Chemical Engineering Science 316 (2025) 121945

Contents lists available at ScienceDirect

Chemical Engineering Science

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

Hybrid modeling of photocatalytic contaminant degradation using nanomaterials synthesized with microalgal extracts

Federico Atzori [a], Bartolomeo Cosenza [b], Federico Zedda [a], Agnieszka Sidorowicz [c], Giacomo Fais [a,c], Giacomo Cao [a,c,d], Alessandro Concas [a,c,*]

[a] Department of Mechanical, Chemical and Materials Engineering, University of Cagliari, Via Marengo 2, 09123 Cagliari, Italy
[b] Department of Civil and Industrial Engineering, University of Pisa, Largo Lucio Lazzarino, 56122 Pisa, PI, Italy
[c] Interdepartmental Centre of Environmental Science and Engineering (CINSA), University of Cagliari, Via San Giorgio 12, 09124 Cagliari, Italy
[d] Center for Advanced Studies, Research and Development in Sardinia (CRS4), Loc. Piscina Manna, Building 1, 09050 Pula, CA, Italy

ARTICLE INFO

Keywords:
Photocatalytic degradation
Silver nanoparticles
Microalgal synthesis
Hybrid modeling
Artificial neural networks
Synthetic data generation

ABSTRACT

Synthetic dyes released through industrial effluents pose significant environmental risks due to their persistence and toxicity. Photocatalytic degradation through metal nanoparticles offers a promising, eco-friendly remediation approach. This study presents a hybrid modeling framework for simulating the visible-light-driven degradation of Brilliant Blue R using silver nanoparticles (AgNPs) synthesized through extracts of *Haematococcus pluvialis*, *Spirulina platensis*, and *Chlorella vulgaris*. The biogenic AgNPs exhibited favorable physicochemical properties, including crystalline sizes of 13–16 nm and band gap energies between 2.17 and 2.33 eV.

A simplified deterministic model was first developed, accounting for adsorption–desorption equilibrium and degradation kinetics, which enables analytical estimation of key kinetic parameters. These parameters were used to train artificial neural networks (ANNs) that map experimental conditions such as light intensity, dye concentration, nanoparticle dosage, and pH to degradation kinetics.

To overcome the limited size of datasets obtained through experiments, a novel data augmentation strategy was implemented using Gaussian noise derived from measurement uncertainty and confidence intervals of the deterministic model's parameters. This strategy enabled the significant augmentation of data enhancing the ANN performance. Indeed, the global mean squared error dropped from 5.6×10^{-4} to 1.3×10^{-5} for AgNPs from *H. pluvialis*, from 1.6×10^{-2} to 3.3×10^{-6} for *C. vulgaris*, and from 2.4×10^{-3} to 4.2×10^{-4} for *S. platensis* when using both input and output augmentation.

The proposed hybrid framework couples mechanistic interpretability with data-driven prediction providing a reliable tool for optimizing photocatalytic degradation processes via sustainable nanomaterials of microalgal origin.

1. Introduction

Several manufacturing processes, including textile dyeing, paper printing, and pharmaceutical production, utilize synthetic dyes that are released into industrial effluents. The inadequate treatment of these effluents results in wastewater contamination, as synthetic dyes can hinder microbial photosynthesis by reducing oxygen availability and limiting sunlight penetration. Some dyes also exhibit carcinogenic and mutagenic properties, posing risks to living organisms (Peyghami et al., 2023). Their complex aromatic structures confer high chemical stability, rendering them poorly biodegradable. So today there is a call for economically sustainable processes to degrade dyes in contaminated water and wastewater.

Photocatalytic degradation has emerged as a promising technique for reducing organic compound concentrations in aqueous solutions (Dutta et al., 2010). This advanced oxidation process (AOP) utilizes semiconductor properties to generate electron-hole pairs under surface irradiation, leading to the formation of strong oxidizing agents capable of degrading adsorbed dye molecules. Metallic nanoparticles, with their high surface area-to-volume ratio, significantly enhance pollutant adsorption and exhibit unique physicochemical properties, making them efficient catalysts for dye photodegradation. Metal oxides such as ZnO

* Corresponding author at: Via Marengo 2, 09123 Cagliari, Italy.
E-mail address: alessandro.concas@unica.it (A. Concas).

https://doi.org/10.1016/j.ces.2025.121945
Received 15 April 2025; Received in revised form 26 May 2025; Accepted 26 May 2025
Available online 28 May 2025

and TiO_2 are widely studied for this purpose due to their low toxicity, cost-effectiveness, and high stability. However, their relatively large band gap limits photocatalytic activation to UV radiation, restricting the use of the more sustainable solar light (Pal and Kryschi, 2016; Wu et al., 2010). To overcome this limitation, researchers have explored methods to extend the absorption spectrum, including metal deposition (Ag, Au, Fe) on metal oxide surfaces (de Souza and Corio, 2013; Pal and Kryschi, 2016; Pulido Melián et al., 2012) or the use of noble metal nanoparticles, such as silver (AgNPs), which demonstrate high efficiency, photocatalytic stability, and a lower band gap, making them suitable for environmental pollutant degradation (Syahirah Kamarudin et al., 2018).

Among the various methods for AgNP production, chemical reduction of Ag salt remains the most widely used due to its cost-effectiveness and rapid synthesis compared to physical techniques (Vidyasagar et al., 2023). However, its main limitation stems from the environmental hazards associated with highly reactive reductants and toxic stabilizers used in the process (Joseph and Mathew, 2015, 2014). Consequently, research efforts focus on developing greener synthesis strategies employing non-toxic reductants and sustainable stabilizing agents. In this context, biological approaches utilizing metabolites as both reducing and stabilizing agents have gained attention for significantly lowering the environmental footprint of AgNP production (Sidorowicz et al., 2023). Notably, microalgae, characterized by rapid growth and metabolite abundance, represent a promising feedstock for biogenic AgNP synthesis (Rajkumar et al., 2021).

In our recent studies, methanolic extracts of *Spirulina platensis* (Sidorowicz et al., 2024a), *Haematococcus pluvialis* (Sidorowicz et al., 2025), and *Chlorella vulgaris* (Sidorowicz et al., 2024b) were utilized for AgNP synthesis, which were subsequently evaluated as catalysts for the photodegradation of Brilliant Blue R (BBR) dye under visible light. The effects of light intensity (I), initial dye concentration ($[Dye]_o$), nanoparticle concentration ([NP]), and solution pH (pH) were systematically investigated. While these studies confirmed high BBR removal efficiencies under specific conditions, an optimal combination of the operating variables I, $[Dye]_o$, [NP] and pH could not be identified despite their systematic variation in the experimental trials. These limitations stem from logistical and economic constraints in the experiments, highlighting the crucial role of theoretical modeling in optimizing operating conditions, process control, and scalability of the proposed technique for BBR degradation. In this regard, pseudo-first-order models are commonly employed to describe the photodegradation process due to its mathematical simplicity, making it a practical choice for process modeling (Frontistis et al., 2012; Irineu et al., 2022; Jiang et al., 2020). However, their inability to account for multiple kinetic phenomena limits its capacity to accurately capture pollutant degradation dynamics (Rasoulifard et al., 2016). Photodegradation involves various processes, including dye and water molecule adsorption on the catalyst surface, electron-hole pair formation, and degradation reactions (Daneshvar et al., 2004; Turchi and Ollis, 1990). Pseudo-first-order models are only appropriate when a single phenomenon prevails in controlling the reaction. Experimental evidence, however, suggests that photocatalytic organic degradation rarely follows a single kinetic pathway (Dutta et al., 2010; Frontistis et al., 2012; Parmar and Srivastava, 2022; Rasoulifard et al., 2016).

This study presents a novel deterministic model for the quantitative interpretation of experimental data on BBR photodegradation using AgNPs synthesized from microalgal extracts. While still simplified, the model incorporates adsorption/desorption equilibrium, allowing the reaction term to be expressed as a function of the adsorbed dye concentration ([Dye]) rather than its bulk concentration. Despite the increased mathematical complexity, it preserves the key advantage of pseudo-first-order kinetics, i.e. an analytical solution, that enables efficient identification of kinetic parameters.

An artificial neural network (ANN) was then developed as a fast surrogate simulation tool based on the previously described deterministic model, but offering greater portability and ease of use. The ANN employs I, $[Dye]_o$, [NP] and pH as input variables, while the output consists of photodegradation and adsorption/desorption rate constants (**k**), obtained through the deterministic model. This ensures that the ANN output remains physically meaningful, as the predicted **k** values are directly linked to the main physical phenomena underlying the experimental evidence.

Given the limited availability of experimental data for the effective training and validation of ANN, a data augmentation methodology was implemented to expand the dataset by generating synthetic data that closely mimics real ones (Cosenza et al., 2024). The approach here proposed for data augmentation is based on introducing noise into both inputs and outputs, preserving the physical consistency of the system by considering the intrinsic uncertainty of experimental measurements and the confidence intervals of **k** values derived from model fitting. Beyond mitigating overfitting (Arslan et al., 2019; Rasmussen, 2004), this methodology further enhances the physical relevance of the ANN-based simulation tool. Consequently, the physics-aware ANN here developed might be used to predict the efficiency of the BBR photodegradation process using microalgae-based AgNPs, even under operating conditions (I, $[Dye]_o$, [NP], and pH) beyond the experimentally investigated ranges.

Table 1
Conditions tested during BBR photodegradation with AgNPs obtained from microalgal strains.

Strain	ID*	[NP] [mg L^{-1}]	I [μE s^{-1} m^{-2}]	pH [/]	$[Dye]^0$ [mg L^{-1}]	y^0 [/]
H. pluvialis	HP_1	250	150	7	7.7	0.23
	HP_2	250	150	7	18.0	0.54
	HP_3	250	150	7	29.7	0.89
	HP_4	250	0	7	17.3	0.52
	HP_5	250	75	7	17.5	0.525
	HP_6	250	300	7	20.7	0.62
	HP_7	500	150	7	18.3	0.55
	HP_8	125	150	7	16.3	0.49
	HP_9	250	150	11	20.0	0.60
	HP_10	250	150	3	18.0	0.54
S. platensis	SP_1	1000	150	7	4.0	0.12
	SP_2	1000	150	7	18.0	0.54
	SP_3	1000	150	7	26.3	0.79
	SP_4	1000	0	7	20.0	0.60
	SP_5	1000	75	7	18.7	0.56
	SP_6	1000	300	7	18.3	0.55
	SP_7	2000	150	7	16.0	0.48
	SP_8	500	150	7	18.0	0.54
	SP_9	1000	150	11	17.3	0.52
	SP_10	1000	150	3	18.3	0.55
C. vulgaris	CV_1	125	150	7	4.0	0.12
	CV_2	125	150	7	19.0	0.57
	CV_3	125	150	7	33.3	1.00
	CV_4	125	0	7	21.0	0.63
	CV_5	125	75	7	20.0	0.60
	CV_6	125	300	7	17.3	0.52
	CV_7	250	150	7	16.7	0.50
	CV_8	62.5	150	7	20.0	0.60
	CV_9	125	150	11	18.0	0.54
	CV_10	125	150	3	18.7	0.56

* ID of the set of operating conditions adopted during each photodegradation experiment.

2. Materials and methods

For the sake of clarity, a summary of the experimental procedures is provided in what follows. For a comprehensive account of the materials and methods employed in the experimental trials, readers are referred to the detailed descriptions in Sidorowicz et al. (2025, 2024a,b).

2.1. Synthesis of nanoparticles

AgNPs were synthesized using a biological synthesis approach, employing microalgal extracts as a source of reducing and stabilizing agents. The methodology following the procedure was previously described in Sidorowicz et al. (2025, 2024a,b). Briefly, aqueous microalgal extracts were mixed with a precursor salt solution ($AgNO_3$, Carlo Erba®) under constant stirring at 85 °C for 90 min. The synthesized nanoparticles were collected, washed, and dried at 80 °C.

2.2. Photocatalytic activity evaluation

The photocatalytic degradation of BBR dye was studied under different operating conditions, including variations in light intensity I, catalyst dosage [NP], initial dye concentration $[Dye]_0$, and pH. First, a calibration curve was established by evaluating the dye's absorbance and concentration correlation. Before each photocatalytic experiment, 50 mL of double-distilled water (ddH_2O) was mixed with Ag NPs and sonicated for 15 min in a sonication bath (Soltec® Sonica® 2400 ETH S3) to ensure uniform dispersion. Initially, a baseline was recorded for the dispersed Ag NP solution. The reaction mixture was then prepared by adding varying concentrations of BBR dye (Sigma-Aldrich®) in ethanol (500 mg/L) to the solution. The mixture was gently stirred in the dark for 30 min to establish equilibrium. Subsequently, the solution was irradiated using a warm white 10.5 W LED bulb (Philips), with light intensity monitored via a luxmeter (HD2302.0 Delta-Ohm, Padua, Italy). A peristaltic pump (Keenso DC) circulated the solution between the reaction flask and a 10 mm flow-through cuvette connected to a UV–Vis spectrophotometer (Cary 50, Varian®), which operated in the 400–750 nm range. UV–Vis spectra were recorded every 10 min. Total irradiation time of the samples was 90 min for all conducted experiments. Excluding repeated trials conducted for reproducibility, approximately 30 distinct experimental conditions were assessed by combining the parameter values presented in Table 1. This resulted in an ANN dataset of equivalent size. A brief discussion on the main outcomes of such activity will be discussed in the next parts of the work.

In Table 1, also the variable $y^0 = [Dye]^0/[Dye]^{max}$ is reported because the modeling results will be represented in dimensionless form. $[Dye]^{max}$ represents the highest initial dye concentration among the experimentally evaluated values, i.e. $[Dye]^{max} = 33.33$ (mg/L).

The efficiency of the BBR dye degradation by Ag NPs was calculated as shown in Eq. (1):

$$\eta(\%) = \left(1 - \frac{C}{C_0}\right)100$$

where η is the degradation efficiency, C_0 is the initial BBR dye concentration, and C is the final concentration of the dye at the end of the degradation period. The results are shown in the supplementary Table S1.

3. Mathematical modeling

3.1. Deterministic model

A deterministic mathematical model is first developed to describe the photodegradation process and support the implementation of the ANN-based predictive algorithm.

First-principles models of pollutants' photocatalytic degradation, which account for all kinetic phenomena and chemical species, result in ODE systems that lack analytical solutions (Daneshvar et al., 2004; Mills et al., 2015; Turchi and Ollis, 1990). The inherent mathematical complexity of such models often makes them impractical, prompting the adoption of simplified approaches, such as pseudo-first-order (Frontistis et al., 2012; Jiang et al., 2020; Rasoulifard et al., 2016; Seyed Dorraji et al., 2017) or pseudo-second-order models (Zulfiqar et al., 2019), which require significantly less computational effort. However, these approximations often fail to accurately describe the experimental degradation dynamics, which arise from the complex interplay of multiple kinetic phenomena (Dutta et al., 2010; Frontistis et al., 2012; Rasoulifard et al., 2016).

To address these limitations, a physically meaningful model of intermediate complexity is here proposed, offering a balance between simplicity and theoretical accuracy. Inspired by (Rasoulifard et al., 2016), this model conceptualizes photodegradation as a two-stage process: initially, pollutant removal occurs through adsorption onto the catalyst surface, followed by photocatalytic degradation of the adsorbed molecules once adsorption sites become saturated. These phenomena can be schematized as follows:

$$Dye \overset{K_1}{\rightleftharpoons} Dye_{ads} \qquad \text{R.1}$$

$$Dye_{ads} \overset{k_3}{\rightarrow} I_S \qquad \text{R.2}$$

where Dye_{ads} represents the species adsorbed on catalyst surface and I_s is the inactive product. The adsorption phenomenon is modelled as a reversible reaction (R.1), whose equilibrium constant, K_1, is defined as the ratio between the kinetic constants of the forward reaction (adsorption), k_1, and the kinetic constant of the inverse reaction (desorption), k_2. The degradation reaction, R.2, is modelled as an irreversible reaction, with k_3 being the kinetic constants of dye degradation.

According to the reactive scheme outlined above, the present model accounts for the adsorption/desorption equilibrium between bulk dye molecules ([Dye]) and adsorbed species ($[Dye]_{ads}$), as well as the degradation reaction involving $[Dye]_{ads}$. The following model equations are written by an isothermal photocatalytic reactor characterized by a perfect mixing condition.

$$\frac{d[Dye]}{dt} = -k_1[Dye] + k_2[Dye]_{ads} \qquad [Dye]|_{t=0} = [Dye]^0 \qquad (2)$$

$$\frac{d[Dye]_{ads}}{dt} = k_1[Dye] - k_2[Dye]_{ads} - k_3[Dye]_{ads} \qquad [Dye]_{ads}|_{t=0} = 0 \qquad (3)$$

Where [Dye] and $[Dye]_{ads}$ represent just the bulk and the adsorbed concentrations of BBR, respectively, while $[Dye]^0$ is the initial concentration of the pollutant. The two ordinary differential equations (ODE) of the model, Eqs. (2)–(3), constitute a first-order, linear and autonomous ODE system whose solution can be analytically computed.

The analytical solution of Eqs. (2)–(3), derived as detailed in the supplementary materials, is presented in Eq. (4) for the dimensionless variable $y = [Dye]/[Dye]^{max}$ where $[Dye]^{max}$ represents the highest initial dye concentration among the experimentally evaluated values, i.e. $[Dye]^{max} = 33.33$ (mg/L).

$$y(t) = \frac{y^0}{2\sqrt{\Delta}}\left[\left(-k_1 + k_2 + k_3 + \sqrt{\Delta}\right)e^{\frac{-\Sigma+\sqrt{\Delta}}{2}t} + \left(k_1 - k_2 - k_3 + \sqrt{\Delta}\right)e^{\frac{-\Sigma-\sqrt{\Delta}}{2}t}\right] \qquad (4)$$

Where:

$$\Sigma = k_1 + k_2 + k_3 \qquad (5)$$

$$\Delta = \Sigma^2 - 4k_1k_3 \qquad (6)$$

Eq. (4) describes the time evolution of the dimensionless dye concentration, (y) in the bulk solution during the photodegradation process. This analytical solution retains the computational efficiency of pseudo-first-order and pseudo-second-order models while incorporating the key macroscopic phenomena governing process dynamics.

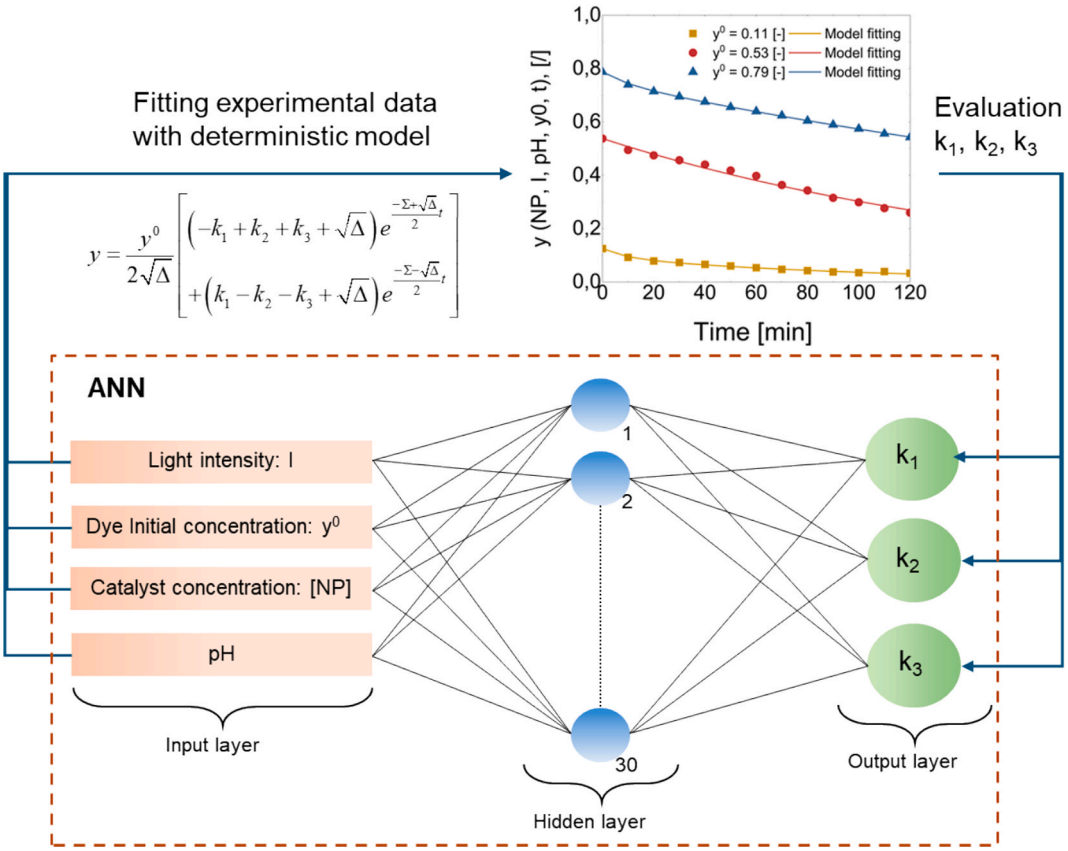

Fig. 1. Structure of the ANN and its interaction with the deterministic mathematical model.

3.2. ANN's dataset definition

The model solution, Eq. (3), and the experimental data were used to define the dataset for the neural network model training. Specifically, the kinetic parameters of Eq. (3), k_1, k_2 and k_3, were identified using the local optimization tool, fminsearchbnd (D'Errico, 2025), in MATLAB. Fminsearchbnd, as fminsearch, is adopted to solve nondifferentiable problems or problems with discontinuities, implementing the simplex search method (Lagarias et al., 1998). This way, a triplet of values k_1, k_2 and k_3 is obtained for each combination of experimental conditions experimentally evaluated (I, y^0, NP and pH). The resulting dataset, that is subsequently used to train and validate the ANN, consists of 30 input/output pairs, 10 for every microalgal strain evaluated, where the input consists of the operating conditions I, y^0, NP and pH and the output of the triplet, k_1, k_2 and k_3.

In this way, a triplet of kinetic parameters, k_1, k_2 and k_3, is determined for each combination of experimentally assessed conditions, including irradiance (I), initial biomass concentration (y^0), nanoparticle concentration (NP), and pH. The resulting dataset, subsequently utilized for the training and validation of the ANN, includes 30 input–output pairs, with 10 pairs corresponding to each evaluated microalgae strain. The input variables include I, y^0, NP and pH, while the output corresponds to the kinetic parameter triplet k_1, k_2 and k_3.

The confidence interval of each kinetic parameter (k_1, k_2 and k_3) was evaluated through Eq. (7) (Seber and Lee, 2003).

$$CI_{k_i} = t_{v,\gamma} \cdot se_{k_j} \tag{7}$$

where $t_{v,\gamma}$ is the upper $\gamma/2$ quantile for Student's T distribution with v degrees of freedom. The confidence level is set to $\gamma = 0.95$, while the number of degrees of freedom is given by v = n-p, with n representing the experimental data numerosity and p the number of parameters. The standard deviation of the specific parameter (se_{kj}) was computed from the diagonal elements of covariance matrix, as reported in Eq. (8).

$$se_{k_j} = \sqrt{cov(k_j)_{i,i}} \qquad \text{with} \qquad cov(k_j) = s^2 (J^T J)^{-1} \tag{8}$$

Where Jacobian matrix, J, is composed by the partial derivatives of the parameters, while s^2 is the Residual Sum of Square, normalized with respect to the number of degrees of freedom (v) (Seber and Lee, 2003).

3.3. Artificial neural network model

The relation between the process kinetic constants k_1, k_2, k_3 (output) and the operating conditions considered as inputs (I, y^0, [NP] and pH), even if complex and hardly explainable with deterministic models, should not include discontinuities. Therefore, according to the universal approximation theorem (Augustine, 2024), it can be described with a neural network with a single hidden layer implementing a nonlinear activation function (Augustine, 2024; Hornik et al., 1989). So, the multilayer perceptron (MLP) model, schematized in Fig. 1, is chosen as the simplest architecture to find a mapping function between the input and output space. It is worth noting that the microalgae strain used to produce catalyst nanoparticles would introduce discontinuity, if included as an input. Accordingly, three separate ANN models are implemented in parallel, one for each microalgae strain, and trained and validated using the corresponding dataset.

Specifically, every MLP model was built to learn and predict the relationship between the four operating conditions, (I, y^0, [NP] and pH) and the rate constants, k_1, k_2 and k_3. Both the input and output data are normalized by dividing each variable by its maximum value. This process ensures that the data is scaled appropriately for subsequent analysis The ANN model was implemented in MATLAB by considering a number of nodes for the hidden layer equal to N = 30. The MLP architecture can be expressed as [4: N: 3] ([input: hidden nodes or neurons: output]) and

visualized in Fig. 1.

Both hidden and output layers employ hyperbolic tangent sigmoid as activation function (tansig), shown in Eq. (9), allowing the network to capture complex, non-linear relationships in the data.

$$tansig(x) = \frac{2}{1+\exp(2x)} - 1 \tag{9}$$

For the training, the Levenberg-Marquardt backpropagation was used, which is MATLAB's default algorithm for feedforward networks. The training process was set to run for up to 2000 epochs, with an early stopping mechanism in place to prevent overfitting. The network's weights and biases were initially set to random values, providing a starting point for the learning process. To ensure robust learning and evaluation, the data was divided into three sets: 70 % for training, 15 % for validation, and 15 % for testing. Before feeding the data into the network, it's normalized by dividing each value by the maximum in its respective category, helping to standardize the input and output ranges.

Mean square error was selected as objective function to be minimized in the network training and it is accordingly used for evaluating the model predictive performance. Since the model is designed for multiple outputs, the mean squared error (MSE) is computed as the average of the MSE values corresponding to each parameter, as defined in Eq. (10).

$$MSE = \frac{MSE_{k_1} + MSE_{k_2} + MSE_{k_3}}{3} = \frac{1}{3}\sum_{j=1}^{3}\frac{1}{N}\sum_{i=1}^{N}\left(k_{j,i} - k_{j,i}^{exp}\right)^2 \tag{10}$$

Where j refers the three kinetic constants: j {1,2,3}, while i corresponds to the specific experimental trial. The mean squared error for every output (MSE_{k_j}) is computed on the basis of the difference between the predicted value ($k_{j,i}$) and the experimental one, $k_{j,i}^{exp}$, where N represents the dataset numerosity. Consistent with the definition of three separate datasets for the three microalgae strains, three neural networks with identical topology and hyperparameters were implemented (Fig. 1).

3.4. Synthetic data generation

The training and validation of neural network models are highly dependent on both the quality and quantity of available experimental data. In cases where real-world data are limited, as in this study, data augmentation techniques can be employed to generate artificial datasets that preserve the statistical properties and characteristics of the original data (Lu et al., 2023). Among the various methods used to synthesize data and enhance the performance of machine learning algorithms, noise-based augmentation strategies introduce controlled variations to the input data, thereby mitigating overfitting and improving model generalization (Arslan et al., 2019; Rasmussen, 2004).

In this work, Gaussian noise was incorporated into the existing datasets, as many real-world variables exhibit Gaussian distributions. This assumption is supported by the central limit theorem, which states that the sum of a sufficiently large number of independent and identically distributed random variables tends to follow a normal distribution. Consequently, Gaussian noise serves as an appropriate model for real-world imperfections, ensuring a realistic augmentation approach. Indeed, noise introduces controlled randomness into data, capturing the inherent imperfections and natural variability observed in real-world scenarios. This variability may arise from measurement errors, environmental influences, or intrinsic fluctuations within the data. By integrating noise into the dataset, the augmented data more accurately reflects the stochastic nature of practical applications, enhancing model robustness in the presence of unavoidable real-world disturbances.

Based on the consideration above we added Gaussian noise to both the input and the output of the ANN. As for the input data, noise was added by considering the uncertainty arising from measurement instruments or data acquisition techniques. Every measuring instrument has some associated error degree due to various factors such as the precision of the instrument, environmental conditions, and the measurement techniques used. This uncertainty is unavoidable and represents a sort of intrinsic "noise" in the collected data which can be exploited for synthetic data generation according to Eq. (11).

Table 2
Measurement errors of instruments involved in the experimental activity for data collection.

Symbol	Lux meter (Delta OHM, 2024) [%]	Scale (KERN & Sohn GmbH, 2017) [mg]	pH (Crison Instruments S.A., 2010) [/]
ε	5	0.2	0.01

$$\boldsymbol{X}_{aug} = \boldsymbol{X} + \boldsymbol{N}\left(0, \boldsymbol{\sigma}^{(\boldsymbol{X})}\right) \tag{11}$$

Where N represents the normal distribution function in Eq. (12):

$$N(X : \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2\right] \tag{12}$$

A synthetic input data, $\mathbf{X}_{aug}$ = [I, y^0, [NP], pH]$_{aug}$, is a vector generated from real experimental data, $\mathbf{X}$ = [I, y^0, [NP], pH], perturbed by a Gaussian noise whose standard deviation ($\boldsymbol{\sigma}^{(\mathbf{X})}$) corresponds to the measurement error that characterizes the specific measurement devices. Specifically, the single element of real data, X_k, is augmented by adding a gaussian noise, $N(0, \sigma_k^{(X)})$, where the superscript (X) refers to the input augmentation and the subscript k refers to the k^{th} input. Accordingly, $\sigma_k^{(X)}$ corresponds to the error associated to the measurement device used to measure X_k. Accordingly, $\sigma_k^{(X)}$ can be different for each measurement device. In this work, the measurement errors (ε) are those of the instruments used in the experimental campaign (Table 2). The measurement error for the lux meter is given in percentage form and thus the absolute value of the corresponding $\sigma_k^{(X)}$ is obtained by multiplying ε_k/100 for the real experimental value. Moreover, to obtain $\sigma_k^{(X)}$ associated to [NP] and y^0, the scale error is multiplied by $(1/V_r)$ and $1/([Dye]_{max} \cdot V_r)$, respectively.

Where V_r is the reactor volume where the dye degradation takes place. To the best of our knowledge, the described approach for input data augmentation, which further strengthens the ANN model's alignment with the underlying physics of the problem, has not been previously utilized. The results of a single ANN input data augmentation is reported in Fig. 2A–D only for the strain *C. vulgaris* for the sake of clarity. For each variable, a representative original point (blue) is shown with its 100 augmented points (red). The spread of augmented points reflects the specific noise level for each variable. The density plots show the Gaussian distribution of the augmented points around their original values, demonstrating how the augmentation preserves the statistical properties while accounting for measurement uncertainties. Dotted lines indicate the noise bounds for each original point.

A new approach is also proposed to generate a synthetic output vector, $\mathbf{Y}_{aug} = [k_1, k_2, k_3]_{aug}$ associated to $\mathbf{X}_{aug}$. Output augmentation is achieved through the introduction of Gaussian noise, with the standard deviation $\sigma_j^{(Y)}$, where the superscript (Y) refers to the output augmentation and the subscript j refers to the j^{th} output. Specifically, $\sigma_j^{(Y)}$ is determined based on the confidence interval of the parameter k_j, obtained during the fitting of deterministic model to experimental data (Eq. (13)).

$$\boldsymbol{Y}_{aug} = \boldsymbol{Y} + \boldsymbol{N}\left(0, \boldsymbol{\sigma}^{(\boldsymbol{Y})}\right) \tag{13}$$

Also in this case, $\sigma_j^{(Y)}$ can be different for the different elements of the output vector ($\mathbf{Y}$) since the confidence interval, computed with Eq. (7), is independently calculated for every single output, Y_j (Fig. 2E–G). In this study, the original dataset is augmented by generating 100 synthetic values for each real input data. First the same output value (equal to the experimental one) is associated with each synthetic input generated

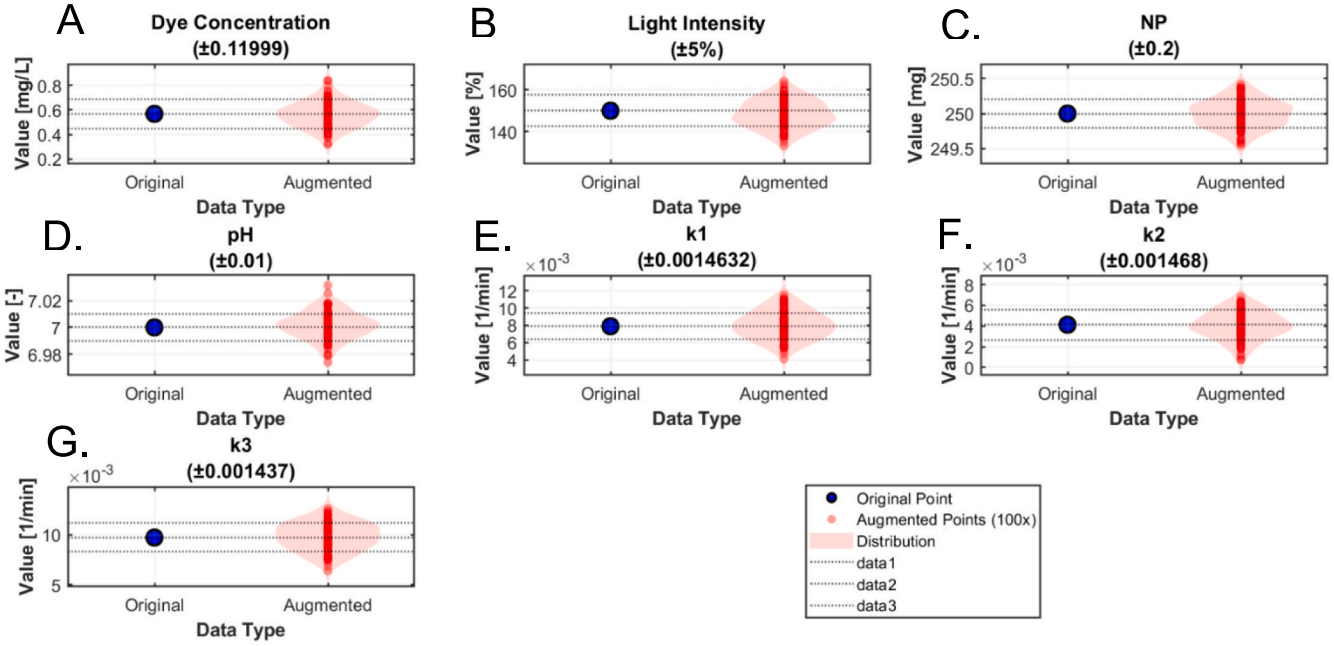

Fig. 2. Detailed view of augmentation for *C. vulgaris* Dataset: (A–D) Input, (E–F) Output data.

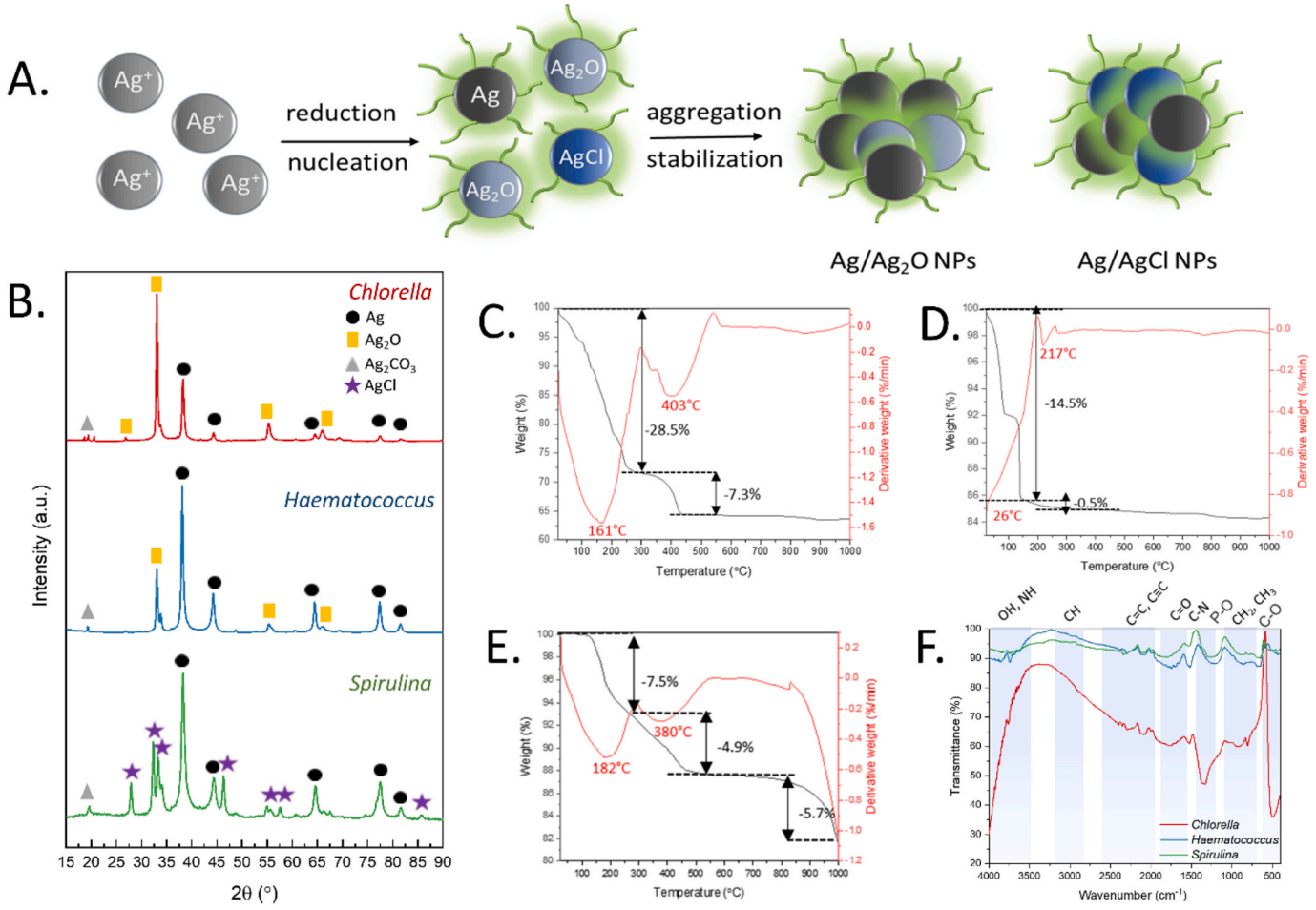

Fig. 3. Biological synthesis of Ag NPs and their characterization, (A) proposed mechanism of nanoparticle formation, (B) XRD analysis, (C) TGA analysis of Ag NPs synthesized using *C. vulgaris* extract, (D) TGA analysis of Ag NPs synthesized using *H. pluvialis* extract, (E) TGA analysis of Ag NPs synthesized using *S. platensis* extract, (F) FTIR analysis.

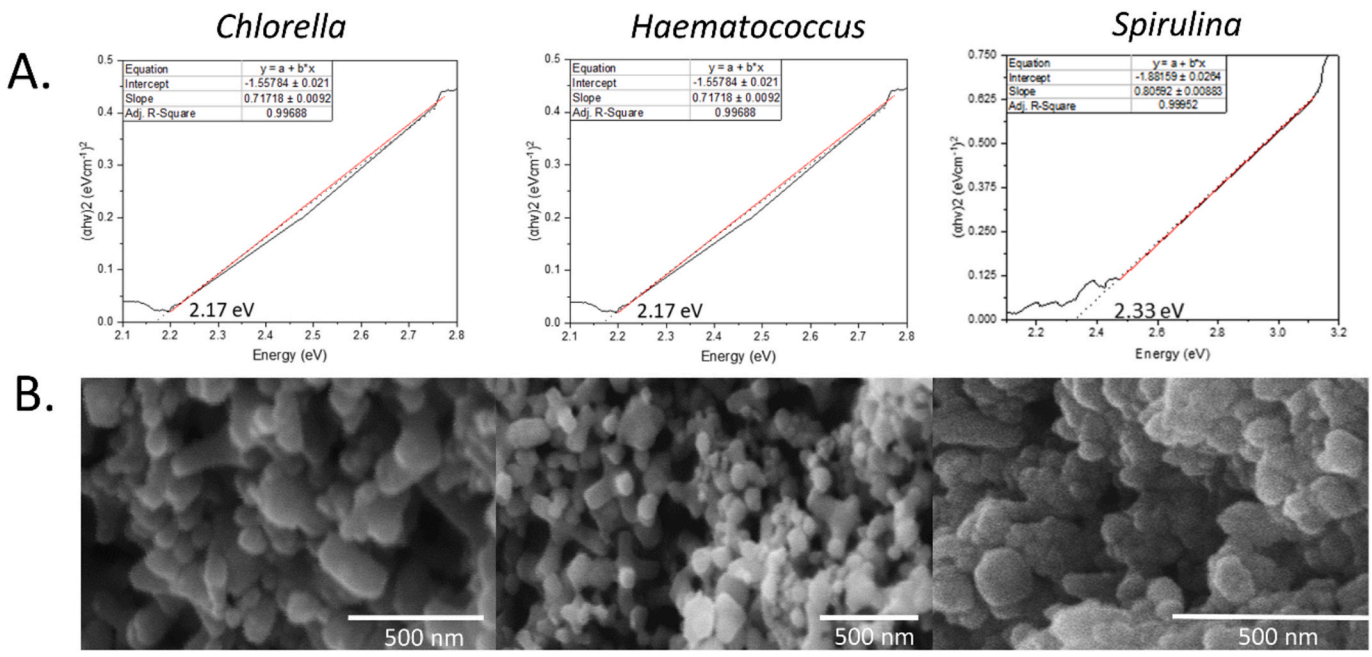

Fig. 4. Optical properties and morphology of Ag NPs synthesized using C. vulgaris, H. pluvialis and S. platensis, respectively, (A) Tauc plots, (B) SEM images.

starting from the same real input. Subsequently, Gaussian noise is randomly added to the real output associated with each synthetic input. Basically, synthetic input/output pairs are generated by adding gaussian noise to both input and output real data. This way, the variability of output associated to each synthetic input is ensured. The above procedure is applied to the 3 datasets available for each microalgae strain (*S. Platensis*, *H. Pluvialis* and *C. Vulgaris*) tested in experimental activity. A total of nine datasets were thus obtained.

By employing this approach, the realism of the synthetic data is enhanced, while also improving the ability of models trained on this data to handle variations and uncertainties encountered in real-world applications. This methodological choice is intentional and based on a fundamental principle: "the introduced synthetic noise is designed to replicate the expected variability in real data."

4. Results and discussion

4.1. Synthesis of Ag NPs with extracts from different microalgal strains

Using the methodology earlier described, AgNPs were successfully synthesized utilizing extracts from three distinct microalgal strains: *H. pluvialis* (Sidorowicz et al., 2025), *S. platensis* (Sidorowicz et al., 2024a), and *C. vulgaris* (Sidorowicz et al., 2024b). A graphical synopsis illustrating the key processes involved in the synthesis, along with the primary physicochemical characteristics of the resulting nanoparticles, is presented in Fig. 3. The characterization data presented in Fig. 3 were reproduced from our previous publications (Sidorowicz et al., 2025, 2024a,b) to provide reference for comparison.

The proposed mechanism of biological nanoparticle synthesis suggests that biomolecules from the microalgal extract act as reducing and stabilizing agents (Fig. 3A). The biological synthesis of NPs follows a process involving nucleation, growth, and stabilization. In the early stages, biomolecules such as proteins, polyphenols, and flavonoids act as reducing agents, facilitating the conversion of metal ions into metal atoms (Sidorowicz et al., 2023, 2021). The biomolecules selectively bind to specific crystal planes, guiding the aggregation of atomic clusters into structured crystalline NPs (El-Seedi et al., 2019; Sidorowicz et al., 2024c). As the latter grow, the biomolecules play a stabilizing role, adsorbing onto the NPs surface and preventing excessive aggregation. The formation of an organic capping layer introduces steric and electrostatic repulsion, ensuring nanoparticle dispersion and stability in solution.

The detailed characterization and results of the Ag NPs synthesis were previously studied (Sidorowicz et al., 2025, 2024a,b), and a summary of the key findings is provided here. The XRD pattern confirms the successful formation of crystalline nanoparticles, with characteristic diffraction peaks corresponding to the Ag, Ag_2O, Ag_2CO_3, and AgCl phases (Fig. 3B). The sharp and intense peaks indicate good crystallinity, while peak broadening suggests nanoscale dimensions. No additional impurity peaks were detected, confirming the purity of the synthesized NPs. The crystalline size for the obtained products was 16.07 nm, 14.27 nm, and 13.02 nm for Ag NPs obtained using *C. vulgaris*, *H. pluvialis*, and *S. platensis*, respectively (Sidorowicz et al., 2025, 2024a,b).

The organic content was analyzed by TGA and FTIR analyses (Fig. 3C–F). The TGA curves revealed the thermal stability and composition of Ag NPs. An initial weight loss at lower temperatures is attributed to the evaporation of residual moisture and volatile organic compounds. A second significant weight loss at higher temperatures corresponds to the decomposition of organic stabilizing agents, further confirming the involvement of biomolecules in NPs stabilization. The remaining residue after complete degradation represents the stable inorganic phase. The final weight for Ag NPs was recorded at around 65 %, 84 %, and 80 % for *C. vulgaris, H. pluvialis*, and *S. platensis*, respectively (Sidorowicz et al., 2025, 2024a,b).

The organic groups on the surface were further analyzed by FTIR analysis. The FTIR spectrum displays characteristic absorption bands corresponding to functional groups in the biological extract. The spectra were used to identify functional groups and confirm the presence of phytochemicals on the nanoparticle surface. The conclusions were drawn based on the presence or absence of characteristic absorption bands. Peaks around 3200–3500 cm^{-1} indicate O–H stretching vibrations from hydroxyl groups, while bands at 1600–1700 cm^{-1} correspond to C═O stretching, suggesting the presence of carboxyl groups (Sidorowicz et al., 2025, 2024a,b). The abundance of hydroxyl and carbonyl groups is especially important during Ag NPs synthesis. Hydroxyl groups donate electrons to reduce metal ions, while carbonyl and carboxyl groups facilitate metal binding, influencing nucleation and growth (Gebre, 2023). Differences among the samples suggest variations

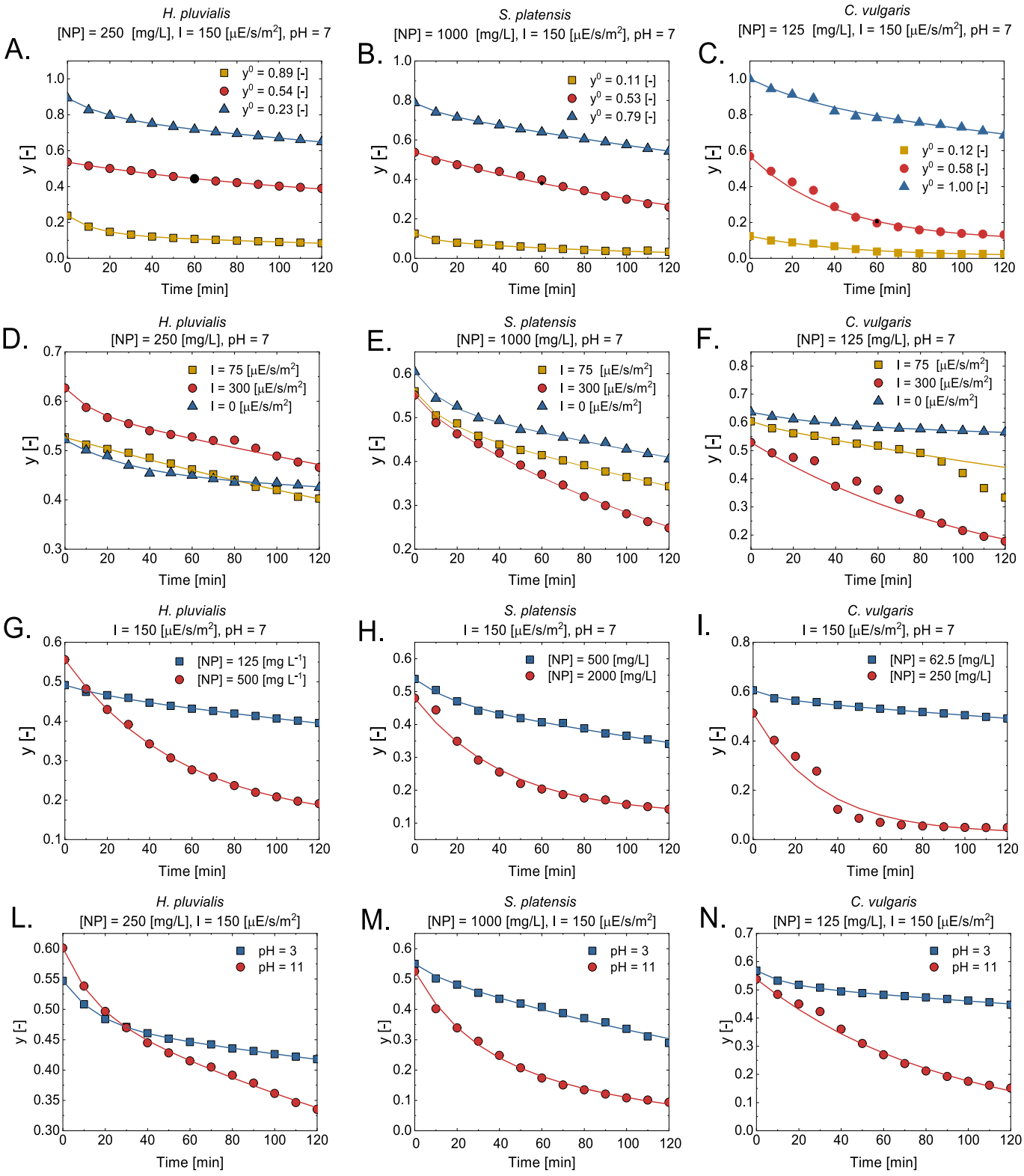

Fig. 5. Comparison of experimental data (dots) and model predictions (lines) for all evaluated operating conditions using Ag NPs synthesized from *H. pluvialis*, *S. platensis*, and *C. vulgaris* extracts.

of nature and abundance of biomolecules from different extracts, influencing the capping efficiency and interaction with the nanoparticle surface.

The optical properties of Ag NPs were examined by UV–Vis absorption study (Fig. 4A). The UV–Vis absorption spectrum exhibits a strong surface plasmon resonance (SPR) peak, characteristic of NP formation. The Tauc plot was used to determine the optical bandgap, showing a bandgap value within the expected range for the synthesized material. The calculated values for Ag NPs were 2.17 eV, 2.26 eV, and 2.33 eV for *C. vulgaris, H. pluvialis*, and *S. platensis*, respectively (Sidorowicz et al., 2025, 2024a,b). A lower bandgap value suggests enhanced light absorption in the visible spectrum, which can improve photocatalytic

Table 3
Values of k_1, k_2 and k_3 by fitting the experimental results with the deterministic model.

Strain	ID*	k_1 [min^{-1}]	k_2 [min^{-1}]	k_3 [min^{-1}]
H. pluvialis	HP_1	0.03578 ± 0.00072	0.038 ± 0.0016	0.0089 ± 0.0011
	HP_2	0.003606 ± 0.00084	0.004620 ± 0.0015	0.001097 ± 0.0001
	HP_3	0.0077 ± 0.0005	0.03612 ± 0.002	0.01103 ± 0.0016
	HP_4	0.005 ± 0.00077	0.02521 ± 0.0009	0.00341 ± 0.0016
	HP_5	0.0023 ± 0.00065	$5.87•10^{-5}$ ± 0.0016	0.08 ± 0.0016
	HP_6	0.00775 ± 0.00048	0.06 ± 0.0016	0.01918 ± 0.0026
	HP_7	0.0136 ± 0.00043	0.004843 ± 0.0022	0.001 ± 0.0032
	HP_8	0.003327 ± 0.00042	0.0238 ± 0.0011	0.02016 ± 0.0033
	HP_9	0.01397 ± 0.00067	0.0438 ± 0.0014	0.017498 ± 0.0014
	HP_10	0.009234 ± 0.00072	0.04763 ± 0.0018	0.00723 ± 0.0009
S. platensis	SP_1	0.03694 ± 0.00068	0.05928 ± 0.0023	0.02974 ± 0.0017
	SP_2	0.005932 ± 0.00073	0.00744 ± 0.0025	0.2463 ± 0.0020
	SP_3	0.007374 ± 0.00075	0.059 ± 0.0025	0.03695 ± 0.0018
	SP_4	0.01236 ± 0.00073	0.06 ± 0.000	0.01472 ± 0.000
	SP_5	0.012 ± 0.00075	0.0526 ± 0.000	0.02147 ± 0.000
	SP_6	0.0119 ± 0.00063	0.06 ± 0.0018	0.0711 ± 0.0011
	SP_7	0.01746 ± 0.00055	0.006012 ± 0.0023	4.38•10-5 ± 0.0015
	SP_8	0.0094 ± 0.00052	0.0370 ± 0.0021	0.01763 ± 0.0019
	SP_9	0.02433 ± 0.00069	0.0112 ± 0.0025	0.01938 ± 0.0011
	SP_10	0.009754 ± 0.0006	0.06 ± 0.002	0.05744 ± 0.0010
C. vulgaris	CV_1	0.0193 ± 0.0015	0.0022 ± 0.0015	(2.30 ± 0.22)• 10^{-5}
	CV_2	0.0201 ± 0.0016	0.004123 ± 0.0015	(3.47 ± 0.34)• 10^{-11}
	CV_3	0.0058 ± 0.0015	0.0166 ± 0.0014	0.0097 ± 0.0012
	CV_4	0.0023 ± 0.0011	0.01752 ± 0.0013	0.00167 ± 0.0009
	CV_5	0.0047 ± 0.0009	0.05952 ± 0.002	0.06911 ± 0.0012
	CV_6	0.0088 ± 0.0016	0.0034 ± 0.0009	0.6893 ± 0.0012
	CV_7	0.0298 ± 0.002	0.0015 ± 0.001	(1.34 ± 0.08)• 10-7
	CV_8	0.0057 ± 0.0012	0.06 ± 0.0009	0.018 ± 0.0013
	CV_9	0.01121 ± 0.0014	0.0033 ± 0.0011	0.6832 ± 0.0009
	CV_10	0.00789 ± 0.0018	0.0588 ± 0.0015	0.01146 ± 0.0012

* Notes: ID of the set of operating conditions adopted as defined in Table 1.

efficiency by generating more electron-hole pairs. (Liza et al., 2024) The slight variation in bandgap values among the samples may be attributed to differences in nanoparticle size, structural defects, or the influence of biomolecules from different extracts.

The morphology of the AgNPs obtained was observed by SEM analysis (Fig. 4B). SEM images reveal the morphological characteristics of the nanoparticles, showing a uniform size distribution and well-defined shapes. (Sidorowicz et al., 2025, 2024a,b) The nanoparticles appear to be spherical or quasi-spherical with minimal agglomeration, indicating effective stabilization by biomolecules. The quasi-spherical shape observed in some samples suggests slight structural distortions, potentially due to the interaction of organic capping agents with the nanoparticle surface during synthesis. While SEM offers useful insights into particle shape and aggregation, it does not provide sufficient resolution for precise size measurement at the nanoscale.

4.2. BBR photodegradation via Ag NPs: experimental results and deterministic modeling

Fig. 5 shows the time evolution of the dye concentration in its dimensionless form $y = [Dye]/[Dye]_{max}$ during the photocatalytic degradation experiments performed with Ag NPs obtained from the different microalgal strain extracts under different sets of operating conditions I, y^0, [NP] and pH. The experimental results in terms of [Dye] evolution Vs time are reported elsewhere (Sidorowicz et al., 2025, 2024a,b) and demonstrated that good BBR removal yields could be achieved with all the tested microalgal strains under specific operating conditions and. Notably, these results were achieved under visible light irradiation, highlighting the advantage of this approach over conventional photocatalytic degradation processes that typically rely on costly UV-based systems.

The deterministic model solution in Eq. (4) and the experimental data were used to define the dataset for the neural network model training. Specifically, the kinetic parameters of Eq. (4), k_1, k_2 and k_3, were identified using the local optimization tool, fminsearchbnd (D'Errico, 2025), in MATLAB. The tool is indeed adopted to solve non-differentiable problems or problems with discontinuities, implementing the simplex search method (Lagarias et al., 1998). Fig. 5A–N also illustrates the comparison between the experimentally obtained profiles and the corresponding model fit used for parameter identification across all tested microalgae strains and experimental conditions. The values of the parameters k_1, k_2 and k_3 allowing the best fitting of experimental data to the deterministic model are reported in Table 3 for each investigated set of operating conditions.

4.3. ANN simulations

Accordingly, the investigated operating conditions, namely light intensity (I), initial dye concentration (y_o), nanoparticle concentration (NP), and pH, summarized in Table 1, were used as input variables for the artificial neural network (ANN). The corresponding kinetic parameters (k_1, k_2, and k_3), detailed in Table 3, served as the output variables.

Accordingly, Table 3 reports the raw dataset (one for each microalgal strain), before any data augmentation,. This dataset was then augmented first by generating synthetic input taking into account the intrinsic errors in the input variables determination and converting into gaussian noise. Subsequently, Gaussian noise was also introduced to the ANN outputs, based on the standard deviations of the kinetic parameters k_1, k_2, and k_3 as reported in Table 3. The resulting augmented datasets were employed to train, validate, and test various feedforward neural network architectures, with the aim of assessing how the progressive integration of synthetic data influences model performance. This analysis was conducted independently for each microalgal strain, using the network architecture previously described.

4.3.1. ANN simulation results with raw dataset

In Fig. 6 the results obtained for each strain by training the ANN with only raw data are displayed.

Specifically, every parity plot shows the comparison between the real output data (observed values of k_1, k_2 and k_3) and the output computed by the network (network value). As clearly demonstrated in the parity plots, the data points do not follow the bisector trend, indicating that the ANN cannot be properly trained. Indeed, the original datasets do not contain enough data to explain how the operating conditions I, $[Dye]^0$, [NP] and pH, influence the value of the kinetic constants. It is worth

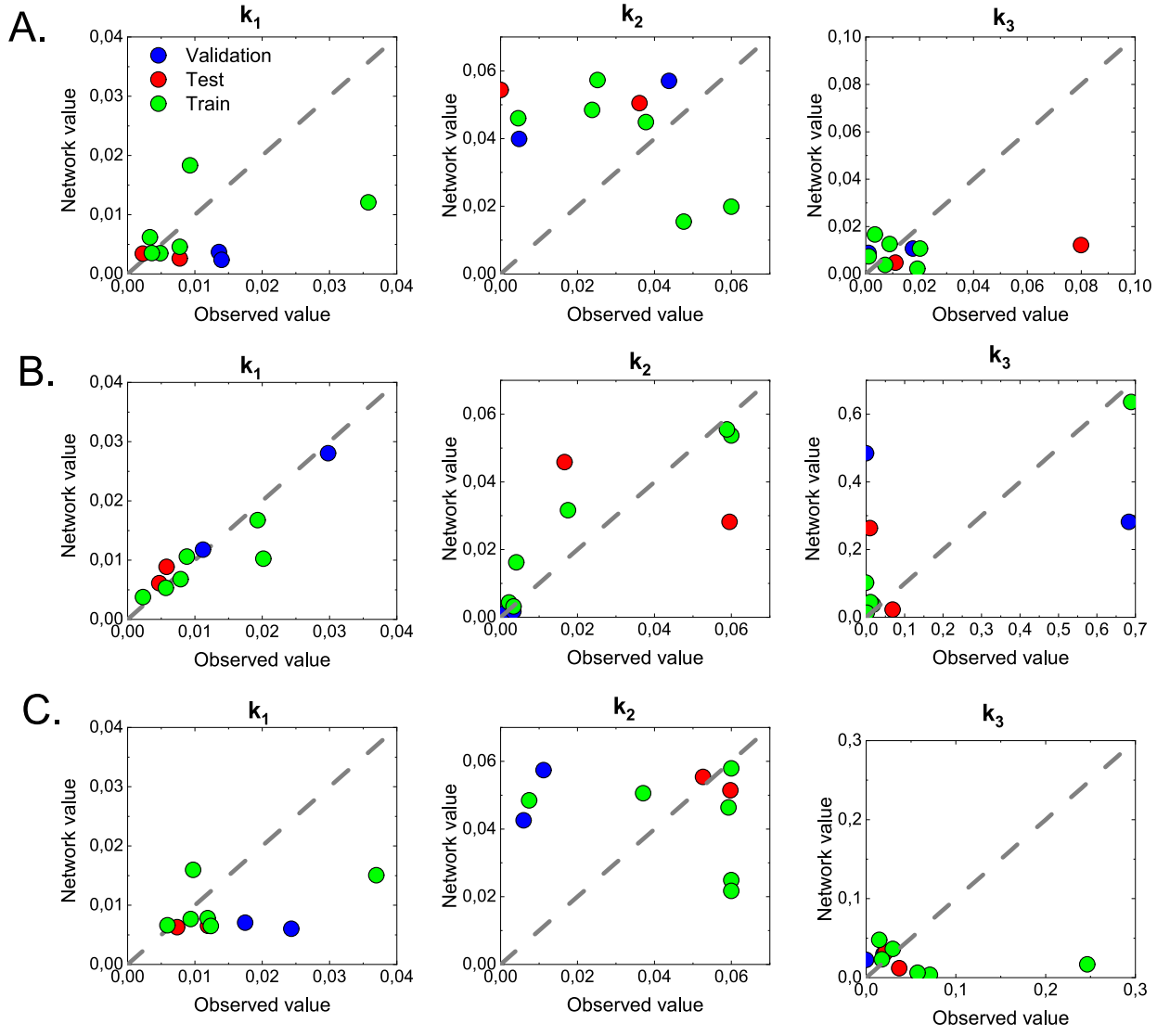

Fig. 6. ANN simulation results in terms of parity plots of kinetic parameters k_1, k_2 and k_3 for (A) *H. pluvialis*, (B) *C. vulgaris* and (C) *S. platensis*.

Table 4
Mean squared error (MSE) and coefficient of determination (R^2) of neural networks trained on the original dataset.

Network	*H. pluvialis*	*C. vulgaris*	*S. platensis*
MSE	7.9×10^{-4}	3.8×10^{-2}	4.5×10^{-4}
R^2	−1.4	0.23	−1.15

noting that the parity plots displayed in Fig. 6 show not only the ANN inaccuracy in predicting the kinetic constants (output) but also the unsatisfactory result of the training procedure, as highlighted by the trend of the training data in the parity plots.

In Table 4 the results of the network's simulation are reported in terms of mean squared error (MSE), computed by Eq. (10) and coefficient of determination (R^2), considering the portion of the dataset dedicated to validation and testing.

It is worth noting that the coefficient of determination, R^2, related to *H. pluvialis* and *S. platensis* is negative, which is mathematically consistent (Chicco et al., 2021) result indicating a poor interpretation of output distribution by the network.

4.3.2. ANN simulation results with input data augmentation

Fig. 7 shows the parity plots corresponding to the validation and test results obtained by training the networks on datasets augmented with Gaussian noise applied to the input variables. Compared to Fig. 6, a marked change in the distribution of data points is evident. The incorporation of synthetically generated input data, accounting physically-consistent measurement uncertainty, improves the network's predictive accuracy, as evidenced by the closer clustering of data points along the bisector.

However, the plots also reveal an additional feature in the data distribution: a pronounced vertical alignment of points for specific values on the x-axis. This artifact arises because the Gaussian noise was added solely to the inputs, while the corresponding outputs remained fixed. Consequently, the training dataset includes multiple input–output pairs with slightly perturbed inputs but identical targets, introducing a non-injective mapping that hampers the network's ability to learn an accurate input–output relationship.

Table 5 Shows the Mean squared error (MSE) and the coefficient of determination (R^2) obtained using the dataset with synthetic data generated in the ANN's input.

If compared with the corresponding values obtained with the raw dataset in Table 4, MSEs and R^2 were dramatically improved by adopting this data augmentation strategy. In particular, a significant increase was observed for the case of *C. vulgaris* and *H. pluvialis*. On the other hand, a pronounced vertical alignment of points for specific values

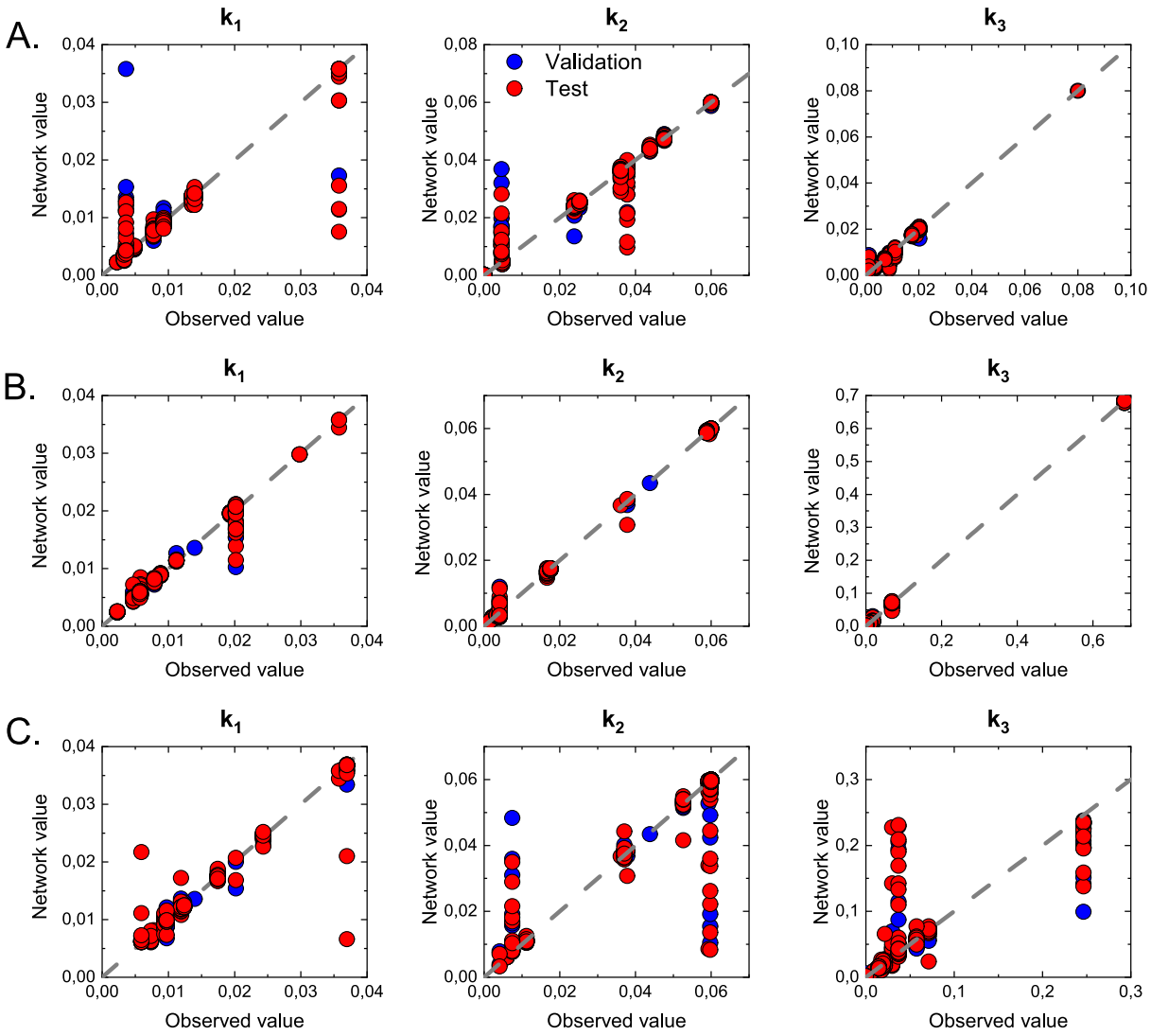

Fig. 7. ANN simulation results in terms of parity plots of kinetic parameters k_1, k_2 and k_3 for *H. pluvialis* (A), *C. vulgaris* (B) and *S. platensis* (C) obtained when using augmented data for the input.

Table 5
MSEs and R^2 of ANNS trained with the dataset obtained by augmenting the input data.

Network	*H. pluvialis*	*C. vulgaris*	*S. platensis*
Global MSE	1.34×10^{-5}	4.56×10^{-6}	4.43×10^{-4}
R^2	0.91	0.99	0.84

on the x-axis was observed.

4.3.3. ANN simulation results with input and output data augmentation

Finally, the training of the network was conducted using synthetic data also for the ANN output, specifically for the kinetic constants k_1, k_2 and k_3. This approach was adopted to investigate its potential in enhancing the network's performance. It is important to note that the Gaussian noise added to the output values was introduced following a physically meaningful rationale. Indeed, the standard deviations used for k_1, k_2 and k_3 were derived from the corresponding confidence values obtained by fitting experimental data with the deterministic model. Fig. 8 presents parity plots comparing the experimentally determined kinetic constants with the corresponding values predicted by the ANNs.

To enhance clarity and maintain consistency with Fig. 7, only validation and test datasets are shown. Notably, the introduction of noise to the synthetic outputs leads to a reduced vertical dispersion of the data points. This is because the noise enhances the injectivity of the input–output mapping, thereby improving the ANN's ability to distinguish between similar input patterns.

Specifically, the unusual condition of the only-input augmented dataset, consisting of different input/output pairs which share the same output value with slightly different input, hinders the learning process of the ANN, as it introduces ambiguities that make it harder to generalize. This condition no longer occurs in the current dataset where the noise is applied to both synthetic input and output. As a direct consequence, the data distribution in Fig. 8 improves compared to Fig. 7, and the vertical trend observed in the parity plots of Fig. 7 tends to vanish. This improvement is also confirmed by the MSE and R^2 coefficients (Table 6), which reflect a slight yet positive change.

4.3.4. Results comparison and discussion

As shown in Fig. 5, the deterministic model (Eq. (4)) accurately describes the experimental data, successfully capturing the time profile of the pollutant concentration as photodegradation evolves. This aspect is a fundamental prerequisite to guarantee that the data used for training the neural network retain physical significance and accurately represent the underlying physical process.

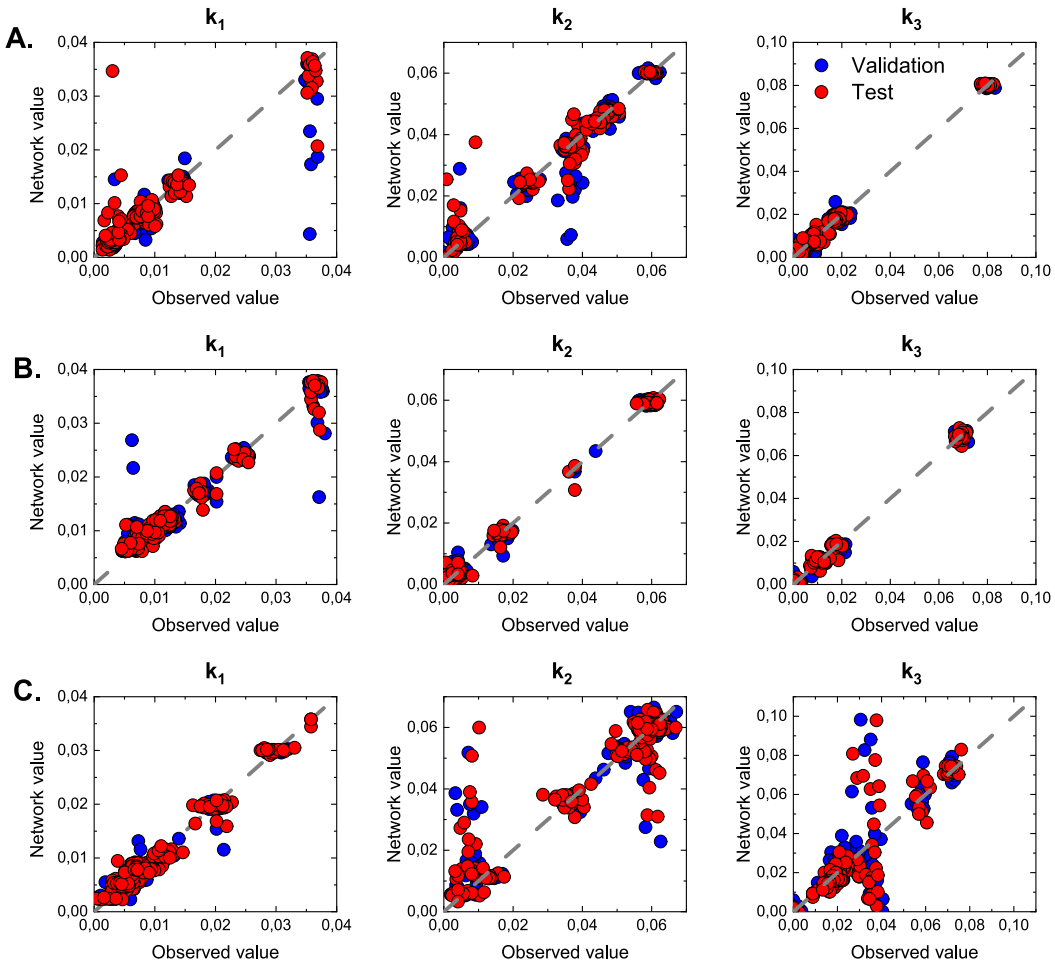

Fig. 8. ANN simulation results in terms of parity plots of kinetic parameters k_1, k_2 and k_3 for (A) *H. pluvialis*, (B) *C. vulgaris* and (C) *S. platensis* obtained when using augmentation for both input and output data.

Table 6
MSEs and R^2 of ANNs trained with the dataset obtained by augmenting both input and output data.

Network	*H. pluvialis*	*C. vulgaris*	*S. platensis*
Global MSE	1.31×10^{-5}	3.26×10^{-6}	4.18×10^{-4}
R^2	0.93	0.99	0.85

However, the available experimental data are not sufficient to train a neural network, even with a minimal architecture in terms of layers and neurons. Therefore, synthetic data are employed, generated by incorporating measurement uncertainty and thus adhering to a paradigm rooted in the physics of the original experiment. The comparison between Fig. 6 and Fig. 7 clearly demonstrates how the use of data augmentation improves the performance of the trained neural networks, allowing its output to better represent the dataset. This graphical improvement is further confirmed by the reduction in MSE and the increase in R^2 which respectively quantify the decrease in error and the improved ability to grasp the data distribution with respect to the operating conditions (input data).

In the final section an additional dataset is defined by integrating synthetic data generated not only by considering measurement uncertainty, but also by incorporating the confidence intervals, that characterize the parameter identification, to generate synthetic output. The comparison between the Fig. 7 and Fig. 8 highlights how the use of the Input-Ouput Augmented (AIO) dataset enables better clustering around the identity line for most parameters compared to the Input Augmented (AI) dataset. Specifically, the vertical spread observed in k_1 and k_2 subplot of Fig. 7A, in k_1 subplot of Fig. 7B and k2 and k3 subplot of Fig. 7C is significantly reduced when Gaussian noise is also applied to the output. Indeed, Fig. 8 displays tighter clustering in every subplot, demonstrating an enhanced ability to produce physically consistent and reliable simulation results without increasing the size training dataset, which remains constant between AI and AIO.

However, it is important to highlight that the corresponding improvement in the statistical parameters (MSE and R^2) is not as pronounced as expected, as shown in Fig. 9.

This suggests that, while the qualitative results are promising, further refinements are needed to achieve more substantial improvements in model accuracy. Accordingly, future efforts should focus on minimizing the confidence intervals that characterize the parameter identification and consequently define the dispersion of the Gaussian noise used to generate the synthetic output data.

5. Conclusion

This work introduces a physics-aware hybrid modeling framework for the photocatalytic degradation of Brilliant Blue R (BBR) dye using Ag NPs synthesized from microalgal extracts. The process was modeled by explicitly considering the main phenomena involved in photocatalytic degradation: adsorption–desorption equilibrium and degradation kinetics. A two-ODEs model framework was developed and proposed as an alternative to pseudo-first-order and pseudo-second-order models, which typically fail to capture the influence of multiple physico-

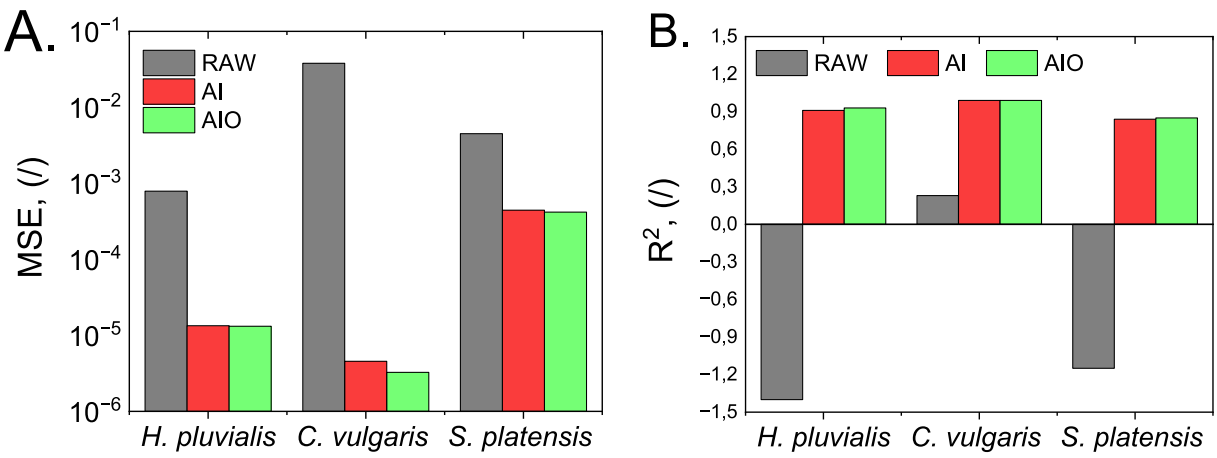

Fig. 9. Comparison between the MSEs (A) and R^2 (B) obtained for the different strains using the adopted datasets. RAW: Raw (not-augmented) dataset; AI: Input-augmented dataset; AIO: Input and output augmented dataset.

chemical phenomena on degradation dynamics. Despite the increased mathematical complexity, the proposed model enables accurate estimation of key kinetic parameters (k_1, k_2, k_3) through analytical solutions. These parameters were then used to train an ANN, providing a fast, portable, and physically consistent simulation tool. To address the limited experimental dataset, a novel data augmentation strategy was implemented by introducing Gaussian noise into both input and output data. This approach preserved the physical relevance of the synthetic data by grounding noise levels in measurement uncertainty and parameter confidence intervals. Comparative evaluation of models trained with raw, input-augmented, and input–output-augmented datasets showed that the dual augmentation strategy significantly improved model performance, reducing mean squared errors and enhancing input–output mapping injectivity. The model was successfully applied to the BBR photodegaradation by AgNPs synthesized from *H. pluvialis*, *S. platensis*, and *C. vulgaris*, demonstrating robustness across catalysts with differing physicochemical properties. The ANN effectively predicts photodegradation behavior beyond experimentally tested conditions, supporting its potential use for process optimization and scale-up. Overall, the proposed methodology advances photocatalytic modeling by coupling mechanistic interpretability with machine learning flexibility, while maintaining physical plausibility through informed data augmentation.

CRediT authorship contribution statement

Federico Atzori: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Bartolomeo Cosenza:** Writing – review & editing, Writing – original draft, Supervision, Software, Investigation, Formal analysis, Data curation, Conceptualization. **Federico Zedda:** Writing – review & editing, Validation, Investigation, Data curation. **Agnieszka Sidorowicz:** Writing – original draft, Visualization, Investigation, Data curation. **Giacomo Fais:** Writing – review & editing, Investigation. **Giacomo Cao:** Writing – review & editing, Resources, Funding acquisition. **Alessandro Concas:** Writing – review & editing, Visualization, Validation, Supervision, Resources, Methodology, Data curation, Conceptualization.

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

F.A. and F.Z. were supported by an International PhD Program in Innovation Sciences and Technologies at the University of Cagliari, Italy.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ces.2025.121945.

Data availability

Data will be made available on request.

References

Arslan, M., Guzel, M., Demirci, M., Ozdemir, S., 2019. SMOTE and gaussian noise based sensor data augmentation. In: International Conference on Computer Science and Engineering. https://doi.org/10.1109/UBMK.2019.8907003.

Augustine, M.T., 2024. A Survey on Universal Approximation Theorems 1–10.

Chicco, D., Warrens, M.J., Jurman, G., 2021. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comput. Sci. 7, 1–24. https://doi.org/10.7717/PEERJ-CS.623.

Cosenza, B., Miccio, M., Pannocchia, G., Sammarco, I., 2024. Advanced simulation model for studying biofuel-producing microalgae populations. Chem. Eng. Trans. 109, 559–564. https://doi.org/10.3303/CET24109094.

Crison Instruments S.A., BASIC 20 Technical Data Sheet., https://www.crisoninstruments.com/site/pdfs_en/basic_20.pdf, 2010 (accessed 2025).

D'Errico, J., 2025. fminsearchbnd, fminsearchcon [WWW Document].

Daneshvar, N., Rabbani, M., Modirshahla, N., Behnajady, M.A., 2004. Kinetic modeling of photocatalytic degradation of Acid Red 27 in UV/TiO2 process. J. Photochem. Photobiol. A Chem. 168, 39–45. https://doi.org/10.1016/j.jphotochem.2004.05.011.

de Souza, M.L., Corio, P., 2013. Effect of silver nanoparticles on TiO2-mediated photodegradation of Alizarin Red S. Appl. Catal. B Environ. https://doi.org/10.1016/j.apcatb.2013.02.012.

Dutta, S., Parsons, S.A., Bhattacharjee, C., Bandhyopadhyay, S., Datta, S., 2010. Development of an artificial neural network model for adsorption and photocatalysis of reactive dye on TiO2 surface. Expert Syst. Appl. https://doi.org/10.1016/j.eswa.2010.06.090.

El-Seedi, H.R., El-Shabasy, R.M., Khalifa, S.A.M., Saeed, A., Shah, A., Shah, R., Iftikhar, F. J., Abdel-Daim, M.M., Omri, A., Hajrahand, N.H., Sabir, J.S.M., Zou, X., Halabi, M.F., Sarhan, W., Guo, W., 2019. Metal nanoparticles fabricated by green chemistry using natural extracts: Biosynthesis, mechanisms, and applications. RSC Adv. 9, 24539–24559. https://doi.org/10.1039/c9ra02225b.

Frontistis, Z., Daskalaki, V.M., Hapeshi, E., Drosou, C., Fatta-Kassinos, D., Xekoukoulotakis, N.P., Mantzavinos, D., 2012. Photocatalytic (UV-A/TiO 2) degradation of 17α- ethynylestradiol in environmental matrices: Experimental studies and artificial neural network modeling. J. Photochem. Photobiol. A Chem. 240, 33–41. https://doi.org/10.1016/j.jphotochem.2012.05.007.

Gebre, S.H., 2023. Bio-inspired synthesis of metal and metal oxide nanoparticles: the key role of phytochemicals. J. Clust. Sci. 34, 665–704. https://doi.org/10.1007/s10876-022-02276-9.

Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. Neural Networks. https://doi.org/10.1016/0893-6080(89)90020-8.

Irineu, M.D., de Aquino, R.V.S., Barbosa, A.A., Do Nascimento Júnior, W.J., da Silva, J. P., Pacheco, J.G.A., da Rocha, O.R.S., 2022. Degradation of textile dye mixture by heterogeneous photocatalysis employing neural network modeling. Desalin. Water Treat. https://doi.org/10.5004/dwt.2022.29059.

Jiang, Z., Hu, J., Zhang, X., Zhao, Y., Fan, X., Zhong, S., Zhang, H., Yu, X., 2020. A generalized predictive model for TiO2–Catalyzed photo-degradation rate constants of water contaminants through artificial neural network. Environ. Res. 187, 109697. https://doi.org/10.1016/j.envres.2020.109697.

Joseph, S., Mathew, B., 2015. Facile synthesis of silver nanoparticles and their application in dye degradation. Mater. Sci. Eng. B. https://doi.org/10.1016/j.mseb.2015.02.007.

Joseph, S., Mathew, B., 2014. Microwave-assisted facile synthesis of silver nanoparticles in aqueous medium and investigation of their catalytic and antibacterial activities. J. Mol. Liq. https://doi.org/10.1016/j.molliq.2014.06.008.

Delta OHM, Operating manual - Photo-radiometer HD2302.0, Version 1.10., https://environmental.senseca.com/wp-content/uploads/document/DeltaOHM_HD2302.0_manual_ENG.pdf, 2024 (accessed 2025).

Lagarias, J.C., Reeds, J.A., Wright, M.H., Wright, P.E., 1998. Convergence properties of the Nelder-Mead simplex method in low dimensions. SIAM J. Optim. https://doi.org/10.1137/S1052623496303470.

Liza, T.Z., Tusher, M.M.H., Anwar, F., Monika, M.F., Amin, K.F., Asrafuzzaman, F.N.U., 2024. Effect of Ag-doping on morphology, structure, band gap and photocatalytic activity of bio-mediated TiO2 nanoparticles. Results Mater. 22, 100559. https://doi.org/10.1016/j.rinma.2024.100559.

KERN & Sohn GmbH, Operating instructions – Analytical balance KERN ADB, Version 2.0., https://dok.kern-sohn.com/manuals/files/English/ADB-BA-e-1720.pdf, 2017 (accessed 2025).

Lu, Y., Shen, M., Wang, H., Wang, X., van Rechem, C., Fu, T., Wei, W., 2023. Machine Learning for Synthetic Data Generation: A Review 14, 1–20.

Mills, A., O'Rourke, C., Moore, K., 2015. Powder semiconductor photocatalysis in aqueous solution: An overview of kinetics-based reaction mechanisms. J. Photochem. Photobiol. A Chem. 310, 66–105. https://doi.org/10.1016/j.jphotochem.2015.04.011.

Pal, N.K., Kryschi, C., 2016. Improved photocatalytic activity of gold decorated differently doped TiO2 nanoparticles: A comparative study. Chemosphere. https://doi.org/10.1016/j.chemosphere.2015.10.060.

Parmar, N., Srivastava, J.K., 2022. Process optimization and kinetics study for photocatalytic ciprofloxacin degradation using TiO2 nanoparticle: A comparative study of artificial neural network and surface response methodology. J. Indian Chem. Soc. https://doi.org/10.1016/j.jics.2022.100584.

Peyghami, A., Moharrami, A., Rashtbari, Y., Afshin, S., Vosuoghi, M., Dargahi, A., 2023. Evaluation of the efficiency of magnetized clinoptilolite zeolite with Fe3O4 nanoparticles on the removal of basic violet 16 (BV16) dye from aqueous solutions. J. Dispers. Sci. Technol. https://doi.org/10.1080/01932691.2021.1947847.

Pulido Melián, E., González Díaz, O., Doña Rodríguez, J.M., Colón, G., Navío, J.A., Macías, M., Pérez Peña, J., 2012. Effect of deposition of silver on structural characteristics and photoactivity of TiO 2-based photocatalysts. Appl. Catal. B Environ. 127, 112–120. https://doi.org/10.1016/j.apcatb.2012.08.007.

Rajkumar, R., Ezhumalai, G., Gnanadesigan, M., 2021. A green approach for the synthesis of silver nanoparticles by Chlorella vulgaris and its application in photocatalytic dye degradation activity. Environ. Technol. Innov. 21, 101282. https://doi.org/10.1016/J.ETI.2020.101282.

Rasmussen, C.E., 2004. Gaussian Processes in machine learning. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics). https://doi.org/10.1007/978-3-540-28650-9_4.

Rasoulifard, M.H., Seyed Dorraji, M.S., Amani-Ghadim, A.R., Keshavarz-Babaeinezhad, N., 2016. Visible-light photocatalytic activity of chitosan/polyaniline/CdS nanocomposite: kinetic studies and artificial neural network modeling. Catal. A Gen. Appl. https://doi.org/10.1016/j.apcata.2016.01.002.

Seber, G.A., Lee, A.J., 2003. Linear Regression Analysis. John Wiley & Sons.

Seyed Dorraji, M.S., Amani-Ghadim, A.R., Rasoulifard, M.H., Daneshvar, H., Sistani Zadeh Aghdam, B., Tarighati, A.R., Hosseini, S.F., 2017. Photocatalytic activity of g-C3N4: An empirical kinetic model, optimization by neuro-genetic approach and identification of intermediates. Chem. Eng. Res. Des. https://doi.org/10.1016/j.cherd.2017.09.012.

Sidorowicz, A., Atzori, F., Zedda, F., Fais, G., Loy, F., Licheri, R., Lai, N., Desogus, F., Cao, G., Concas, A., 2025a. Novel experimental and theoretical study on the synthesis and use of microalgae-derived silver nanomaterials for water purification. J. Water Process Eng. 69. https://doi.org/10.1016/j.jwpe.2024.106831.

Sidorowicz, A., Atzori, F., Zedda, F., Fais, G., Loy, F., Licheri, R., Lai, N., Desogus, F., Cao, G., Concas, A., 2025b. Novel experimental and theoretical study on the synthesis and use of microalgae-derived silver nanomaterials for water purification. J. Water Process Eng. 69, 106831. https://doi.org/10.1016/j.jwpe.2024.106831.

Sidorowicz, A., Fais, G., Casula, M., Borselli, M., Giannaccare, G., Locci, A.M., Lai, N., Orrù, R., Cao, G., Concas, A., 2023. Nanoparticles from microalgae and their biomedical applications. Mar. Drugs 21, 352. https://doi.org/10.3390/md21060352.

Sidorowicz, A., Fais, G., Desogus, F., Loy, F., Licheri, R., Lai, N., Cao, G., Concas, A., 2024a. Eco-friendly photocatalytic treatment of dyes with Ag nanoparticles obtained through sustainable process involving spirulina platensis. Sustainability 16, 8758. https://doi.org/10.3390/su16208758.

Sidorowicz, A., Fais, G., Desogus, F., Loy, F., Licheri, R., Lai, N., Locci, A.M., Cincotti, A., Orrù, R., Cao, G., Concas, A., 2024b. Optimization of Brilliant Blue R photocatalytic degradation by silver nanoparticles synthesized using Chlorella vulgaris. Environ. Sci. Pollut. Res. https://doi.org/10.1007/s11356-024-34967-3.

Sidorowicz, A., Szymański, T., Rybka, J.D., 2021. Photodegradation of biohazardous dye brilliant blue r using organometallic silver nanoparticles synthesized through a green chemistry method. Biology (Basel) 10, 784. https://doi.org/10.3390/biology10080784.

Sidorowicz, A., Yigit, N., Wicht, T., Stöger-Pollach, M., Concas, A., Orrù, R., Cao, G., Rupprechter, G., 2024c. Microalgae-derived Co3O4 nanomaterials for catalytic CO oxidation. RSC Adv. 14, 4575–4586. https://doi.org/10.1039/D4RA00343H.

Syahirah Kamarudin, N., Jusoh, R., Dina Setiabudi, H., Fateha Sukor, N., 2018. Photodegradation of methylene blue using phyto-mediated synthesis of silver nanoparticles: Effect of calcination treatment, in: Materials Today: Proceedings. https://doi.org/10.1016/j.matpr.2018.07.059.

Turchi, C.S., Ollis, D.F., 1990. Photocatalytic degradation of organic water contaminants: Mechanisms involving hydroxyl radical attack. J. Catal. 122, 178–192. https://doi.org/10.1016/0021-9517(90)90269-P.

Vidyasagar, N., Patel, R.R., Singh, S.K., Singh, M., 2023. Green synthesis of silver nanoparticles: methods, biological applications, delivery and toxicity. Mater. Adv. 4, 1831–1849. https://doi.org/10.1039/D2MA01105K.

Wu, Z.C., Zhang, Y., Tao, T.X., Zhang, L., Fong, H., 2010. Silver nanoparticles on amidoxime fibers for photo-catalytic degradation of organic dyes in waste water. Appl. Surf. Sci. https://doi.org/10.1016/j.apsusc.2010.08.022.

Zulfiqar, M., Samsudin, M.F.R., Sufian, S., 2019. Modelling and optimization of photocatalytic degradation of phenol via TiO2 nanoparticles: an insight into response surface methodology and artificial neural network. J. Photochem. Photobiol. A Chem. https://doi.org/10.1016/j.jphotochem.2019.112039.