucleic Acids Res. 46, 7542–7553.
MacIntyre, H.L., Kana, T.M., Anning, T., Geider, R.J., 2002. Photoacclimation of photosynthesis irradiance response curves and photosynthetic pigments in microalgae and cyanobacteria. J. Phycol. 38, 17–38.
Mahanty, B., 2023. Hybrid modeling in bioprocess dynamics: structural variabilities, implementation strategies, and practical challenges. Biotechnol. Bioeng. 120 (8), 2072–2091.
Mairet, F., Bernard, O., 2019. Twelve quick tips for designing sound dynamical models for bioprocesses. PLoS Comput. Biol. 15 (8), e1007222.
Marques, R., Rodríguez-Caballero, A., Oehmen, A., Pijuan, M., 2016. Assessment of online monitoring strategies for measuring N2O emissions from full-scale wastewater treatment systems. Water Res. 99, 171–179.
Martínez, C., Mairet, F., Bernard, O., 2018. Theory of turbid microalgae cultures. J. Theor. Biol. 456, 190–200.
Mašić, A., Santos, A.T.L., Etter, B., Udert, K.M., Villez, K., 2015. Estimation of nitrite in source-separated nitrified urine with UV spectrophotometry. Water Res. 85, 244–254.
Maurya, R.K., Zhu, X., Valverde-Pérez, B., Kiran, B.R., General, T., Sharma, S., Sharma, A.K., Thomsen, M., Mohan, S.V., Mohanty, K., Angelidaki, I., 2022. Advances in algal research for valorization and remediation of industrial wastewater. Bioresour. Technol. 343, 126128.
Mayer, P., Cuhel, R., Nyholm, N., 1997. A simple in vitro fluorescence method for biomass measurements in algal growth inhibition tests. Water Res. 31 (10), 2525–2531.
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.
Meenatchisundaram, K., Gowd, S.C., Lee, J., Barathi, S., Rajendran, K., 2024. Data-driven model development for prediction and optimization of biomass yield of microalgae-based wastewater treatment. Sustain. Energy Technol. Assess 63, 103670.
Mellbye, B.L., Giguere, A., Chaplen, F., Bottomley, P.J., Sayavedra-Soto, L.A., 2016. Steady-state growth under inorganic carbon limitation conditions increases energy consumption for maintenance and enhances nitrous oxide production in Nitrosomonas europaea. Appl. Environ. Microbiol. 82 (11), 3310–3318.
Mendoza, J.L., Granados, M.R., De Godos, I., Acién, F.G., Molina, E., Heaven, S., Banks, C.J., 2013. Oxygen transfer and evolution in microalgal culture in open raceways. Bioresour. Technol. 137, 188–195.
Millero, F.J., Schreiber, D.R., 1982. Use of the ion pairing model to estimate activity coefficients of the ionic components of natural waters. Am. J. Sci. 282 (9), 1508–1540.

Mohamadnia, S., Thygesen, A., Ghofrani-Isfahani, P., Monachese, A.P., Valverde-Pérez, B., Angelidaki, I., 2023. Valorization of potato starch wastewater using anaerobic acidification coupled with Chlorella sorokiniana cultivation. J. Appl. Phycol. 35 (6), 2645–2658.
Monje, V., Junicke, H., Batstone, D.J., Kjellberg, K., Gernaey, K.V., Flores-Alsina, X., 2022. Prediction of mass and volumetric flows in a full-scale industrial waste treatment plant. Chem. Eng. J. 445, 136774.
Monod, J., 2012. The Growth of Bacterial Cultures. Selected Papers in Molecular Biology by Jacques Monod, 139, p. 606.
Morales, M., Aflalo, C., Bernard, O., 2021. Microalgal lipids: A review of lipids potential and quantification for 95 phytoplankton species. Biomass Bioenergy 150, 106108.
Morgado, D., Fanesi, A., Martin, T., Tebbani, S., Bernard, O., Lopes, F., 2024. Non-destructive monitoring of microalgae biofilms. Bioresour. Technol. 398, 130520.
Muñoz, R., Guieysse, B., 2006. Algal–bacterial processes for the treatment of hazardous contaminants: a review. Water Res. 40 (15), 2799–2815.
Mutale-Joan, C., Sbabou, L., Hicham, E.A., 2023. Microalgae and cyanobacteria: how exploiting these microbial resources can address the underlying challenges related to food sources and sustainable agriculture: a review. J. Plant Growth Regul. 42 (1), 1–20.
Natrah, F.M., Bossier, P., Sorgeloos, P., Yusoff, F.M., Defoirdt, T., 2014. Significance of microalgal–bacterial interactions for aquaculture. Rev. Aquacult. 6 (1), 48–61.
Nguyen, B.T., Rittman, B.E., 2018. Low-cost optical sensor to automatically monitor and control biomass concentration in microalgal cultivation. Algal Res. 32, 101–106.
Nikolaou, A., Hartmann, P., Sciandra, A., Chachuat, B., Bernard, O., 2016. Dynamic coupling of photoacclimation and photoinhibition in a model of microalgae growth. J. Theor. Biol. 390, 61–72.
Nordio, R., Rodríguez-Miranda, E., Casagli, F., Sánchez-Zurano, A., Guzmán, J.L., Acién, G., 2024. ABACO-2: a comprehensive model for microalgae-bacteria consortia validated outdoor at pilot-scale. Water Res. 248, 120837.
Nordio, R., Casagli, F., Rodriguez-Miranda, E., Guzman, J.L., Bernard, O., Acien, G., 2025. Benchmarking of ALBA and ABACO-2 models for algae-bacteria wastewater treatment. Algal Res. 89, 104049.
Orth, J.D., Thiele, I., Palsson, B.O., 2010. What is flux balance analysis? Nat. Biotechnol. 28, 245–248.
Ota, S., Kawano, S., 2019. Three-dimensional ultrastructure and hyperspectral imaging of metabolite accumulation and dynamics in Haematococcus and Chlorella. Microscopy 68, 57–68.
Otálora, P., Guzmán, J.L., Acién, F.G., Berenguel, M., Reul, A., 2021. Microalgae classification based on machine learning techniques. Algal Res. 55.
Otálora, P., Guzmán, J.L., Berenguel, M., Acién, F.G., 2023. Data-driven pH model in raceway reactors for freshwater and wastewater cultures. Mathematics 11 (7), 1614.
Oviedo, J.A., Muñoz, R., Donoso-Bravo, A., Bernard, O., Casagli, F., Jeison, D., 2022. A half-century of research on microalgae-bacteria for wastewater treatment. Algal Res. 67, 102828.
Palacios, O.A., López, B.R., de Bashan, L.E., 2022. Microalga growth-promoting Bacteria (MGPB): a formal term proposed for beneficial bacteria involved in microalgal–bacterial interactions. Algal Res. 61, 102585.
Parashar, M., Kurc, T., Klie, H., Wheeler, M.F., Saltz, J.H., Jammoul, M., Dong, R., 2023. Dynamic data- driven application systems for reservoir simulation-based optimization: lessons learned and future trends. In: Handbook of Dynamic Data Driven Applications Systems, 2. Springer International Publishing, Cham, pp. 287–330.
Penaranda, D., Casagli, F., Morales, M., Beline, F., Bernard, O., 2025. Ex–ante LCA for circular resource management of liquid digestate, by predictive modeling of algae–bacterial processes. J. Ind. Ecol. 1–15.
Peng, L., Ni, B.J., Ye, L., Yuan, Z., 2015. N2O production by ammonia oxidizing bacteria in an enriched nitrifying sludge linearly depends on inorganic carbon concentration. Water Res. 74, 58–66.
Perez-Garcia, O., Escalante, F.M.E., de Bashan, L.E., Bashan, Y., 2011. Heterotrophic cultures of microalgae: metabolism and potential products. Water Res. 45 (1), 11–36.
Pessi, B.A., Pruvost, E., Talec, A., Sciandra, A., Bernard, O., 2022. Does temperature shift justify microalgae production under greenhouse? Algal Res. 61, 102579.
Pessi, B.A., Baroukh, C., Bacquet, A., Bernard, O., 2023. A universal dynamical metabolic model representing mixotrophic growth of Chlorella sp. on wastes. Water Res. 229, 119388.
Pisa, I., Santin, I., Morell, A., Vicario, J.L., Vilanova, R., 2019. LSTM-based wastewater treatment plants operation strategies for effluent quality improvement. IEEE Access 7, 159773–159786.
Podevin, M., Fotidis, I.A., Angelidaki, I., 2017. Microalgal process-monitoring based on high-selectivity spectroscopy tools: status and future perspectives. Crit. Rev. Biotechnol. 38 (5), 704–718.
Pozzobon, V., Perré, P., 2020. Multiscale numerical workflow describing microalgae motion and light pattern incidence towards population growth in a photobioreactor. J. Theor. Biol. 498, 110293.
Puyol, D., Barry, E.M., Hülsen, T., Batstone, D.J., 2017. A mechanistic model for anaerobic phototrophs in domestic wastewater applications: photo-anaerobic model (PAnM). Water Res. 116, 241–253.
Qiu, Y., Griffin, C.T., Ekström, S., Valverde-Pérez, B., Smets, B.F., Climent, J., Domingo-Félez, C., Martínez Cuenca, R., Plósz, B.Gy., 2024. Numerical modelling of surface aeration and N_2O emission in biological water resource recovery. Water Res. 255, 121398.
Qiu, Y., Bakos, V., Stewart-Campbell, N., Plósz, B.G., 2024b. Dynamic α-factor models for fine-bubble- and surface aeration – impact of settleability and implications on N2O emission. Chem. Eng. J. 488, 150650.
Ramanan, R., Kim, B.H., Cho, D.H., Oh, H.M., Kim, H.S., 2016. Algae–bacteria interactions: evolution, ecology and emerging applications. Biotechnol. Adv. 34 (1), 14–29.
Ranganathan, P., Pandey, A.K., Sirohi, R., Tuan Hoang, A., Kim, S.-H., 2022. Recent advances in computational fluid dynamics (CFD) modelling of photobioreactors: design and applications. Bioresour. Technol. 350, 126920.
Ras, M., Steyer, J.P., Bernard, O., 2013. Temperature effect on microalgae: a crucial factor for outdoor production. Rev. Environ. Sci. Biotechnol. 12 (2), 153–164.
Rasouli, Z., Valverde-Pérez, B., D'Este, M., De Francisci, D., Angelidaki, I., 2018. Nutrient recovery from industrial wastewater as single cell protein by a co-culture of green microalgae and methanotrophs. Biochem. Eng. J. 134, 129–135.
Raue, A., Karlsson, J., Saccomani, M.P., Jirstrand, M., Timmer, J., 2014. Comparison of approaches for parameter identifiability analysis of biological systems. Bioinformatics 30 (10), 1440–1448.
Refsgaard, J.C., van der Sluijs, J.P., Højberg, A.L., Vanrolleghem, P.A., 2007. Uncertainty in the environmental modelling process – a framework and guidance. Environ. Model Softw. 22 (11), 1543–1556.
Reichert, P., Borchardt, D., Henze, M., Rauch, W., Shanahan, P., Somlyódy, L., Vanrolleghem, P., 2001. River water quality model no. 1 (RWQM1): II. Biochemical process equations. Water Sci. Technol. 43 (5), 11–30.
Richmond, A., 2004. Handbook of microalgal culture: biotechnology and applied phycology. Orton.Catie.Ac.Cr 472.
Rieger, L., Gillot, S., Langergraber, G., Ohtsuki, T., Shaw, A., Takacs, I., Winkler, S., 2013. Guidelines for Using Activated Sludge Models. IWA Publishing.
Robles, Á., Capson-Tojo, G., Galès, A., Ruano, M.V., Sialve, B., Ferrer, J., Steyer, J.P., 2020. Microalgae-bacteria consortia in high-rate ponds for treating urban wastewater: elucidating the key state indicators under dynamic conditions. J. Environ. Manag. 261, 110244.
Rodríguez-Rángel, H., Arias, D.M., Morales-Rosales, L.A., Gonzalez-Huitron, V., Valenzuela Partida, M., García, J., 2022. Machine learning methods modeling carbohydrate-enriched cyanobacteria biomass production in wastewater treatment systems. Energies 15 (7), 2500.
Romero-García, J.M., Acién-Fernández, F.G., Fernández-Sevilla, J.M., 2012. Development of a process for the production of l-amino-acids concentrates from microalgae by enzymatic hydrolysis. Bioresour. Technol. 112, 164–170.
Ross, O.N., Geider, R.J., 2009. New cell-based model of photosynthesis and photo-acclimation: accumulation and mobilisation of energy reserves in phytoplankton. Mar. Ecol. Prog. Ser. 383, 53–71.
Rossi, S., Díez-Montero, R., Rueda, E., Castillo Cascino, F., Parati, K., García, J., Ficara, E., 2020. Free ammonia inhibition in microalgae and cyanobacteria grown in wastewaters: photo-respirometric evaluation and modelling. Bioresour. Technol. 305 (February), 123046.
Rossi, S., Capson-Tojo, G., Sánchez-Zurano, A., Carecci, D., Batstone, D.J., Acien-Fernández, G., Ficara, E., 2025. Recent advances and challenges in mechanistic modelling of photosynthetic processes for wastewater treatment. Water Res. 278, 123216.
Rosso, L., Lobry, J.R., Bajard, S., Flandrois, J.P., 1995. Convenient model to describe the combined effects of temperature and pH on microbial growth. Appl. Environ. Microbiol. 61, 610–616.
Rueda, E., García, J., 2021. Optimization of the phototrophic Cyanobacteria polyhydroxybutyrate (PHB) production by kinetic model simulation. Sci. Total Environ. 800.
Santos-Ballardo, D.U., Rossi, S., Hernández, V., Gómez, R.V., del Carmen Rendón-Unceta, M., Caro-Corrales, J., Valdez-Ortiz, A., 2015. A simple spectrophotometric method for biomass measurement of important microalgae species in aquaculture. Aquaculture 448, 87–92.
Schneider, M.Y., Quaghebeur, W., Borzooei, S., Froemelt, A., Li, F., Saagi, R., Wade, M.J., Zhu, J.J., Torfs, E., 2022. Hybrid modelling of water resource recovery facilities: status and opportunities. Water Sci. Technol. 85 (9), 2503–2524.
Serra-Maia, R., Bernard, O., Gonçalves, A., Bensalem, S., Lopes, F., 2016. Influence of temperature on Chlorella vulgaris growth and mortality rates in a photobioreactor. Algal Res. 18, 352–359.
Shoener, B., Schramm, S., Béline, F., Bernard, O., Martínez von Dossow, C., Plósz, B.Gy., Snowling, S., Steyer, J.P., Valverde-Pérez, B., Wágner, D.S., Guest, J., 2019. Microalgae and cyanobacteria modeling in water resource recovery facilities: a critical review. Water Res. X 2, 100024.
Silkina, A., Flynn, K., Llewellyn, C., Bayliss, C., 2015. Standard Operating Procedures for Analytical Methods and Data Collection in Support of Pilot-Scale Cultivation of Microalgae. Public Output report WP1A3. 01 of the EnAlgae Project.
Sin, G., Gernaey, K.V., Neumann, M.B., van Loosdrecht, M.C.M., Gujer, W., 2009. Uncertainty analysis in WWTP model applications: a critical discussion using an example from design. Water Res. 43 (11), 2894–2906.
Singh, D., Nedbal, L., Ebenhöh, O., 2018. Modelling phosphorus uptake in microalgae. Biochem. Soc. Trans. 46 (2), 483–490.
Solimeno, A., Gómez-Serrano, C., Acién, F.G., 2019. BIO_ALGAE 2: improved model of microalgae and bacteria consortia for wastewater treatment. Environ. Sci. Pollut. Res. 26 (25), 25855–25868.
Solon, K., Flores-Alsina, X., Mbamba, C.K., Volcke, E.I., Tait, S., Batstone, D., Jeppsson, U., 2015. Effects of ionic strength and ion pairing on (plant-wide) modelling of anaerobic digestion. Water Res. 70, 235–245.
Stegemüller, L., Caccavale, F., Valverde-Pérez, B., Angelidaki, I., 2025. Online monitoring of *Haematococcus lacustris* cell cycle using machine and deep learning techniques. Bioresour. Technol. 418, 131976.
Steuer, R., Knoop, H., MacHné, R., 2012. Modelling cyanobacteria: from metabolism to integrative models of phototrophic growth. J. Exp. Bot. 63 (6), 2259–2274.

Syed, T., Krujatz, F., Ihadjadene, Y., Mühlstädt, G., Hamedi, H., Mädler, J., Urbas, L., 2024. A review on machine learning approaches for microalgae cultivation systems. Comput. Biol. Med. 172, 108248.
Takahashi, T., 2018. Applicability of automated cell counter with a chlorophyll detector in routine management of microalgae. Sci. Rep. 8 (1), 4967.
Tchobanoglous, G., Stensel, H., Tsuchihashi, R., Burton, F., 2014. Wastewater Engineering: Treatment and Resource Recovery, Fifth edition. McGraw-Hill Education.
Torfs, E., Nicolaï, N., Daneshgar, S., Copp, J.B., Haimi, H., Ikumi, D., Johnson, B., Plósz, B.Gy., Snowling, S., Townley, Ll, Valverde-Pérez, B., Vanrolleghem, P.A., Vezzaro, L., Nopens, I., 2022. The transition of WRRF models to digital twin applications. Water Sci. Technol. 85 (10), 2840–2853.
Tribe, L.A., Briens, C.L., Margaritis, A., 1995. Determination of the volumetric mass transfer coefficient (k,a) using the dynamic "gas out-gas in" method: analysis of errors caused by dissolved oxygen probes. Biotechnol. Bioeng. 46, 388–392.
Van Daele, T., Van Hoey, S., Gernaey, K.V., Krühne, U., Nopens, I., 2015. A numerical procedure for model identifiability analysis applied to enzyme kinetics. Comput. Aided Process Eng. 37, 575–580.
van Loosdrecht, M.C.M., Nielsen, P.H., Lopez-Vazquez, C.M., Brdjanovic, D., 2016. Experimental Methods in Wastewater Treatment. IWA Publishing.
Varling, A.S., Chrysochoidis, V., Bisinella, V., Valverde-Pérez, B., Christensen, T.H., 2025. Climate change impact of biological treatment of liquid digestate from anaerobic digestion of food waste. Waste Manag. 193 (1), 339–349.
Vulsteke, E., Van Den Hende, S., Bourez, L., Capoen, H., Rousseau, D.P.L., Albrecht, J., 2017. Economic feasibility of microalgal bacterial floc production for wastewater treatment and biomass valorization: a detailed up-to-date analysis of up-scaled pilot results. Bioresour. Technol. 224, 118–129.
Wagner, H., Liu, Z., Langner, U., Stehfest, K., Wilhelm, C., 2010. The use of FTIR spectroscopy to assess quantitative changes in the biochemical composition of microalgae. J. Biophotonics 3 (8-9), 557–566.
Wágner, D.S., Valverde-Pérez, B., Sæbø, M., de la Sotilla, M.B., Van Wagenen, J., Smets, B.F., Plosz, B.G., 2016. Towards a consensus-based biokinetic model for green microalgae–the ASM-A. Water Res. 103, 485–499.
Wágner, D.S., Radovici, M., Smets, B.F., Angelidaki, I., Valverde-Pérez, B., Plósz, B.Gy., 2016b. Harvesting microalgae using activated sludge can decrease polymer dosing and enhance methane production via co-digestion in a bacterial-microalgal process. Algal Res. 20, 197–204.
Wágner, D.S., Valverde-Pérez, B., Plósz, B.G., 2018. Light attenuation in photobioreactors and algal pigmentation under different growth conditions - model identification and complexity assessment. Algal Res. 35, 488–499.
Wágner, D.S., Cazzaniga, C., Steidl, M., Dechesne, A., Valverde-Pérez, B., Plósz, B.Gy., 2021. Optimal influent N-to-P ratio for stable microalgal cultivation in water treatment and nutrient recovery. Chemosphere 262, 127939.
Walter, E., Pronzato, L., 1997. Identification of parametric models: from experimental data. Springer. (Communications and Control Engineering, London.
Wenk, P., Gotovos, A., Bauer, S., Gorbach, N.S., Krause, A., Buhmann, J.M., 2019. April. Fast Gaussian process based gradient matching for parameter identification in systems of nonlinear ODEs. In: The 22nd International Conference on Artificial Intelligence and Statistics. PMLR, pp. 1351–1360.
Wicklein, E., Batstone, D.J., Ducoste, J., Laurent, J., Griborio, A., Wicks, J., Saunders, S., Samstag, R., Potier, O., Nopens, I., 2015. Good modelling practice in applying computational fluid dynamics for WWTP modelling. Water Sci. Technol. 73, 969–982.
Wieland, F.G., Hauber, A.L., Rosenblatt, M., Tönsing, C., Timmer, J., 2021. On structural and practical identifiability. Curr. Opin. Syst. Biol. 25, 60–69.
Williams, P.J.L.B., Laurens, L.M., 2010. Microalgae as biodiesel & biomass feedstocks: Review & analysis of the biochemistry, energetics & economics. Energy Environ. Sci. 3 (5), 554–590.
Wolf, G., Picioreanu, C., van Loosdrecht, M.C.M., 2007. Kinetic modeling of phototrophic biofilms: the PHOBIA model. Biotechnol. Bioeng. 97 (5), 1064–1079.
Wolf-Gladrow, D.A., Zeebe, R.E., Klaas, C., Körtzinger, A., Dickson, A.G., 2007. Total alkalinity: the explicit conservative expression and its application to biogeochemical processes. Mar. Chem. 106 (1–2), 287–300.
Wongburi, P., Park, J.K., 2023. Prediction of wastewater treatment plant effluent water quality using recurrent neural network (RNN) models. Water 15 (19), 3325.
Wu, G.F., Wu, Q.Y., Shen, Z.Y., 2001. Accumulation of poly-β-hydroxybutyrate in cyanobacterium Synechocystis sp. PCC6803. Bioresour. Technol. 76 (2), 85–90.
Yahia, B., Malphettes, L., Heinzle, E., 2015. Macroscopic modeling of mammalian cell growth and metabolism. Appl. Microbiol. Biotechnol. 99 (17), 7009–7024.
Yang, Z., Cheng, J., Ye, Q., Liu, J., Zhou, J., Cen, K., 2016. Decrease in light/dark cycle of microalgal cells with computational fluid dynamics simulation to improve microalgal growth in a raceway pond. Bioresour. Technol. 220, 352–359.
Yang, S.H., Hu, B.G., Cournède, P.H., 2008. Structural identifiability of generalized constraint neural network models for nonlinear regression. Neurocomputing 72 (1–3), 392–400.
Zhang, D., Del Rio-Chanona, E.A., Petsagkourakis, P., Wagner, J., 2019. Hybrid physics-based and data-driven modeling for bioprocess online simulation and optimization. Biotechnol. Bioeng. 116 (11) (2919–293).
Zhang, Y., Wang, J.H., Zhang, J.T., Chi, Z.Y., Kong, F.T., Zhang, Q., 2023. The long overlooked microalgal nitrous oxide emission: characteristics, mechanisms, and influencing factors in microalgae-based wastewater treatment scenarios. Sci. Total Environ. 856, 159153.