

Retrieval of Coastal Biogeochemical Parameters From Near-Surface Hyperspectral Remote Sensing Reflectance Using Physics-Aware Meta-Learning

Yiqing Guo, Nagur R. C. Cherukuru, Eric A. Lehmann, S. L. Kesav Unnithan, Tim J. Malthus, Gemma Kerrisk, Xiubin Qi, Faisal Islam, Tisham Dhar, and Mark J. Doubell

Abstract—Hyperspectral in situ sensing has shown promise in retrieving aquatic biogeochemical (BGC) parameters, such as total suspended solids (TSS), dissolved organic carbon (DOC), and total chlorophyll-a (TChl-a), for cost-effective management of coastal water quality. However, generalising such retrieval algorithms across water bodies remains challenging, as the relationship between remote sensing reflectance (R_{rs}) and BGC parameters can vary considerably from one region to another due to regional distinctions in environmental conditions and biogeochemistry that lead to different BGC ranges and bio-optical properties. In this study, we propose a two-stage physics-aware meta-learning framework for retrieving coastal BGC parameters from near-surface R_{rs} observations. In the first stage, a bio-optical forward model is used to generate a large physics-guided synthetic dataset based on an in situ bio-optical spectral library. This dataset is then used to pretrain a region-agnostic base model with meta-learning, allowing the model to learn fundamental physical relationships. In the second stage, the pre-trained base model is adapted to specific regions via model finetuning with local samples. To evaluate the proposed approach, we collected in situ hyperspectral R_{rs} and BGC measurements from five geographically distinct sites in Australian coastal waters. Our experimental results suggest the following: (1) the BGC parameters and their corresponding hyperspectral R_{rs} signatures exhibit clear regional distinctions among the experimental sites; (2) the synthetic dataset, generated under guidance of the bio-optical forward model and the bio-optical spectral library, is physically plausible and closely aligned with real-world samples in both parameter distributions and inter-parameter correlations; (3) when aggregated across the experimental sites, the proposed approach achieves improvements of 3.2%, 6.3%, and 14.1% in Log- R^2 for TSS, DOC, and TChl-a, respectively, while reducing

Log-MAE by 22.1%, 2.7%, and 14.8% relative to the best benchmark model; and (4) time series of in situ measured and model-predicted BGC parameters show good agreement in both magnitude and temporal dynamics, highlighting the potential of in situ hyperspectral sensing as a cost-effective solution for continuous time-series monitoring of coastal water quality. These results demonstrate that the proposed physics-aware meta-learning framework provides a robust and adaptable approach for accurate BGC retrieval using hyperspectral in situ sensing.

Index Terms—Coastal water quality, biogeochemical parameters, hyperspectral data, physics-aware learning, meta-learning.

I. INTRODUCTION

Water quality is an important indicator of aquatic ecosystem health [1]–[3]. Degradation of water quality in aquatic systems undermines ecosystem functions and services [4]. Timely, accurate, and continuous monitoring of water quality is therefore critical for effective management and mitigation of environmental impacts, particularly in Australian coastal waters where many regions are vulnerable to anthropogenic pressures and climate-driven change [3], [5], [6].

Key indicators of water quality include biogeochemical (BGC) parameters, such as total suspended solids (TSS), dissolved organic carbon (DOC), total chlorophyll-a (TChl-a), and their inherent optical properties (IOPs) [7]–[9]. TSS represents all particulate matter suspended in the water column, including phytoplankton and non-algal particles (NAP) [10]. TChl-a is widely used as an indicator of phytoplankton biomass [11], while NAP comprises the non-phytoplankton fraction of TSS [12]. In addition to suspended particulates, dissolved components also play a critical role in the bio-optical properties of natural waters [13]. Dissolved organic carbon (DOC) represents the carbon content of the diverse pool of dissolved organic matter (DOM) [14], [15], while its optically active fraction, known as chromophoric dissolved organic matter (CDOM), is widely used as a proxy for the source and compositional characteristics of DOM [16]. The IOPs, including absorption and scattering coefficients, serve as the optical fingerprints of BGC parameters, characterising the intrinsic interactions between water constituents and light [17]. The mass-normalised IOPs, known as specific IOPs (SIOPs), describe how efficiently substances in water absorb or scatter light per unit concentration [17]. Accurate local calibration of SIOPs is essential, as they may vary from one region to another due to differences in the composition and

The code and data are available at: <https://github.com/yiqing-csiro/wq-meta>. This work is supported by Commonwealth Scientific and Industrial Research Organisation (CSIRO) AquaWatch Australia Program, CSIRO A14Missions, South Australian Research and Development Institute (SARDI), CSIRO Technology, CSIRO Environment, CSIRO Space and Astronomy, CSIRO Earth Analytics Science and Innovation (EASI) platform, and CSIRO AquaWatch Data Service (ADS). (Corresponding author: Yiqing Guo.)

Yiqing Guo and Eric A. Lehmann are with CSIRO Technology, Acton, ACT 2601, Australia (e-mail: yiqing.guo@csiro.au; eric.lehmann@csiro.au).

Nagur R. C. Cherukuru is with CSIRO Environment, Acton, ACT 2601, Australia (e-mail: nagur.cherukuru@csiro.au).

S. L. Kesav Unnithan is with CSIRO Space and Astronomy, Acton, ACT 2601, Australia (e-mail: kesav.unnithan@csiro.au).

Tim J. Malthus, Gemma Kerrisk, and Faisal Islam are with CSIRO Environment, Dutton Park, QLD 4102, Australia (e-mail: tim.malthus@csiro.au; gemma.kerrisk@csiro.au; faisalislam0407@gmail.com).

Xiubin Qi is with CSIRO Space and Astronomy, Kensington, WA 6151, Australia (e-mail: xiubin.qi@csiro.au).

Tisham Dhar is with CSIRO Space and Astronomy, Adelaide, SA 5000, Australia (e-mail: tisham.dhar@csiro.au).

Mark J. Doubell is with South Australian Research and Development Institute, Aquatic Sciences, West Beach, SA 5024, Australia (e-mail: mark.doubell@sa.gov.au).

physiology of phytoplankton community, the size distributions of suspended particles, and the sources and processing of BGC constituents [18], [19]. Collectively, these properties provide a basis for assessing water quality by optical means.

Traditional water quality monitoring relies on laboratory analysis of grab samples to derive BGC parameters (*e.g.*, [20]). While these methods are accurate and well-established, they are often labour-intensive, costly, and logistically constrained [3], [9]. Remote sensing has emerged as a cost-effective approach to retrieve key BGC parameters remotely [1], and hyperspectral observations are particularly well-suited for capturing the spectral signatures required to assess water quality conditions [2], [7]. Recent advances in data-driven algorithms, particularly those leveraging rigorous machine learning and deep learning frameworks, have shown promise for improving the accuracy and robustness of BGC parameter retrieval from hyperspectral R_{rs} measurements (*e.g.*, [7]–[9]).

Previous studies have identified that algorithms developed and validated within a particular region for hyperspectral BGC retrieval may exhibit limited transferability to other regions with optically distinct waters [21]–[24]. This lack of generalisation stems from the fact that the relationship between hyperspectral R_{rs} and BGC parameters is not universal, but rather depends on regional differences in environmental conditions and biogeochemistry that lead to distinct BGC ranges and SIOP characteristics [7], [18], [25]. Several seminal works have observed such regional distinctions in the relationship between BGC parameters and hyperspectral R_{rs} , and proposed strategies to address it. For example, Mao et al. [26] reported that some regions in the East China Sea exhibit extremely high total suspended matter (TSM), exceeding the valid range of standard retrieval algorithms. To address this, the authors proposed a complex-proxy TSM model that blends four R_{rs} -based indices to transform the non-linear TSM– R_{rs} relationship into a quasi-linear one, enabling robust retrievals across the full concentration range. Ogashawara et al. [27] observed that the Funil and Itumbiara reservoirs in Brazil are dominated by CDOM as the primary optically active constituent, whereas standard quasi-analytical algorithms (QAAs), originally designed for phytoplankton-dominated waters, failed to accurately retrieve the IOPs. To overcome this limitation, the QAA was re-parameterised by Ogashawara et al. [27] to suit these CDOM-dominated conditions. In a study to retrieve chlorophyll-*a* concentrations in 185 inland and coastal water bodies worldwide, Neil et al. [25] demonstrated that improved retrieval accuracies could be achieved when algorithms are optimised for individual optical water types. These regional bio-optical distinctions contribute to the non-unique inverse problem [7], as the same (or nearly identical) spectral signatures of R_{rs} may correspond to different values of BGC parameters across water bodies. Therefore, robust algorithms for retrieving BGC parameters require explicit accommodation of such regional variations in bio-optical conditions.

In addition to handling regional bio-optical distinctions, data-driven water quality retrieval from hyperspectral R_{rs} measurements faces further challenges. The first challenge is that data-driven algorithms often require more data for

training than empirical and semi-analytical models, while the in situ sampling process is labour and time-consuming. Several strategies have been implemented to handle this challenge, including: (1) self-supervised pretraining on unlabelled hyperspectral R_{rs} spectra is used to learn general representations, thereby reducing the number of labelled BGC– R_{rs} pairs required for subsequent model fine-tuning [9]; and (2) aggregating samples from multiple sources into large, curated datasets, *e.g.* GLORIA [2], to facilitate the development of generalisable and adaptive models [7]. The second challenge is that existing data-driven algorithms for hyperspectral retrieval of BGC parameters often rely on observational data only for model training without integrating knowledge of the underlying physical principles of underwater optics. This may limit the generalisability and interpretability of the developed models.

To alleviate these challenges, we propose a two-stage training framework in this study: (1) a physics-aware pretraining stage, and (2) a region-specific adaptation stage. In the pretraining stage, a physics-based bio-optical model is used to synthesise a BGC– R_{rs} dataset from a bio-optical library of in situ measured BGC parameters and IOPs/SIOPs with broad representativeness of Australian coastal waters. This physics-guided synthetic dataset is applied to train a region-agnostic base model within a physics-aware meta-learning framework. In the adaptation stage, the pretrained base model is subsequently adapted to individual regions using in situ measured BGC– R_{rs} samples specific to each region. The base model can regulate regional adaptation by leveraging physical knowledge learned from pretraining. The proposed algorithm is designed to incorporate the following features that distinguish it from existing data-driven approaches to BGC retrieval from hyperspectral R_{rs} observations:

- 1) Integration of physical principles within a data-driven deep learning framework, enabling more robust modelling of the BGC– R_{rs} relationship and ensuring that the retrieved BGC parameters remain physically plausible;
- 2) Adaptability to account for region-specific bio-optical distinctions in the BGC– R_{rs} relationship, thereby improving retrieval accuracies within each region.

Supported by CSIRO’s AquaWatch Australia initiative¹, we established four experimental sites in Australian coastal waters, namely the Fitzroy Estuary and Keppel Bay sites in Queensland, the Boston Bay site in South Australia, and the Cockburn Sound site in Western Australia. Time-series measurements of in situ BGC parameters and hyperspectral R_{rs} were collected at these sites across multiple seasons from 2022 to 2025. In situ measurements were also sourced from a long-term monitoring site, namely the Lucinda Jetty Coastal Observatory in Queensland, from 2019 to 2023. These five sites cover several climate zones, including tropical, subtropical, and temperate, and represent a range of hydrodynamic regions, including riverine, estuarine, and coastal. As a result, the waters at these sites exhibit distinct bio-optical characteristics.

¹<https://www.csiro.au/en/about/challenges-missions/AquaWatch>

TABLE I
SYMBOLS AND ACRONYMS USED IN THIS STUDY.

Symbol	Unit	Definition
<i>Biogeochemical (BGC) Parameters</i>		
TSS	mg/L	Total suspended solids concentration
TChl-a	$\mu\text{g/L}$	Total chlorophyll-a concentration
NAP	mg/L	Non-algal particles concentration
DOC	mg/L	Dissolved organic carbon concentration
<i>Inherent Optical Properties (IOPs)</i>		
a	m^{-1}	Total absorption coefficient
a_w	m^{-1}	Absorption coefficient of pure water
a_d	m^{-1}	Absorption coefficient of NAP
a_{ph}	m^{-1}	Absorption coefficient of phytoplankton
a_y	m^{-1}	Absorption coefficient of Chromophoric dissolved organic matter (CDOM)
S_y	nm^{-1}	Spectral slope of CDOM absorption coefficient
b_b	m^{-1}	Total backscattering coefficient
b_{bw}	m^{-1}	Backscattering coefficient of pure water
b_{bp}	m^{-1}	Backscattering coefficient of particulate matter
S_{bbp}	–	Spectral slope of particulate backscattering coefficient
<i>Specific IOPs (SIOPs)</i>		
a_d^*	$\text{m}^2 \text{g}^{-1}$	TSS-normalised NAP absorption coefficient
a_{ph}^*	$\text{m}^2 \text{mg}^{-1}$	TChl-a-normalised phytoplankton absorption coefficient
a_y^*	$\text{m}^2 \text{g}^{-1}$	DOC-normalised CDOM absorption coefficient
b_{bp}^*	$\text{m}^2 \text{g}^{-1}$	TSS-normalised particulate backscattering coefficient
<i>Apparent Optical Properties (AOPs)</i>		
R_{rs}	sr^{-1}	Above-water remote sensing reflectance
r_{rs}	sr^{-1}	Subsurface remote sensing reflectance
<i>Ancillary Variables</i>		
T	$^{\circ}\text{C}$	Water temperature
S	–	Water salinity

The proposed model for retrieving BGC parameters was evaluated using data collected from the aforementioned sites. Through quantitative analysis of the evaluation results, we aim to address primarily the following questions:

- 1) Is there a clear regional distinction among the experimental sites in terms of BGC parameters and their hyperspectral R_{rs} signatures?
- 2) How does the proposed approach perform in hyperspectral retrieval of BGC parameters compared to benchmark algorithms?

The rest of this article is organised as follows. Section II describes the study sites, in situ hyperspectral R_{rs} and BGC measurements, and the bio-optical spectral library used in this study. Section III details the proposed physics-aware meta-learning framework, including the bio-optical forward model, synthetic data generation, model pretraining, and region-specific adaptation. Section IV presents the experimental results and discusses the retrieval performance, regional variability, and time-series monitoring capabilities of the proposed approach. The key findings, limitations, and future directions are also discussed in Section IV. Finally, Section V concludes this article with a summary overview of this study. To facilitate readability, Table I lists the commonly used symbols and acronyms in this article.

TABLE II
LIST OF EXPERIMENTAL SITES.

Site name	State	Nominal geo-coordinates	Climate zone	Mean salinity
Fitzroy Estuary	Queensland	23°30'07"S, 150°48'03"E	Subtropical	27.1
Keppel Bay	Queensland	23°27'54"S, 150°56'39"E	Subtropical	36.1
Boston Bay	South Australia	34°43'03"S, 135°54'02"E	Temperate	36.3
Cockburn Sound	Western Australia	32°15'03"S, 115°43'42"E	Temperate	35.5
Lucinda Jetty	Queensland	18°31'11"S, 146°23'10"E	Tropical	34.8

II. STUDY SITES AND DATASETS

A. Study sites

This study focused on five sites located in Australian coastal waters, as shown in Fig. 1. The Fitzroy Estuary and Keppel Bay sites are located near the mouth of Fitzroy River, Queensland, with the former situated inside the estuary and the latter downstream in open coastal waters in Keppel Bay. The Boston Bay site is located in the strait between Boston Island and the mainland town of Port Lincoln, within Boston Bay in Spencer Gulf, South Australia. The Cockburn Sound site is located within Cockburn Sound, Western Australia, between Garden Island and the Perth metropolitan coast. The last site, Lucinda Jetty, is located in the coastal waters of the Great Barrier Reef World Heritage Area, close to the Herbert River Estuary and the Hinchinbrook Channel in Queensland. These geographically distinct sites span tropical, subtropical, and are influenced to varying degrees by terrestrial freshwater discharge as reflected by their salinity levels (Table II).

B. Near-surface hyperspectral R_{rs} observations

A HydraSpectra instrument (Fig. 2a) was deployed to measure hyperspectral water-leaving R_{rs} at each of the Fitzroy Estuary, Keppel Bay, Boston Bay, and Cockburn Sound sites. These instruments were calibrated in the laboratory prior to deployment. It recorded time-series hyperspectral measurements at 10-minute intervals, including downwelling solar irradiance ($E_d(\lambda)$), upwelling water-leaving radiance ($L_w(\lambda)$), and diffuse skylight radiance ($L_{sky}(\lambda)$). These measurements were used to calculate the water-leaving R_{rs} following the method described in [28]. To ensure a high signal-to-noise ratio for the HydraSpectra sensor, we selected R_{rs} observations acquired around local noontime (10:00am – 2:00pm), when solar irradiance is usually stronger than at other times of the day, for our analysis. Each R_{rs} spectrum spanned over the spectral range of 400–700 nm, with a spectral resolution of approximately 3 nm and a sampling interval of 1 nm. Data outside this range were excluded from our analysis due to low signal-to-noise ratios. These HydraSpectra instruments were regularly serviced and calibrated by the CSIRO servicing team. At the Lucinda Jetty site, water-leaving hyperspectral R_{rs} was derived from measurements acquired using a Hyperspectral Ocean Colour Radiometer (HyperOCR). These data are available through the Integrated Marine Observing System

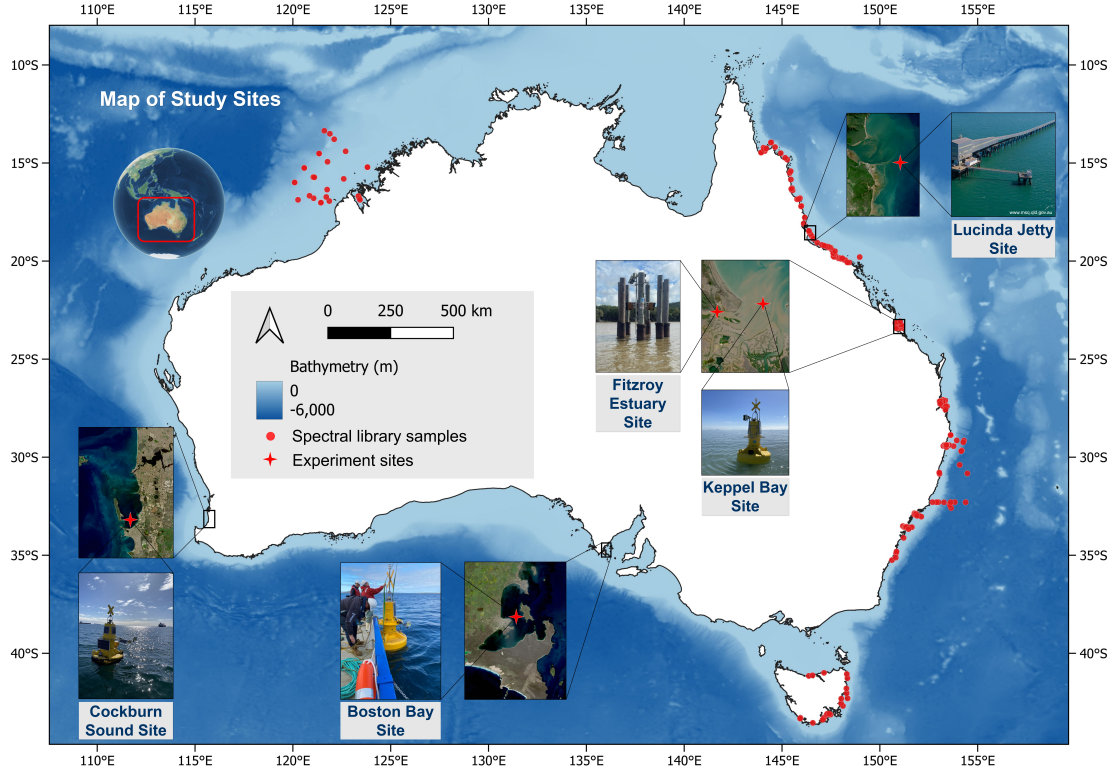


Fig. 1. This study was conducted across five experimental sites in Australian coastal waters: Fitzroy Estuary, Keppel Bay, Boston Bay, Cockburn Sound, and Lucinda Jetty. The map also shows the locations (red dot points) where the bio-optical spectral library samples were collected. Insets show satellite images of the experimental sites and adjacent areas, along with photos of the instruments taken during or after deployment.

(IMOS) via the Australian Ocean Data Network (AODN) portal². Similar to the HydraSpectra data, the spectral range from 400 to 700 nm in the HyperOCR measurements was used for analysis.

C. In situ BGC parameter measurements

At the experimental sites of Fitzroy Estuary, Keppel Bay, Boston Bay, and Cockburn Sound, a YSI EXO2 multiparameter sonde and an OPUS UV Spectral Sensor (Fig. 2b) were installed at a nominal depth of 1 m below the water surface, at the same location as the above-water HydraSpectra deployment. Both sensors were configured to record BGC measurements at 10 min intervals. For the Lucinda Jetty site, grab samples of water were collected at the same location as the above-water HyperOCR deployment, and are available through the IMOS AODN portal³. The number of in situ BGC samples and the corresponding sampling period at each experimental site are listed in Table III.

D. Bio-optical spectral library

We compiled a bio-optical spectral library using 247 in situ bio-optical samples collected from several regions along Australia's coastline, including coastal waters in Queensland,

TABLE III
SUMMARY OF IN SITU BIOGEOCHEMICAL SAMPLES COLLECTED AT EACH EXPERIMENTAL SITE.

Site	No. of Samples	Sampling Period
Fitzroy Estuary	296	26 Apr 2023 – 23 Sep 2024
Keppel Bay	377	1 Jun 2023 – 28 Nov 2024
Boston Bay	774	1 Sep 2022 – 2 Feb 2025
Cockburn Sound	172	15 Jul 2023 – 18 Feb 2024
Lucinda Jetty	104	12 Nov 2019 – 1 Nov 2023

New South Wales, Tasmania, and Western Australia, as shown in Fig. 1. These in situ bio-optical measurements cover a wide range of optical water types and hydro-optical conditions with broad representativeness of Australian coastal waters. The library is built upon multiple field campaigns conducted between 2002 and 2022, including: (1) seven voyages in the inshore estuarine, lagoonal, and reef waters of the Great Barrier Reef during four dry and one wet tropical seasons between October 2002 and September 2005 [29], [30]; (2) a voyage in Tasmanian coastal waters from 21 to 30 May 2007 [31]; (3) a voyage in the western Tasman Sea, off the New South Wales coast, conducted from 15 to 30 October 2010 [32]; (4) a voyage across the Kimberley shelf off northwest Western Australia from 14 April to 5 May 2010 [33]; (5) five voyages in Moreton Bay, Queensland, in 2011 following a major flood event [34]; (6) a voyage in Princess Charlotte Bay, Queensland, conducted from 30 January to 1 February 2013 after the passage of Tropical Cyclone Oswald [35]; and

²<https://thredds.aodn.org.au/thredds/catalog/IMOS/SRS/OC/LJCO/HyperOCR-hourly/catalog.html>

³<https://thredds.aodn.org.au/thredds/catalog/IMOS/SRS/OC/BODBAW/catalog.html>

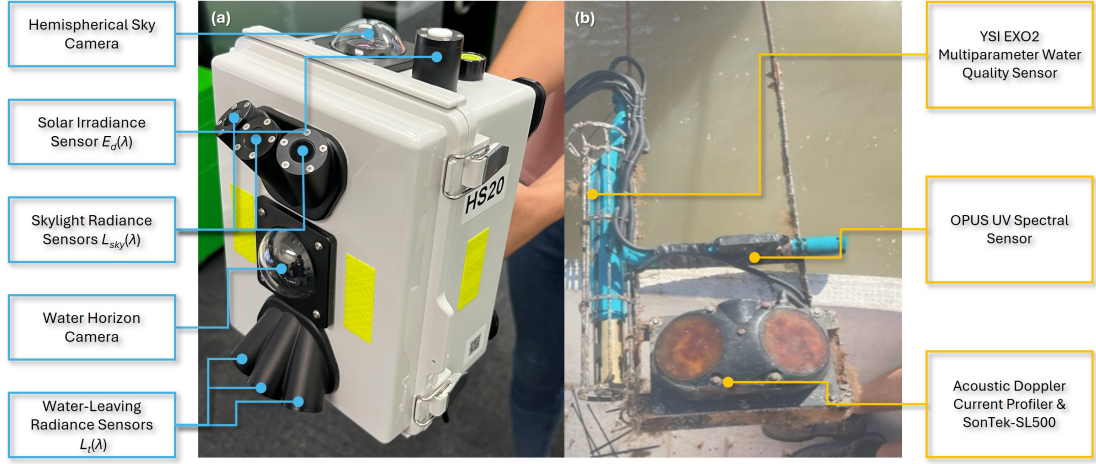


Fig. 2. In situ sensors deployed at the experimental sites of Fitzroy Estuary, Keppel Bay, Boston Bay, and Cockburn Sound. (a) HydraSpectra instrument for recording water-leaving hyperspectral R_{rs} . (b) In-water sensors for collecting biogeochemical measurements and monitoring water conditions.

(7) two voyages across three estuaries along the New South Wales coast in March and July 2022, respectively [3].

For each water sample in the bio-optical library, the following BGC parameters, IOPs, and ancillary physical variables were measured: (1) water temperature (in unit of $^{\circ}\text{C}$), (2) water salinity (unitless), (3) TSS (in unit of mg/L), (4) TChl-a (in unit of $\mu\text{g/L}$), (5) DOC (in unit of mg/L), (6) absorption coefficient of CDOM at 440 nm ($a_y(\lambda_{440})$, in unit of m^{-1}) and its spectral slope ($S_y(\lambda_{440})$, in unit of nm^{-1}), (7) backscattering coefficient of particulate matter at 550 nm ($b_{bp}(\lambda_{550})$, in unit of m^{-1}) and its spectral slope ($S_{b_{bp}}(\lambda_{550})$, unitless), (8) absorption coefficient of particulate matter within 400–700 nm ($a_p(\lambda)$, in unit of m^{-1}), (9) absorption coefficient of NAP within 400–700 nm ($a_d(\lambda)$, in unit of m^{-1}), and (10) absorption coefficient of phytoplankton within 400–700 nm ($a_{ph}(\lambda)$, in unit of m^{-1}). The SIOPs were then derived, including: (1) mass-specific absorption coefficient of CDOM ($a_y^*(\lambda)$, in unit of $\text{m}^2 \text{g}^{-1}$), (2) mass-specific backscattering coefficient of particulate matter ($b_{bp}^*(\lambda)$, in unit of $\text{m}^2 \text{g}^{-1}$), (3) mass-specific absorption coefficient of NAP ($a_d^*(\lambda)$, in unit of $\text{m}^2 \text{g}^{-1}$), and (4) mass-specific absorption coefficient of phytoplankton ($a_{ph}^*(\lambda)$, in unit of $\text{m}^2 \text{mg}^{-1}$).

III. METHODS

A. Overview

We aim to develop a region-adaptable approach for retrieving BGC parameters from in situ hyperspectral R_{rs} measurements. Given that the relationship between R_{rs} and BGC parameters may vary from one region to another due to distinct regional bio-optical characteristics, we propose a two-stage framework for model training, as shown in Fig. 3. In the first stage, a physics-based bio-optical forward model was adopted to simulate a large physics-guided synthetic dataset linking BGC parameters and their SIOPs with hyperspectral R_{rs} . The simulation was conducted based on the bio-optical spectral library. Then, a region-agnostic base model was trained with the synthetic dataset using meta-learning. In the second stage, the base model was fine-tuned for each target region using in

situ BGC- R_{rs} samples, resulting in a region-specific model for each region. This two-stage framework was designed to integrate physical modelling with data-driven learning, with the potential to accommodate regional bio-optical distinctions. The proposed framework is described in detail in the subsequent Subsections.

B. Bio-optical forward model

We use a physics-based bio-optical model to link BGC parameters and their SIOPs with hyperspectral R_{rs} . Specifically, the total spectral absorption coefficient of the water, a , is expressed as:

$$a(\lambda, T, S) = a_w(\lambda, T, S) + a_d(\lambda) + a_y(\lambda) + a_{ph}(\lambda), \quad (1)$$

where a_w , a_d , a_y , and a_{ph} represent absorption by water, NAP, CDOM, and phytoplankton, respectively.

To parameterise a_d , a_y , and a_{ph} using BGC measurements, we derived empirical SIOPs by normalising these component absorption coefficients by the corresponding TSS, DOC, and TChl-a concentrations:

$$a(\lambda, T, S) = a_w(\lambda, T, S) + \text{TSS} \times a_d^*(\lambda) + \text{DOC} \times a_y^*(\lambda) + \text{TChl-a} \times a_{ph}^*(\lambda). \quad (2)$$

Here, a_d^* , a_y^* , and a_{ph}^* are the TSS-normalised a_d , DOC-normalised a_y , and TChl-a-normalised a_{ph} , respectively.

Given that most of our experimental sites are located in coastal waters influenced by freshwater-seawater mixing, the spectral curve of a_w is computed accounting for variations in water salinity (S) and temperature (T) to enhance accuracy:

$$a_w(\lambda, T, S) = a_w^{\text{ref}}(\lambda) + (T - T_w^{\text{ref}}) \times \Psi_T(\lambda) + S \times \Psi_S(\lambda), \quad (3)$$

where a_w^{ref} is the reference water absorption spectrum, adapted from [36]; $T_w^{\text{ref}} = 22^{\circ}\text{C}$ is the reference temperature; Ψ_T and Ψ_S are the correction coefficients for temperature and salinity, respectively, adapted from [37].

The total spectral backscattering coefficient, b_b , is expressed as:

$$b_b(\lambda, S) = b_{bw}(\lambda, S) + \text{TSS} \times b_{bp}^*(\lambda), \quad (4)$$

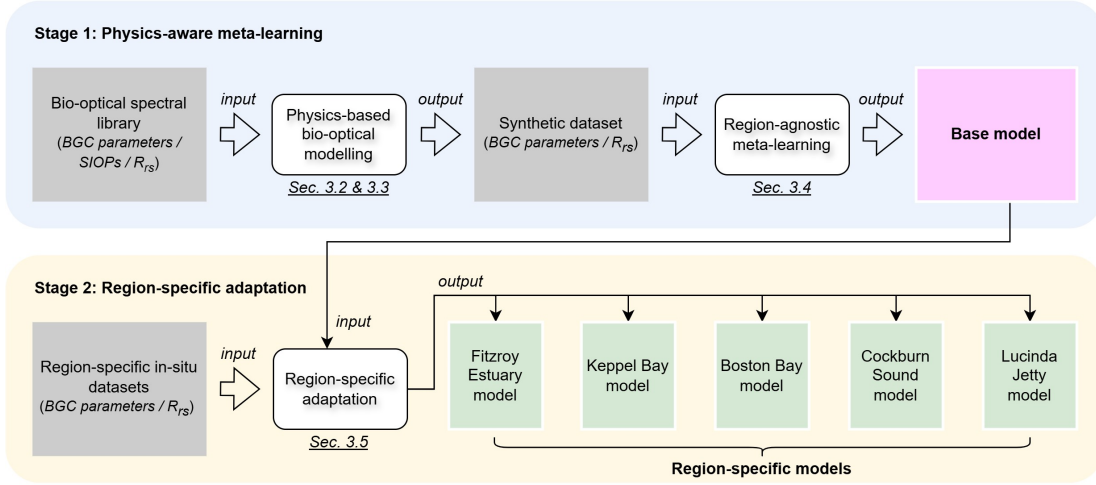


Fig. 3. Flowchart of the proposed physics-aware meta-learning approach for hyperspectral retrieval of biogeochemical parameters. It consists of two major stages: (1) a physics-aware pretraining stage (detailed in Subsections III-B, III-C, and III-D), and (2) a region-specific adaptation stage (detailed in Subsection III-E). Please refer to the nomenclature in Table I for definitions of symbols and acronyms.

where b_{bp}^* is the TSS-specific particulate backscattering coefficient. Following the molecular scattering formulation of [38], the water volume-scattering function is calculated as:

$$A(\lambda, S) = 1.38 \times 10^{-4} \times \left(\frac{\lambda}{500} \right)^{-4.32} \times \left(1 + 0.3 \times \frac{S}{37} \right). \quad (5)$$

The backscattering coefficient of water, b_{bw} , is then obtained by integrating the volume-scattering function over the backward hemisphere:

$$b_{bw}(\lambda, S) = 2\pi \times A(\lambda, S) \times \left[1 + \frac{1 - \delta}{3(1 + \delta)} \right], \quad (6)$$

where $\delta = 0.09$ is the molecular depolarisation ratio.

Following a standard radiative-transfer approximation [17], the subsurface R_{rs} at 0^- , r_{rs} , is related to the total backscattering albedo, u , by:

$$r_{rs}(\lambda) = g_0 \times u(\lambda) + g_1 \times u^2(\lambda), \quad (7)$$

where $g_0 = 0.082$ and $g_1 = 0.17$ are empirical coefficients derived from the HydroLight radiative transfer model [39], and u is expressed as:

$$u(\lambda) = \frac{b_b(\lambda)}{a(\lambda) + b_b(\lambda)}. \quad (8)$$

To obtain the above-water R_{rs} at 0^+ , R_{rs} , we apply a standard air–water interface correction:

$$R_{rs}(\lambda) = \frac{0.52 \times r_{rs}(\lambda)}{1 - 1.7 \times r_{rs}(\lambda)}. \quad (9)$$

where the correction coefficients 0.52 and 1.7 follow the values given in [17].

C. Synthetic data generation

We generate a physics-based synthetic dataset with bio-optical modelling. For a given set of input parameters (TSS, DOC, TChl-a, a_d^* , a_y^* , a_{ph}^* , b_{bp}^* , T , and S), the bio-optical

forward model, as described in Subsection III-B, is able to simulate the corresponding hyperspectral R_{rs} . Randomly selected sets of input values may result in physically unrealistic simulations. To enable physically consistent and plausible simulations, the values for simulation inputs are sampled with the guidance of the bio-optical spectral library. Specifically, we first characterise the statistical distribution of the spectral library measurements and then draw simulation inputs from this distribution, aiming to align the sampled values with the observed measurements. These sampled simulation inputs are subsequently passed to the bio-optical forward model to generate hyperspectral R_{rs} . This procedure is detailed in the following paragraphs.

We first apply two preprocessing steps to the spectral library measurements before modelling their statistical distribution. The first step is to transform the values of TSS, DOC, TChl-a, a_d^* , a_y^* , a_{ph}^* , and b_{bp}^* into the $\log_{10}$ scale. Because the spectral library measurements of these BGC parameters and SIOPs are skewed towards small values, $\log_{10}$ -scaling helps normalise their distributions and stabilise subsequent modelling, as recommended by [40]. We denote the $\log_{10}$ -scaled TSS, DOC, and TChl-a as $\mathbf{x}_{TSS}, \mathbf{x}_{DOC}, \mathbf{x}_{TChl-a} \in \mathbb{R}^N$, where $N = 247$ is the number of spectral library measurements. Similarly, the $\log_{10}$ -scaled a_d^* , a_y^* , a_{ph}^* , and b_{bp}^* are denoted as $\mathbf{A}_d, \mathbf{A}_y, \mathbf{A}_{ph}, \mathbf{B}_{bp} \in \mathbb{R}^{M \times N}$, with $M = 301$ spectral bands covering 400–700 nm with a 1 nm interval. The T and S measurements are not strongly skewed and therefore left unscaled, and are denoted as $\mathbf{x}_T, \mathbf{x}_S \in \mathbb{R}^N$.

The second preprocessing step is to handle the spectral dependency of a_d^* , a_y^* , a_{ph}^* , and b_{bp}^* . Unlike other input variables to the bio-optical forward model, these SIOPs are wavelength-dependent. For each of these variables, the values across adjacent wavelengths are highly correlated, forming smooth and continuous spectral shapes in nature. To preserve these correlations across wavelengths, we transform a_d^* , a_y^* , a_{ph}^* , and b_{bp}^* from the original spectral space into a principal component (PC) space using principal component analysis

(PCA). The transformed data are represented in the PC space as $\mathbf{X}_d$, $\mathbf{X}_y$, $\mathbf{X}_{ph}$, and $\mathbf{X}_{bbp}$. This PCA transformation captures the dominant modes of spectral variability while preserving the correlations across wavelengths, as the spectral shapes are represented by a compact set of uncorrelated PCs.

We then model the joint statistical distribution of TSS, DOC, TChl-a, a_d^* , a_y^* , a_{ph}^* , b_{bp}^* , T , and S for the spectral library samples based on their preprocessed values. We adopt an infinite Gaussian mixture model called the Dirichlet Process Bayesian Gaussian Mixture Model (DP-GMM), which is a non-parametric Bayesian framework suitable for modelling non-Gaussian multivariate distributions. Specifically, we first collate the preprocessed input variables as $\mathbf{Y} \in \mathbb{R}^{(5+4 \times P) \times N}$:

$$\mathbf{Y} = [\mathbf{x}_{\text{TSS}}, \mathbf{x}_{\text{DOC}}, \mathbf{x}_{\text{TChl-a}}, \mathbf{X}_d, \mathbf{X}_y, \mathbf{X}_{ph}, \mathbf{X}_{bbp}, \mathbf{x}_T, \mathbf{x}_S]^\top. \quad (10)$$

Let $\mathbf{y}_n \in \mathbb{R}^{(5+4 \times P)}$ denote the n -th measurement in $\mathbf{Y}$, and let $\mathbf{y} \in \mathbb{R}^{(5+4 \times P)}$ denote a generic measurement. We then use DP-GMM to model the distribution of $\mathbf{y}$, $p(\mathbf{y})$, as an unbounded mixture of Gaussian components:

$$p(\mathbf{y}) = \sum_{i=1}^{\infty} \pi_i \mathcal{N}(\mathbf{y} \mid \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i). \quad (11)$$

where $\mathcal{N}(\cdot)$ denotes the multivariate Gaussian distribution; π_i , $\boldsymbol{\mu}_i$, and $\boldsymbol{\Sigma}_i$ are learnable parameters denoting the mixture weight, mean vector, and covariance matrix of the i -th Gaussian component, respectively.

The joint distribution $p(\mathbf{y})$ given in Eq. (11) is derived from a large number of in situ measurements in the spectral library, and thus serves as a realistic representation of the distribution and inter-correlations of BGC parameters and their SIOPs. We draw a total of K samples from $p(\mathbf{y})$ to obtain physically plausible inputs for bio-optical modelling:

$$\{\mathbf{y}^{(k)}\}_{k=1}^K \sim p(\mathbf{y}), \quad (12)$$

where $\mathbf{y}^{(k)}$ is the k -th sample drawn from the distribution, consisting of a set of parameters, including the $\log_{10}$ -scaled TSS, DOC, and TChl-a, the $\log_{10}$ -scaled and PCA-transformed a_d^* , a_y^* , a_{ph}^* , b_{bp}^* , and T and S . These parameters are then scaled and transformed back to their original scales and spaces, and denoted as $\text{TSS}^{(k)}$, $\text{DOC}^{(k)}$, $\text{TChl-a}^{(k)}$, $a_d^{*(k)}$, $a_y^{*(k)}$, $a_{ph}^{*(k)}$, $b_{bp}^{*(k)}$, $T^{(k)}$, and $S^{(k)}$. For each drawn sample, we input them into the bio-optical forward model and calculate the corresponding R_{rs} :

$$R_{rs}^{(k)} = f(\text{TSS}^{(k)}, \text{DOC}^{(k)}, \text{TChl-a}^{(k)}, a_d^{*(k)}, a_y^{*(k)}, a_{ph}^{*(k)}, b_{bp}^{*(k)}, T^{(k)}, S^{(k)}), \quad (13)$$

where $f(\cdot)$ is the bio-optical forward model as described in Subsection III-B. To adequately cover the distribution of the bio-optical variables, we sampled $K = 10,000$ instances. Based on empirical test runs, this sample size was found to provide sufficient coverage of the parameter space, with larger sample sizes yielding only marginal gains for pretraining the base model. For each sample, the corresponding R_{rs} was

calculated, resulting in a synthetic dataset $\mathcal{D}$ that links BGC parameters with R_{rs} :

$$\mathcal{D} = \left\{ \left(\text{TSS}^{(k)}, \text{DOC}^{(k)}, \text{TChl-a}^{(k)}, a_d^{*(k)}, a_y^{*(k)}, a_{ph}^{*(k)}, b_{bp}^{*(k)}, T^{(k)}, S^{(k)}, R_{rs}^{(k)} \right) \right\}_{k=1}^K. \quad (14)$$

D. Pretraining of the base model

We pretrain a region-agnostic base model with the synthetic dataset $\mathcal{D}$ generated in Subsection III-C. We formulate the pretraining of the base model as a meta-learning problem and develop an approach based on the Model-Agnostic Meta-Learning (MAML) architecture [41]. The base model is pre-trained on a collection of tasks, where each task is formed from a subset of synthetic samples in $\mathcal{D}$. As $\mathcal{D}$ is constructed with physics-based bio-optical modelling (Subsection III-C), the base model is able to learn the general physical relationship linking BGC parameters and hyperspectral R_{rs} through this pretraining process. Moreover, tasks are constructed with diverse SIOP settings, enabling the base model to adapt effectively to new regions characterised by distinct SIOP conditions. The pretraining of the base model is detailed in the following.

We denote the base model as $g_\theta(\cdot) : \mathbb{R}^M \rightarrow \mathbb{R}^3$ with model parameters θ , mapping a hyperspectral R_{rs} spectrum $R_{rs} \in \mathbb{R}^M$ to the BGC parameters TSS, DOC, and TChl-a. In our study, the number of spectral bands is $M = 301$, covering 400–700 nm with a 1 nm interval. We pretrain $g(\cdot)$ on a distribution of tasks $p(\mathcal{T})$ sampled from $\mathcal{D}$. Each task τ_i is constructed by sampling a subset $\mathcal{D}_i \subset \mathcal{D}$ of $2K_i$ samples and splitting it evenly into a support (training) set $\mathcal{S}_i$ and a query (test) set $\mathcal{Q}_i$, with each having K_i samples. The experimental sites considered in this study may present distinct SIOPs, due to differences in regional biogeochemistry, such as phytoplankton community composition and physiology, suspended-particle size distributions, and CDOM sources and processing. To equip $g(\cdot)$ with the ability to adapt to different SIOP scenarios, each task τ_i is sampled in a way that its synthetic samples from $\mathcal{D}$ share similar SIOPs. Specifically, for each task, we pick a random sample from $\mathcal{D}$, compute its Euclidean distances to all other samples in $\mathcal{D}$ in terms of SIOPs, and select the closest $2K_i$ ones. In this way, different tasks correspond to distinct SIOP scenarios. The adaptation ability of the base model, $g_\theta(\cdot)$, is learned by adapting its parameters to each of the sampled tasks. Here, we distinguish the parameters of the base model, θ , which serve as meta-parameters shared across all tasks, from the task-specific parameters, ϕ_i , which are obtained by adapting θ to the i -th task τ_i . In other words, θ represents the general knowledge shared across tasks, while ϕ_i captures the additional task-specific knowledge learned from the particular task τ_i . For the i -th task τ_i , ϕ_i are obtained by adapting θ to the support set $\mathcal{S}_i$, via minimising the following loss function $\mathcal{L}_{\mathcal{S}_i}$ that measures the discrepancy between the model-predicted and

the true values of TSS, DOC, and TChl-a:

$$\mathcal{L}_{\mathcal{S}_i}(\phi_i) = \frac{1}{K_i} \sum_{(R_{rs}, [\text{TSS}, \text{DOC}, \text{TChl-a}]) \in \mathcal{S}_i} \|g_{\phi_i}(R_{rs}) - [\text{TSS}, \text{DOC}, \text{TChl-a}]\|_2^2. \quad (15)$$

The accuracy of this adaptation is then measured on the query set with the same type of loss function $\mathcal{L}_{\mathcal{Q}_i}$. Then, the parameters of the base model, θ , are optimised by minimising the meta loss, $\mathcal{J}(\theta)$, which is expressed as the average query loss across all tasks:

$$\mathcal{J}(\theta) = \frac{1}{B} \sum_{i=1}^B \mathcal{L}_{\mathcal{Q}_i}(\phi_i(\theta)), \quad (16)$$

where B is the total number of tasks, and $\phi_i(\theta)$ stands for the adapted parameters ϕ_i to the i -th task τ_i under the meta-parameters θ .

E. Region-specific model adaptation

We adapt the pretrained base model, $g_{\theta}(\cdot)$, to each experimental site using local in situ samples. Let $D_r = \{(R_{rs}^{(i)}, \mathbf{y}^{(i)})\}_{i=1}^{N_r}$ denote the region-specific dataset collected from the experimental site r , where $R_{rs}^{(i)} \in \mathbb{R}^M$ represents the hyperspectral R_{rs} spectrum and $\mathbf{y}^{(i)} = [\text{TSS}, \text{DOC}, \text{TChl-a}] \in \mathbb{R}^3$ represents the corresponding BGC parameters measured in situ, with N_r denoting the number of samples available for that site. Following the meta-learning paradigm introduced in Subsection III-D, we treat each site as a new task and initialise the model parameters using the pretrained meta-parameters θ . The region-specific parameters ϕ_r are then obtained through iterative gradient updates using the regional dataset. Specifically, for a given set of parameters ϕ_r , the adaptation loss is defined as:

$$\mathcal{L}_r(\phi_r) = \frac{1}{N_r} \sum_{i=1}^{N_r} \|g_{\phi_r}(R_{rs}^{(i)}) - \mathbf{y}^{(i)}\|_2^2. \quad (17)$$

This procedure iteratively adjusts the model parameters so that the predictions align with the local BGC- R_{rs} relationships observed in the target experimental site.

F. Accuracy assessment

In this study, a 10-fold cross-validation scheme was adopted for accuracy assessment. To mitigate temporal autocorrelation, a block-based training/test strategy was employed in which the dataset from each experimental site was partitioned into 10 temporally separate blocks.

The proposed approach was benchmarked against five existing models, including three classic empirical models and two recent deep-learning-based models. The empirical models are the TSS model by [42], the DOC model by [43], and the TChl-a model by [33]. The deep-learning-based models include DLRS [3], which was used for benchmarking TSS and DOC, and HyperEST [9], which was used for benchmarking TSS and TChl-a. These benchmark models were recalibrated/fine-tuned for each experimental site using our dataset. Consistent

TABLE IV
STATISTICS OF THE BIOGEOCHEMICAL PARAMETERS, SPECIFIC INHERENT OPTICAL PROPERTIES, AND ANCILLARY PHYSICAL VARIABLES IN THE BIO-OPTICAL SPECTRAL LIBRARY.

Parameter	Min	Med	Max	Mean $\pm$ Std Dev
T ($^{\circ}\text{C}$)	12.128	25.190	31.404	24.301 ± 5.137
S (-)	0.080	33.420	39.400	28.388 ± 9.851
TSS (mg/L)	0.133	2.600	69.830	6.892 ± 10.637
DOC (mg/L)	0.240	1.300	14.250	2.177 ± 2.200
TChl-a ($\mu\text{g/L}$)	0.059	0.963	22.037	2.204 ± 3.330
$a_y^*(\lambda_{440})$ ($\text{m}^2 \text{g}^{-1}$)	0.004	0.230	2.552	0.307 ± 0.288
$b_{bp}^*(\lambda_{550})$ ($\text{m}^2 \text{g}^{-1}$)	0.001	0.008	0.042	0.008 ± 0.006
$a_d^*(\lambda_{440})$ ($\text{m}^2 \text{g}^{-1}$)	0.002	0.031	0.257	0.037 ± 0.031
$a_{ph}^*(\lambda_{440})$ ($\text{m}^2 \text{mg}^{-1}$)	0.010	0.059	0.400	0.076 ± 0.060

with the proposed method, a temporal block-based 10-fold cross-validation scheme was employed to evaluate model performance.

Log-scaled metrics have been recommended for accuracy assessment in water quality studies mainly because BGC measurements are often skewed towards small values. Following the practice by [3], [9], [40], we adopted logarithmic coefficient of determination ($\text{Log-}R^2$) and logarithmic mean absolute error (Log-MAE) as assessment metrics:

$$\text{Log-}R^2 = 1 - \frac{\sum_{i=1}^n [\log_{10}(y_{\text{true},i}) - \log_{10}(y_{\text{pred},i})]^2}{\sum_{i=1}^n [\log_{10}(y_{\text{true},i}) - \overline{\log_{10}(y_{\text{true}})}]^2}, \quad (18)$$

$$\text{Log-MAE} = 10^{\frac{1}{n} \sum_{i=1}^n |\log_{10}(y_{\text{pred},i}) - \log_{10}(y_{\text{true},i})|} - 1. \quad (19)$$

IV. RESULTS AND DISCUSSION

A. Overview

The following Subsections are organised as follows. Subsections IV-B to IV-D show the analysis results of the spectral library, forward model, and generated dataset. Subsection IV-E analyses the regional differences in BGC parameters and hyperspectral R_{rs} among the experimental sites. Subsection IV-F shows the BGC retrieval accuracies obtained with the proposed approach, in comparison with the benchmark models. Finally, Subsection IV-G analyses the time-series variations in R_{rs} and BGC parameters and compares the predicted values with in situ measurements.

B. Analysis of the bio-optical spectral library

The statistics of the spectral library samples are summarised in Table IV. As shown in the table, the spectral library covers a broad dynamic range for each parameter. In particular, the samples extend from near-freshwater conditions (0.08) to fully marine salinities (39.4), and from relatively cool (12.1 $^{\circ}\text{C}$) to warm waters (31.4 $^{\circ}\text{C}$). This indicates that the spectral library encompasses a wide range of environmental conditions representative of both river-affected estuarine and open-ocean waters. The broad representativeness of the spectral library samples provides a basis to guide the pretraining of the base model, allowing it to learn generalised BGC- R_{rs} relationships across diverse water types and environmental conditions.

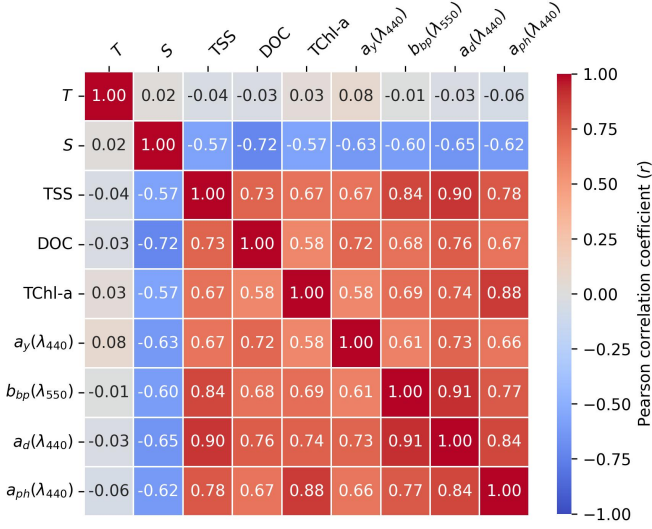


Fig. 4. Pearson correlation matrix among BGC parameters (TSS, DOC, and TChl-a), IOPs ($a_y(\lambda_{440})$, $b_{bp}(\lambda_{550})$, $a_d(\lambda_{440})$, and $a_{ph}(\lambda_{440})$) and ancillary physical variables (T and S) for samples in the bio-optical spectral library.

Fig. 4 presents the Pearson correlation matrix for the spectral library measurements, where the coefficients of correlation (r) were computed for pairwise combinations of T , S , TSS, DOC, TChl-a, $a_y(\lambda_{440})$, $b_{bp}(\lambda_{550})$, $a_d(\lambda_{440})$, and $a_{ph}(\lambda_{440})$. It was found that water temperature T exhibits weak correlations with BGC parameters and IOPs with $|r| \leq 0.08$, which is expected since water constituents are primarily governed by hydrological inputs and transport processes rather than by temperature itself. In contrast, water salinity S is generally negatively correlated with BGC parameters and IOPs with r ranging between -0.72 to -0.57 , as freshwater inflows reduce seawater salinity and concurrently introduce elevated loads of particulates and dissolved matter from terrestrial sources. TSS exhibits strong correlations with particulate IOPs: it is highly correlated with $b_{bp}(\lambda_{550})$ ($r = 0.84$), $a_d(\lambda_{440})$ ($r = 0.90$), and $a_{ph}(\lambda_{440})$ ($r = 0.78$). This is physically consistent, as TSS represents the concentration of suspended particulate matter, encompassing both inorganic NAP and organic phytoplankton components, and therefore influences both light backscattering and absorption. DOC shows a strong positive correlation with $a_y(\lambda_{440})$ ($r = 0.72$). This relationship is expected because DOC is closely associated with CDOM concentration, which highly affects light absorption in the blue spectral region. TChl-a shows reasonably strong correlation with $a_{ph}(\lambda_{440})$ ($r = 0.88$), reflecting the dependence of phytoplankton absorption on chlorophyll concentration. The results of this correlation analysis indicate that the underlying physical correlations among variables are well preserved in the spectral library measurements.

The CIE 1931 chromaticity diagram in Fig. 5 shows the colour distribution of in situ samples in the bio-optical spectral library. The distribution of these samples follows a trajectory that closely resembles the chromaticity progression of the Forel-Ule (FU) colour scale (e.g., [44], [45]), and spans the FU range from blue and green to yellow and brown waters.

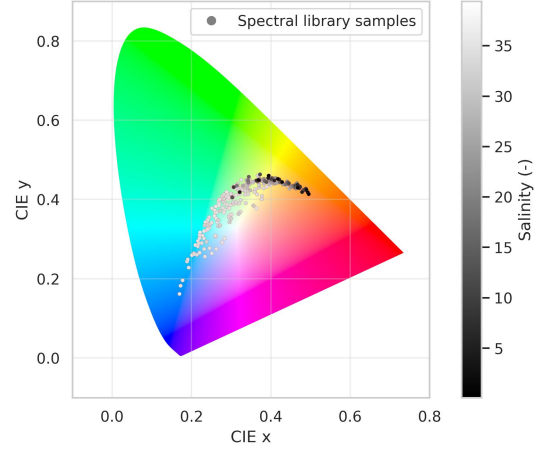


Fig. 5. The CIE 1931 chromaticity diagram showing the colour distribution of spectral library samples. Sample chromaticity coordinates were computed from R_{rs} spectra using the CIE 1931 2° Standard Observer under CIE Standard Daylight D65 illumination. Sample points are shaded in grayscale according to their salinity.

It was also observed from that figure that salinity exhibits a systematic gradient along the chromaticity trajectory, with higher-salinity samples generally associated with greener to bluer colours, and lower-salinity samples tending towards yellower to browner colours. These findings indicate that the bio-optical spectral library employed in this study provides a representative coverage of diverse water colours.

C. Analysis of the bio-optical forward model

Fig. 6 shows the sensitivity indices of the bio-optical forward model as computed with the Extended Fourier Amplitude Sensitivity Test (EFAST) [46]. It was found that the sensitivity of R_{rs} to TSS increases toward longer wavelengths, reflecting the higher influence of particle backscattering in the red region. In contrast, the sensitivity of R_{rs} to DOC is relatively higher in the blue region than in the red region, which is consistent with the relatively higher absorption of CDOM at shorter wavelengths. Compared with TSS and DOC, the sensitivity of R_{rs} to TChl-a displays overall lower sensitivity, with relatively higher sensitivity in the blue and near 680 nm regions corresponding to known chlorophyll absorption features in these regions. Previous studies have shown that absorption is generally more pronounced in the blue region than in the red due to CDOM and phytoplankton pigments, whereas scattering effects associated with suspended particles tend to become relatively more important at longer wavelengths (e.g., [47]). Compared with these well-established spectral characteristics of BGC parameters, the sensitivity analysis results in Fig. 6 suggest that the forward model is built upon assumptions that align well with the bio-optical properties of water constituents.

D. Analysis of the synthetic dataset

Fig. 7 shows the distribution of 10,000 samples from the synthetic dataset $\mathcal{D}$ in comparison with samples from the in situ spectral library. The histogram comparisons in Figs. 7a–c show that the BGC parameters (TSS, DOC, and TChl-a)

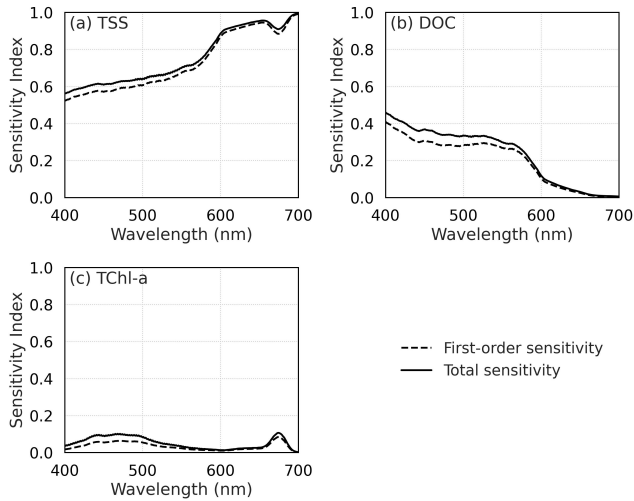


Fig. 6. Sensitivity indices of R_{rs} to variations in (a) TSS, (b) DOC, and (c) TChl-a across the 400–700 nm spectral range.

in the synthetic dataset not only span a similar range of values as the spectral library measurements, but also generally follow their distributional patterns. The Wasserstein distance between the synthetic and spectral library distributions is 1.77 mg/L for TSS, 0.57 mg/L for DOC, and 0.75 $\mu\text{g/L}$ for TChl-a. The spectral comparisons in Figs. 7d–g indicate that the synthetic SIOP spectra (a_d^* , a_y^* , a_{ph}^* , and b_{bp}^*) largely preserve the spectral correlations across wavelengths (*e.g.*, spectral smoothness and absorption features) observed in the spectral library measurements, and their shapes and magnitudes also generally resemble those of the measured spectra. The spectrally averaged Wasserstein distance is 0.0059 $\text{m}^2 \text{g}^{-1}$ for a_d^* , 0.0505 $\text{m}^2 \text{g}^{-1}$ for a_y^* , 0.0051 $\text{m}^2 \text{mg}^{-1}$ for a_{ph}^* , and 0.0016 $\text{m}^2 \text{g}^{-1}$ for b_{bp}^* . In particular, the chlorophyll-a absorption peaks in a_{ph}^* at 430–440 nm and 665–675 nm, evident in the spectral library spectra, are well preserved in the synthetic spectra, as shown from Fig. 7f. The simulated R_{rs} spectra also closely resemble those from the spectral library, in terms of both spectral shape and magnitude, as shown in Fig. 7h, with a spectrally averaged Wasserstein distance of 0.0009 sr^{-1} . The close agreement between the distributions of the synthetic samples and the spectral library samples, as shown in Fig. 7, demonstrates that the synthetic dataset is representative of Australian coastal waters.

While Fig. 7 presents distributional comparisons between the synthetic and spectral library samples for individual parameters, we further assessed whether the inter-parameter correlations observed in the spectral library are preserved in the synthetic dataset. Fig. 8 presents the joint distributions of selected parameter pairs for both synthetic and spectral library samples. The figure shows that the correlation patterns between parameters are similar for the synthetic and spectral library samples. A comparison of Pearson correlation coefficients (r) indicates that, for each parameter pair, the r values between the spectral library and synthetic data are closely aligned, with differences ranging from 0.01 (Figs. 8a and f) to 0.09 (Fig. 8g). The results shown in Fig. 8 indicate that the

inter-parameter correlations observed in the spectral library have been largely preserved in the synthetic dataset.

E. Analysis of regional distinctions

Fig. 9 shows the regional distinctions among the experimental sites in their BGC and R_{rs} distributions. Fig. 9a shows the distribution of TSS, DOC, and TChl-a concentrations in the BGC feature space across the five experimental sites. Fitzroy Estuary exhibits the highest TSS and DOC concentrations overall, with moderate to elevated TChl-a levels. Keppel Bay is characterised by moderate TSS and DOC concentrations but comparatively lower TChl-a values. Boston Bay and Cockburn Sound cluster towards lower TSS and DOC concentrations, with Boston Bay showing a broader spread in TChl-a than Cockburn Sound. Lucinda Jetty shows moderate TSS, DOC, and TChl-a values as compared with other sites.

In the CIE 1931 chromaticity diagram shown in Fig. 9b, the R_{rs} measurements display distinct colour groupings that reflect differences in apparent water colour among the regions. Waters from Fitzroy Estuary display brown colours, characteristic of turbid estuarine conditions. Keppel Bay exhibits less brown colours due to moderate river influence. Boston Bay and Cockburn Sound are associated with clearer green to blue-green waters, typical of less turbid coastal environments with lower suspended sediment loads. Lucinda Jetty is more green-coloured than the clearer coastal sites but less brown than the estuarine waters.

In summary, the regional distinctions observed in Fig. 9 are clearly reflected in both the BGC parameters and their corresponding optical responses across the experimental sites. This variability suggests that constituent dynamics and R_{rs} characteristics are region-dependent. These findings underpin the need to develop region-adaptable approaches for BGC parameter retrieval to ensure robust, accurate, and transferable performance across diverse coastal and estuarine water environments.

F. Accuracies of BGC parameter retrieval

Fig. 10 shows the relationship between TSS, DOC, and TChl-a measurements and those predicted from R_{rs} using the proposed approach. Across the five experimental sites, TSS showed a Log- R^2 of 0.970 and a Log-MAE of 0.264. DOC showed a Log- R^2 of 0.645 and a Log-MAE of 0.390. TChl-a showed a Log- R^2 of 0.663 and a Log-MAE of 0.512. The per-site accuracies are given in Table V, with Log- R^2 ranging from 0.194 to 0.895 and Log-MAE from 0.014 to 0.792 across sites and parameters.

Table VI compares the predictive accuracy of the proposed method against the benchmark models. It was observed from the table that the proposed approach generally outperforms the benchmark models. For TSS, the proposed model improves the Log- R^2 from 0.940 to 0.970, representing a 3.2% increase over the second-best DL-RS model, while reducing the Log-MAE from 0.339 to 0.264, corresponding to a 22.1% reduction. For DOC, it achieves a 6.3% higher Log- R^2 (0.645 vs 0.607) and a 2.7% lower Log-MAE (0.390 vs 0.401) than the second-best DL-RS model. A higher improvement was observed for

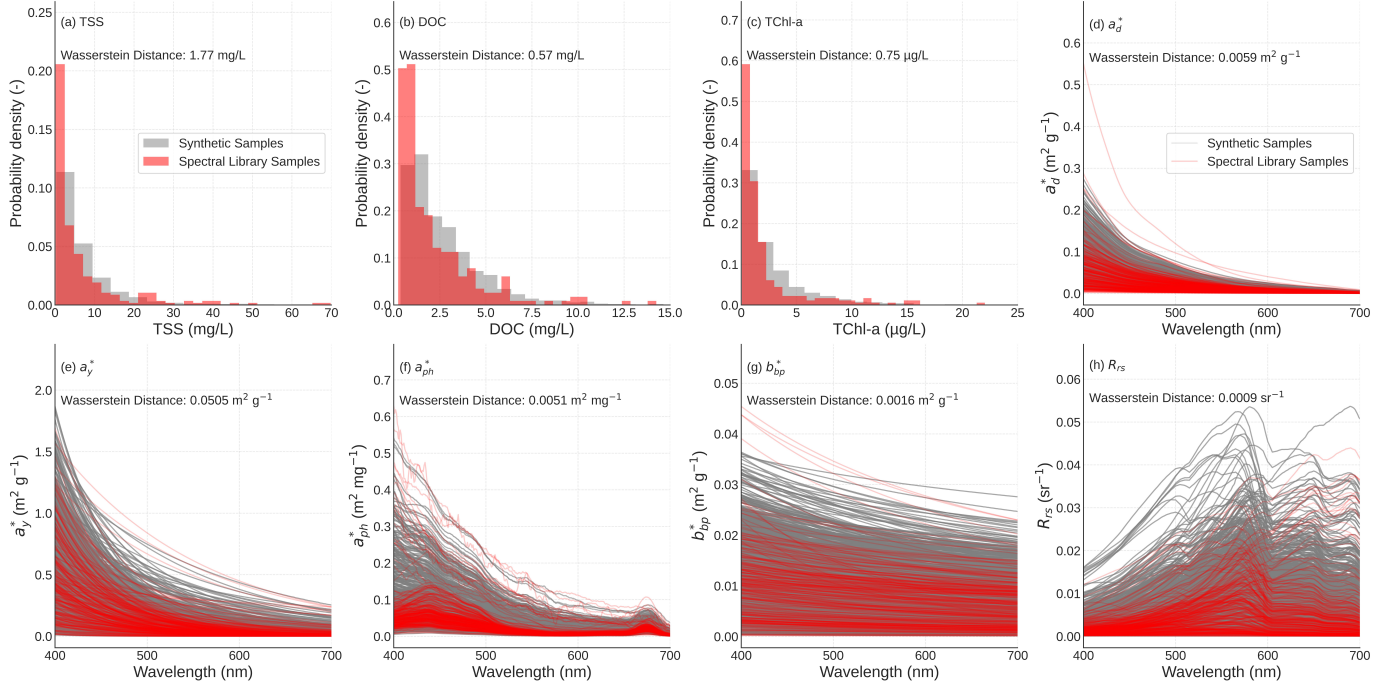


Fig. 7. Comparisons between samples from the synthetic dataset $\mathcal{D}$ with those from the in situ spectral library for (a–c) BGC parameters (TSS, DOC, and TChl-a), (d–g) SIOPs (a_d^* , a_y^* , a_{ph}^* , and b_{dp}^*), and (h) the corresponding R_{rs} . The Wasserstein distance is calculated to quantify the distributional difference.

TABLE V
PER-SITE ACCURACIES OF BIOGEOCHEMICAL PARAMETERS PREDICTED WITH THE PROPOSED APPROACH.

Site	Parameter	Log- R^2	Log-MAE
Fitzroy Estuary	TSS	0.539	0.390
	DOC	0.515	0.375
	TChl-a	0.523	0.184
Keppel Bay	TSS	0.673	0.335
	DOC	0.458	0.296
	TChl-a	0.194	0.372
Boston Bay	TSS	0.869	0.186
	DOC	0.349	0.462
	TChl-a	0.515	0.792
Cockburn Sound	TSS	0.668	0.296
	DOC	0.272	0.386
	TChl-a	0.437	0.436
Lucinda Jetty	TSS	0.895	0.174
	DOC	0.654	0.014
	TChl-a	0.478	0.628

TChl-a, where the proposed method outperforms HyperEST by increasing the Log- R^2 from 0.581 to 0.663 (14.1% improvement) and reducing the Log-MAE from 0.601 to 0.512 (14.8% reduction).

The ablation study results are shown in Table VII where we tested the effects of removing the physics-aware pretraining stage (no prior) and region-specific adaptation stage (no adaptation) from the full model. Removing the pretrained prior reduced Log- R^2 from 0.970 to 0.950 for TSS, from 0.645 to 0.566 for DOC, and from 0.663 to 0.571 for TChl-a, while increasing the corresponding Log-MAE values from 0.264, 0.390, and 0.512 to 0.355, 0.448, and 0.607, respectively. This indicates that the physics-aware prior provides a more effective initialisation for learning regional BGC- R_{rs} relationships. A

TABLE VI
ACCURACY COMPARISON BETWEEN THE PROPOSED APPROACH AND BENCHMARK MODELS. THE BEST AND SECOND-BEST RESULTS ARE SHOWN IN BOLD AND UNDERLINED, RESPECTIVELY.

Parameter	Model	Log- R^2	Log-MAE
TSS	Empirical TSS [42]	0.911	0.525
	DL-RS [3]	<u>0.940</u>	<u>0.339</u>
	HyperEST [9]	0.923	0.458
	Proposed	0.970	0.264
DOC	Empirical DOC [43]	0.401	0.545
	DL-RS [3]	<u>0.607</u>	<u>0.401</u>
	Proposed	0.645	0.390
TChl-a	Empirical TChl-a [33]	0.271	0.787
	HyperEST [9]	<u>0.581</u>	<u>0.601</u>
	Proposed	0.663	0.512

TABLE VII
ABLATION STUDY RESULTS FOR THE PROPOSED APPROACH.

Model Setup	TSS		DOC		TChl-a	
	Log- R^2	Log-MAE	Log- R^2	Log-MAE	Log- R^2	Log-MAE
Full	0.970	0.264	0.645	0.390	0.663	0.512
No prior	0.950	0.355	0.566	0.448	0.571	0.607
No adapt.	0.696	1.055	0.126	0.650	0.220	0.942

decrease in performance was also observed when region-specific adaptation was removed: Log- R^2 decreased to 0.696, 0.126, and 0.220 for TSS, DOC, and TChl-a, respectively, whereas Log-MAE increased to 1.055, 0.650, and 0.942. These results show that both stages of the proposed approach contribute to the retrieval performance.

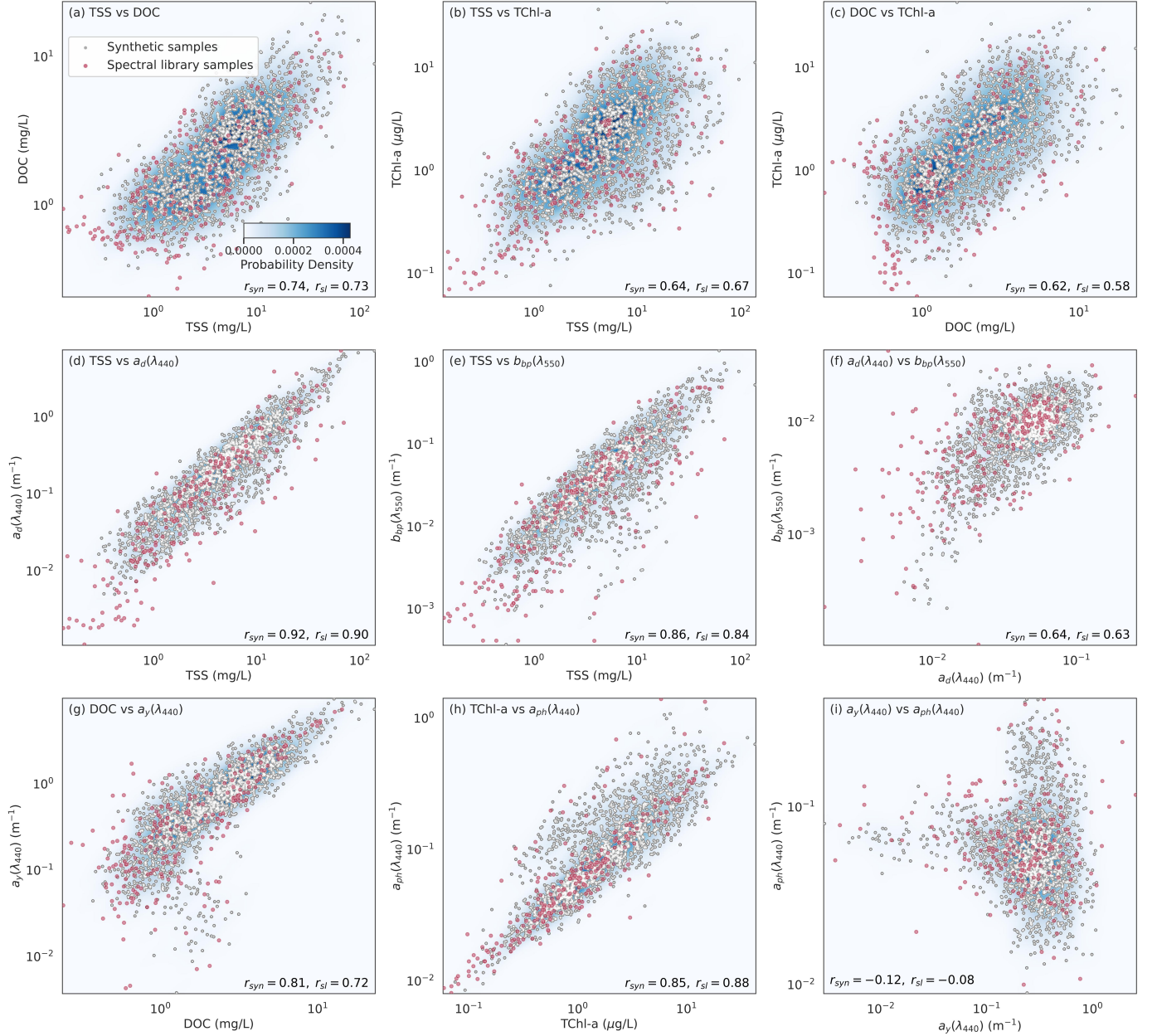


Fig. 8. Comparison of parameter correlations between synthetic samples and spectral library samples for the joint distributions of (a) TSS and DOC, (b) TSS and TChl-a, (c) DOC and TChl-a, (d) TSS and $a_d(\lambda_{440})$, (e) TSS and $b_{bp}(\lambda_{550})$, (f) $a_d(\lambda_{440})$ and $b_{bp}(\lambda_{550})$, (g) DOC and $a_y(\lambda_{440})$, (h) TChl-a and $a_{ph}(\lambda_{440})$, and (i) $a_y(\lambda_{440})$ and $a_{ph}(\lambda_{440})$. Background colour indicates the probability density of the joint distributions. r_{syn} and r_{sl} stand for the Pearson correlation coefficients between synthetic samples and spectral library samples, respectively.

G. Time-series analysis of BGC and R_{rs} data

Fig. 11 shows time-series variations in hyperspectral R_{rs} and corresponding BGC parameters (TSS, DOC, and TChl-a) and ancillary physical variables (temperature and salinity) at the experimental site of Fitzroy Estuary, with Fig. 11a showing data over the entire experimental period from 26 April 2023 to 23 September 2024, and Fig. 11b showing the zoomed-in view for the period between 10 March and 31 May 2024. It was observed from the figure that, the Fitzroy Estuary site, characterised by high river influence and a subtropical climate (Table II), exhibits high TSS concentrations with a periodic variability over time, alongside elevated and highly

dynamic levels of DOC and TChl-a. The periodic variability likely reflects the combined influence of tidal cycles and fluctuations in Fitzroy River discharge, as evidenced by the salinity variations shown in the figure, which indicate varying degrees of mixing between river water and seawater.

The time series of R_{rs} at the Fitzroy Estuary site (Fig. 11) shows relatively high R_{rs} values, reaching up to approximately 0.014 sr^{-1} , particularly in the green to red spectral region (550–650 nm). The time-series R_{rs} measurements also exhibit pronounced temporal variability, including a periodic pattern over the observation period (Fig. 11). Given that the Fitzroy Estuary is a river-influenced system characterised by relatively

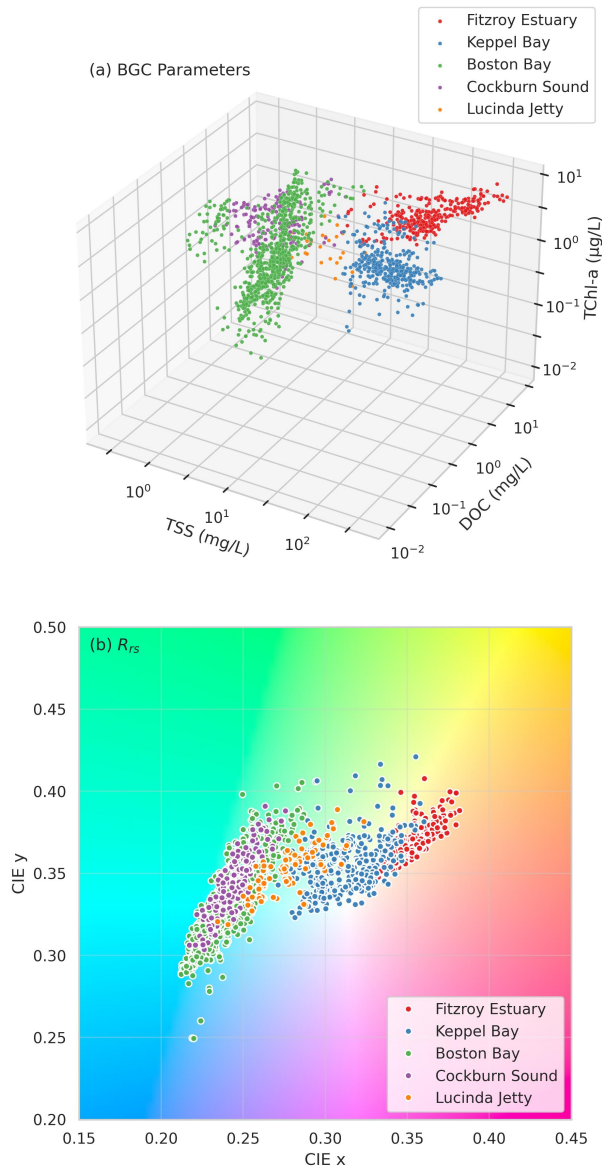


Fig. 9. (a) Distribution of BGC parameters, including TSS, DOC, and TChl-a concentrations, at the experimental sites. (b) The CIE 1931 chromaticity diagram showing the distribution of water-leaving R_{rs} measured at the experimental sites.

high TSS concentrations, the elevated R_{rs} in the green-to-red region is likely driven by enhanced particulate backscattering associated with high sediment loads. The observed periodic variability in R_{rs} may further reflect temporal changes in BGC properties, influenced by tidal dynamics and variations in river discharge.

The time-series plots for other sites (Keppel Bay, Boston Bay, and Cockburn Sound) are shown in Fig. 12. It is worth noting that the Lucinda Jetty result is not shown in the figure because the data were not sampled continuously in time at this site. Fig. 12a shows time-series variations in R_{rs} , TSS, DOC, TChl-a, temperature, and salinity at the Keppel Bay site over the experimental period from 1 June 2023 to 28 November 2024. This site is located downstream of the Fitzroy Estuary site and is closer to the open ocean. Accordingly,

higher salinity was observed at Keppel Bay compared with the Fitzroy Estuary site. For the BGC parameters, concentrations of TSS, DOC, and TChl-a at the Keppel Bay site are generally lower than those at the Fitzroy Estuary site. Periodic variations are also evident in the time-series BGC measurements at the Keppel Bay site, although they are less pronounced than those observed at the Fitzroy Estuary site. As compared with the Fitzroy Estuary site, lower levels of R_{rs} were observed at the Keppel Bay site, as shown in Fig. 12a. The peak R_{rs} is generally below 0.008 sr^{-1} and centred in the green spectral region around 550 nm.

Fig. 12b shows the time-series data for the Boston Bay site over the experimental period from 1 September 2022 to 2 February 2025. Different from the subtropical Fitzroy Estuary and Keppel Bay sites, the Boston Bay site is a temperate site without direct river influence (Table II). Accordingly, it was observed from Fig. 12b that the Boston Bay site is characterised by consistently low TSS and DOC and minimal temporal variability, indicative of clearer and more optically stable waters. The R_{rs} levels at the Boston Bay site are generally lower than the Fitzroy Estuary and Keppel Bay sites, with the peak R_{rs} below 0.003 sr^{-1} and centred in the green/blue spectral region of 450–550 nm.

An important observation from Fig. 12b is the clear seasonal pattern in TChl-a concentrations at the Boston Bay site, which increase during the winter months. This seasonal rise begins around March/April each year over the experimental period, remains elevated throughout winter, and declines to lower levels by July/August. To verify this pattern, grab samples were collected at the Boston Bay site on several occasions during the experiment, and TChl-a concentrations were analysed by the CSIRO Hydrochemistry Laboratories in Hobart, Tasmania. As shown in Fig. 12b, the laboratory-analysed TChl-a values are in good agreement with the in situ time-series measurements, consistently showing higher concentrations between March/April and July/August. In 2025, the year following the experimental period of this study, a harmful algal bloom broke out in March along the South Australian coast⁴, including the Boston Bay region, resulting in mass fish deaths and significant impacts on marine ecosystems, fisheries, aquaculture, tourism, and coastal activities. The observation reported in this study may therefore provide important baseline information on seasonal phytoplankton dynamics and could contribute to a better understanding of environmental conditions that precede or favour the development of such bloom events.

Fig. 12c shows the time-series data for the Cockburn Sound site over the experimental period from 15 July 2023 to 18 February 2024. As observed from the figure, Cockburn Sound, being a temperate embayment with low freshwater influence (Table II), exhibits generally low to moderate concentrations of TSS, DOC, and TChl-a over the experimental period, with some temporal variability. Correspondingly, this site displays relatively low R_{rs} values, with peak R_{rs} typically below 0.003 sr^{-1} and centred in the green spectral region around 500–550 nm.

⁴https://en.wikipedia.org/wiki/2025_algal_bloom_in_South_Australia

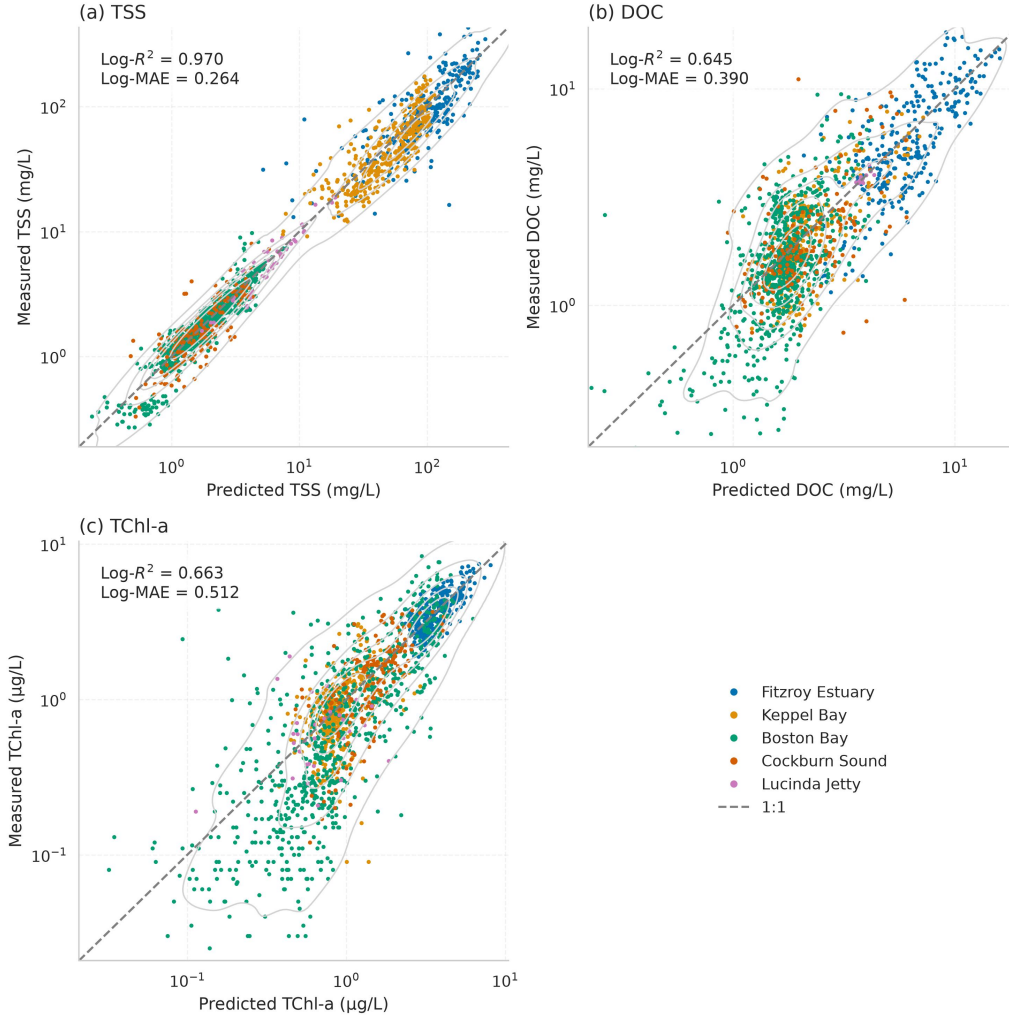


Fig. 10. Comparison between predicted and measured concentrations of (a) TSS, (b) DOC, and (c) TChl-a across the five experimental sites. Axes are shown on $\log_{10}$ scales. Grey lines represent contours of equal probability mass for the purpose of enhancing visualisation of the data distribution.

It was also observed from Figs. 11 and 12 that, despite clear regional differences among the experimental sites, the BGC parameters predicted from hyperspectral R_{rs} observations using the proposed method generally agree well with the in situ BGC measurements in both magnitude and temporal dynamics. This consistency demonstrates the capability of in situ hyperspectral sensing to reliably capture BGC variability over time, showing its potential as a cost-effective approach for continuous time-series monitoring of water quality.

V. CONCLUSION

This study presented a region-adaptable framework for retrieving key BGC parameters, including TSS, DOC, and TChl-a, from in situ hyperspectral R_{rs} measurements. The proposed method integrates physical knowledge with a data-driven learning strategy through a two-stage physics-aware meta-learning framework. In the first stage, a physics-based bio-optical forward model was used to generate a large physics-guided synthetic dataset from a continental bio-optical spectral library. This dataset was then used to pretrain a region-agnostic base model with meta-learning, allowing the model to

learn fundamental physical relationships. In the second stage, the pretrained base model was adapted to individual regions using local in situ observations, allowing the framework to account for region-specific bio-optical variability.

Evaluation across five experimental sites showed clear regional differences in both BGC parameters and hyperspectral R_{rs} signatures. In the site-aggregated comparison, the proposed approach outperformed five benchmark models. Relative to the second-best model, it improved $\text{Log-}R^2$ by 3.2%, 6.3%, and 14.1% for TSS, DOC, and TChl-a, respectively, while reducing Log-MAE by 22.1%, 2.7%, and 14.8%. The ablation results showed that both physics-aware pretraining and region-specific adaptation contributed to retrieval accuracy. The predicted time series also agreed well with in situ measurements in both magnitude and temporal variation, showing the potential of the framework for continuous coastal water-quality monitoring.

While the results are promising, there remain opportunities for further improvement. The current evaluation focuses on a limited number of coastal and estuarine sites, which may not fully represent the broader diversity of optical water

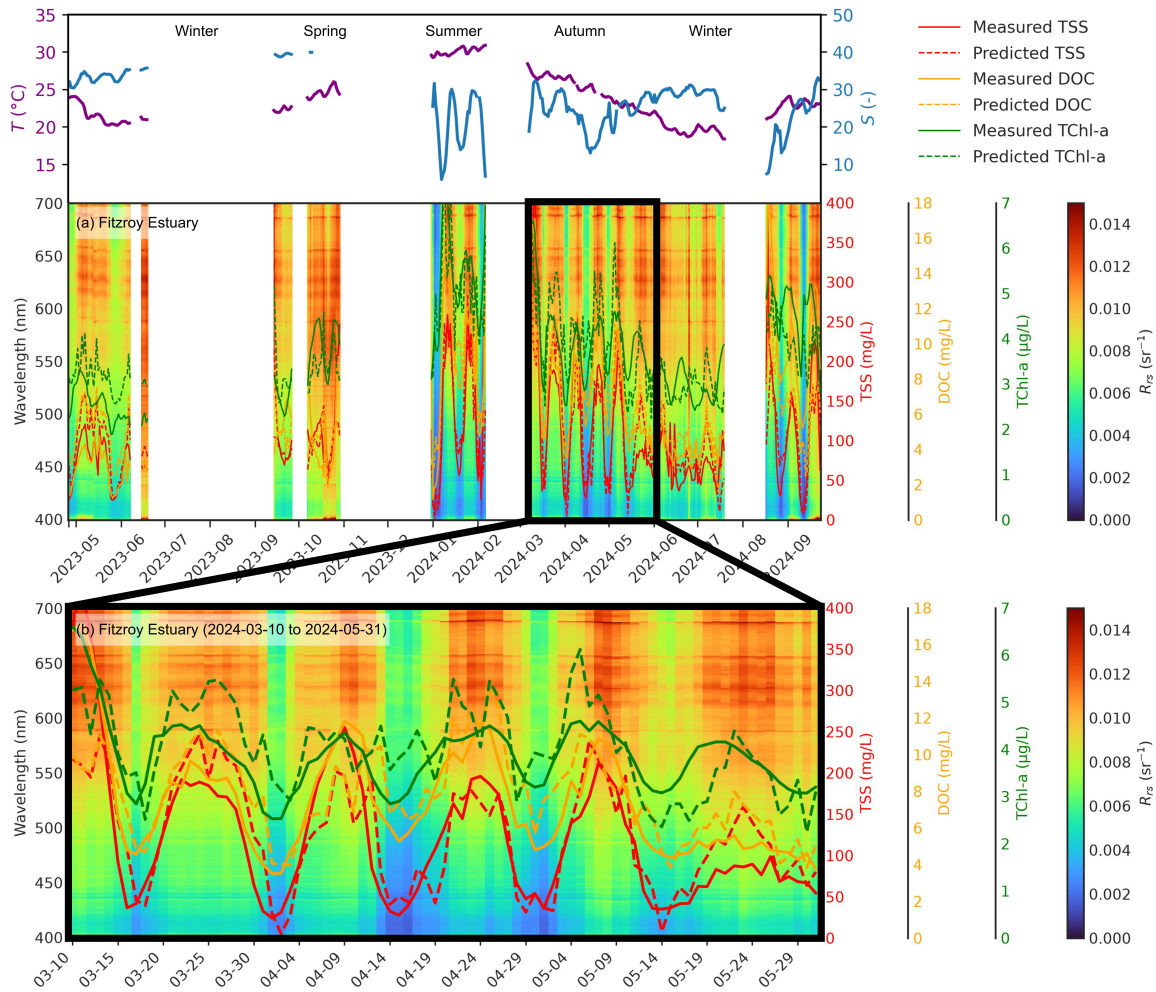


Fig. 11. (a) Temporal variation of water-leaving R_{rs} and corresponding in situ measurements of temperature (T), salinity (S), TSS, DOC, and TChl-a at the experimental site of Fitzroy Estuary over the experimental period from 26 April 2023 to 23 September 2024. The TSS, DOC, and TChl-a predictions generated by the proposed approach from hyperspectral R_{rs} observations in the test folds are also shown. (b) Zoomed-in view for the period between 10 March and 31 May 2024. Background colour shows hyperspectral R_{rs} as a function of wavelength (400–700 nm) and time. Blank regions in the plots denote periods of instrument malfunction.

types, particularly inland waters with stronger spatiotemporal variability. Future work would be on extending validation to additional aquatic environments, and exploring the application of the proposed framework to satellite-based hyperspectral observations.

ACKNOWLEDGEMENT

The code and datasets of this work are publicly available at: <https://github.com/yiqing-csiro/wq-meta>. We would like to acknowledge the in situ data from Lucinda Jetty Coastal Observatory (Principal Investigator: Dr Thomas Schroeder). These data were sourced from Australia's Integrated Marine Observing System (IMOS). IMOS is enabled by the National Collaborative Research Infrastructure Strategy (NCRIS). It is operated by a consortium of institutions as an unincorporated joint venture, with the University of Tasmania as Lead Agent. We acknowledge the SARDI field team (Mr Ian Moody and Mr Paul Malthouse) for their help with Boston Bay buoy setup and maintenance. We are grateful to Dr Albertina Dias with

CSIRO Hydrochemistry Laboratories, Hobart, Tasmania, for analysing the grab water samples.

REFERENCES

- [1] W. Zhi, A. P. Appling, H. E. Golden, J. Podgorski, and L. Li, "Deep learning for water quality," *Nature Water*, vol. 2, no. 3, pp. 228–241, 2024.
- [2] M. K. Lehmann, D. Gurlin, N. Pahlevan, K. Alikas, T. Conroy, J. Anstee, S. V. Balasubramanian, C. C. Barbosa, C. Binding, A. Bracher *et al.*, "GLORIA - A globally representative hyperspectral in situ dataset for optical sensing of water quality," *Scientific Data*, vol. 10, no. 1, p. 100, 2023.
- [3] S. K. Unnithan, N. Cherukuru, T. Ingleton, E. Lehmann, M. Paget, Y. Guo, N. Drayson, and G. Kerrisk, "Mapping total suspended solids (TSS) and dissolved organic carbon (DOC) in complex coastal waters using deep learning enhanced remote sensing," *Ecological Informatics*, vol. 90, p. 103276, 2025.
- [4] L. B. Medina, D. Nguyen, K. Joehnk, Y. Guo, R. C. Deo, and S. Prasad, "Data-driven prediction of dissolved oxygen to identify fish kill factors: Case study in Redbank weir at Murrumbidgee River, Australia," *Ecological Informatics*, vol. 95, p. 103778, 2026.
- [5] B. Schaffelke, J. Carleton, M. Skuza, I. Zagorskis, and M. J. Furnas, "Water quality in the inshore Great Barrier Reef lagoon: Implications for long-term monitoring and management," *Marine Pollution Bulletin*, vol. 65, no. 4-9, pp. 249–260, 2012.

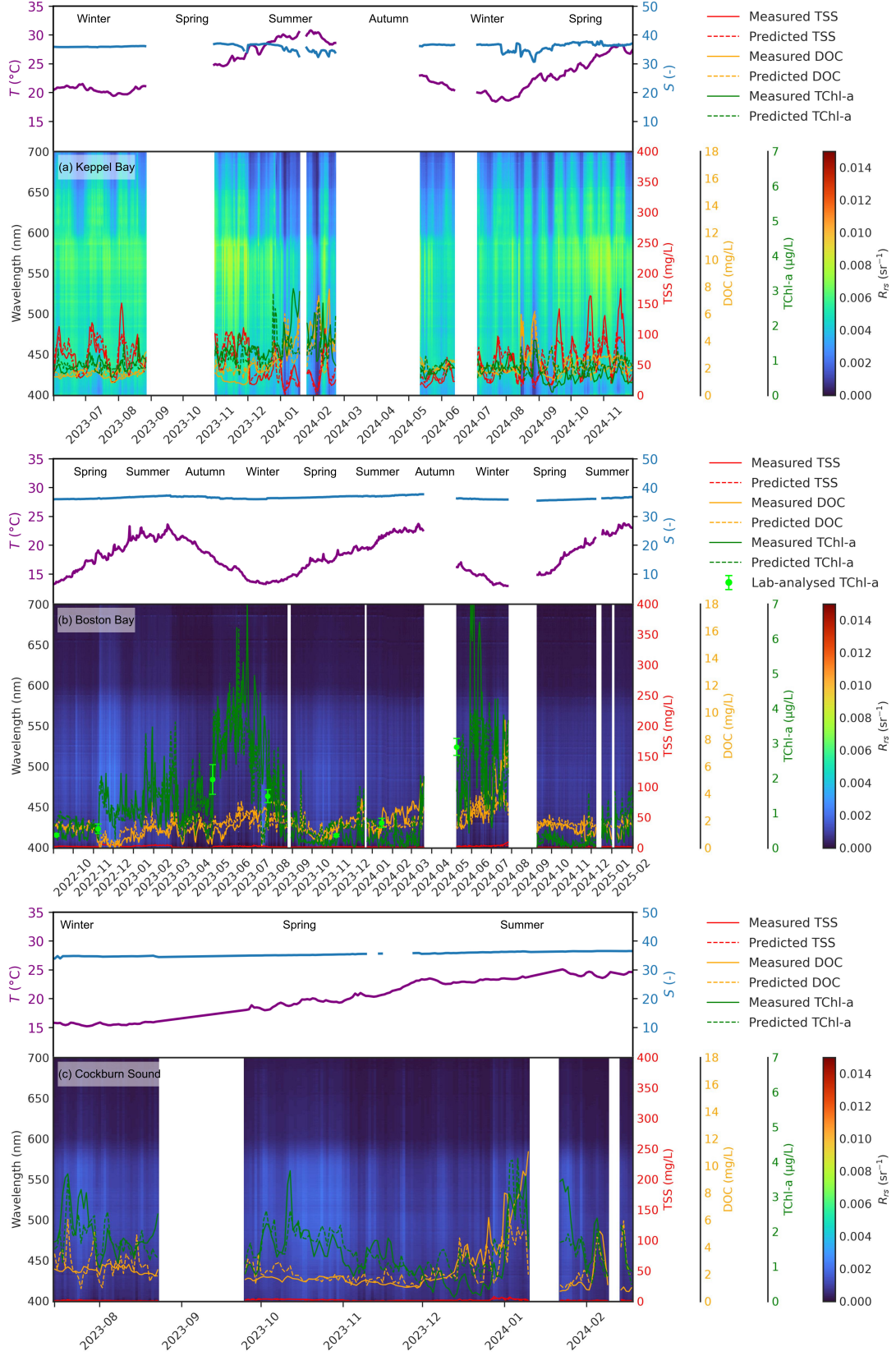
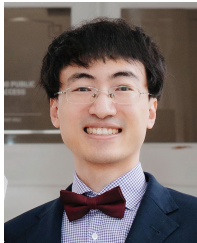


Fig. 12. Temporal variation of water-leaving R_{rs} and corresponding in situ measurements of temperature (T), salinity (S), TSS, DOC, and TChl-a at the experimental sites of (a) Keppel Bay, (b) Boston Bay, and (c) Cockburn Sound. Background colour shows hyperspectral R_{rs} as a function of wavelength (400–700 nm) and time. The TSS, DOC, and TChl-a predictions generated by the proposed approach from hyperspectral R_{rs} observations in the test folds are also shown. Laboratory-analysed grab samples of TChl-a (mean $\pm$ standard deviation) are shown as light green markers. Blank regions in the plots denote periods of instrument malfunction.

- [6] Y. Guo, N. Cherukuru, E. Lehmann, X. Qi, M. J. Doubell, S. Kesav Unnithan, and M. Feng, "Decadal analysis of sea surface temperature patterns, climatology, and anomalies in temperate coastal waters with Landsat-8 TIRS observations," *GIScience & Remote Sensing*, vol. 62, no. 1, p. 2518623, 2025.
- [7] R. E. O'Shea, N. Pahlevan, B. Smith, E. Boss, D. Gurlin, K. Alikas, K. Kangro, R. M. Kudela, and D. Vaičiūtė, "A hyperspectral inversion framework for estimating absorbing inherent optical properties and biogeochemical parameters in inland and coastal waters," *Remote Sensing of Environment*, vol. 295, p. 113706, 2023.
- [8] J. Lou, B. Liu, Y. Xiong, X. Zhang, and X. Yuan, "Variational autoencoder framework for hyperspectral retrievals (Hyper-VAE) of phytoplankton absorption and chlorophyll a in coastal waters for NASA's EMIT and PACE missions," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 63, pp. 1–16, 2025.
- [9] C. Luo, W. Xiang, K. Han, L. Yu, Y. Guo, S. K. Unnithan, X. Qi, and N. Cherukuru, "HyperEst: Context-aware self-supervised pretraining for hyperspectral and multispectral water quality estimation," *International Journal of Applied Earth Observation and Geoinformation*, vol. 143, p. 104761, 2025.
- [10] E. T. Baker and J. W. Lavelle, "The effect of particle size on the light attenuation coefficient of natural suspensions," *Journal of Geophysical Research: Oceans*, vol. 89, no. C5, pp. 8197–8203, 1984.
- [11] Y. Huot, M. Babin, F. Bruyant, C. Grob, M. Twardowski, and H. Claustre, "Relationship between photosynthetic parameters and different proxies of phytoplankton biomass in the subtropical ocean," *Biogeosciences*, vol. 4, no. 5, pp. 853–868, 2007.
- [12] M. Babin, D. Stramski, G. M. Ferrari, H. Claustre, A. Bricaud, G. Obolensky, and N. Hoepffner, "Variations in the light absorption coefficients of phytoplankton, nonalgal particles, and dissolved organic matter in coastal waters around Europe," *Journal of Geophysical Research: Oceans*, vol. 108, no. C7, p. 3211, 2003.
- [13] N. B. Nelson and D. A. Siegel, "The global distribution and dynamics of chromophoric dissolved organic matter," *Annual Review of Marine Science*, vol. 5, no. 1, pp. 447–476, 2013.
- [14] D. A. Hansell and C. A. Carlson, "Deep-ocean gradients in the concentration of dissolved organic carbon," *Nature*, vol. 395, no. 6699, pp. 263–266, 1998.
- [15] D. A. Hansell, C. A. Carlson, D. J. Repeta, and R. Schlitzer, "Dissolved organic matter in the ocean: A controversy stimulates new insights," *Oceanography*, vol. 22, no. 4, pp. 202–211, 2009.
- [16] R. Del Vecchio and N. V. Blough, "Spatial and seasonal distribution of chromophoric dissolved organic matter and dissolved organic carbon in the Middle Atlantic Bight," *Marine Chemistry*, vol. 89, no. 1–4, pp. 169–187, 2004.
- [17] Z. Lee, K. L. Carder, and R. A. Arnone, "Deriving inherent optical properties from water color: a multiband quasi-analytical algorithm for optically deep waters," *Applied Optics*, vol. 41, pp. 5755–5772, 2002.
- [18] P. J. Werdell, B. A. Franz, S. W. Bailey, G. C. Feldman, E. Boss, V. E. Brando, M. Dowell, T. Hirata, S. J. Lavender, Z. Lee, H. Loisel, S. Maritorena, F. Mélin, T. S. Moore, T. J. Smyth, D. Antoine, E. Devred, D. Doxaran, O. H. Fanton d'Andon, and A. Mangin, "Generalized ocean color inversion model for retrieving marine inherent optical properties," *Applied Optics*, vol. 52, no. 10, pp. 2019–2037, 2013.
- [19] N. Cherukuru, P. Martin, N. Sanwlani, A. Mujahid, and M. Müller, "A semi-analytical optical remote sensing model to estimate suspended sediment and dissolved organic carbon in tropical coastal waters influenced by peatland-draining river discharges off Sarawak, Borneo," *Remote Sensing*, vol. 13, no. 1, p. 99, 2020.
- [20] N. Cherukuru, V. E. Brando, D. Blondeau-Patissier, P. W. Ford, L. A. Clementson, and B. J. Robson, "Impact of wet season river flood discharge on phytoplankton absorption properties in the southern Great Barrier Reef region coastal waters," *Estuarine, Coastal and Shelf Science*, vol. 196, pp. 379–386, 2017.
- [21] F. Mélin and V. Vantrepotte, "How optically diverse is the coastal ocean?" *Remote Sensing of Environment*, vol. 160, pp. 235–251, 2015.
- [22] E. Spyarakos, R. O'donnell, P. D. Hunter, C. Miller, M. Scott, S. G. Simis, C. Neil, C. C. Barbosa, C. E. Binding, S. Bradt *et al.*, "Optical types of inland and coastal waters," *Limnology and Oceanography*, vol. 63, no. 2, pp. 846–870, 2018.
- [23] X. Zhou, C. Liu, A. Akbar, H. Wang, and D. Carrion, "Spectral similarity-driven transfer learning for cross-regional retrieval of non-optically active water quality parameters," *IEEE Transactions on Geoscience and Remote Sensing*, 2026.
- [24] Y. Guo, K. Mokany, S. R. Levick, J. Yang, and P. Moghadam, "Spatio-former: a geo-encoded transformer for large-scale plant species richness prediction," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 63, pp. 1–16, 2025.
- [25] C. Neil, E. Spyarakos, P. D. Hunter, and A. N. Tyler, "A global approach for chlorophyll-a retrieval across optically complex inland waters based on optical water types," *Remote Sensing of Environment*, vol. 229, pp. 159–178, 2019.
- [26] Z. Mao, J. Chen, D. Pan, B. Tao, and Q. Zhu, "A regional remote sensing algorithm for total suspended matter in the East China Sea," *Remote Sensing of Environment*, vol. 124, pp. 819–831, 2012.
- [27] I. Ogashawara, D. R. Mishra, R. F. Nascimento, E. H. Alcantara, M. Kampel, and J. L. Stech, "Re-parameterization of a quasi-analytical algorithm for colored dissolved organic matter dominant inland waters," *International Journal of Applied Earth Observation and Geoinformation*, vol. 53, pp. 128–145, 2016.
- [28] C. D. Mobley, "Estimation of the remote-sensing reflectance from above-surface measurements," *Applied Optics*, vol. 38, no. 36, pp. 7442–7455, 1999.
- [29] K. Oubelkheir, L. A. Clementson, I. T. Webster, P. W. Ford, A. G. Dekker, L. C. Radke, and P. Daniel, "Using inherent optical properties to investigate biogeochemical dynamics in a tropical macrotidal coastal system," *Journal of Geophysical Research: Oceans*, vol. 111, no. C7, 2006.
- [30] D. Blondeau-Patissier, V. E. Brando, K. Oubelkheir, A. G. Dekker, L. A. Clementson, and P. Daniel, "Bio-optical variability of the absorption and scattering properties of the Queensland inshore and reef waters, Australia," *Journal of Geophysical Research: Oceans*, vol. 114, no. C5, p. C05003, 2009.
- [31] N. Cherukuru, V. E. Brando, T. Schroeder, L. A. Clementson, and A. G. Dekker, "Influence of river discharge and ocean currents on coastal optical properties," *Continental Shelf Research*, vol. 84, pp. 188–203, 2014.
- [32] N. Cherukuru, P. L. Davies, V. E. Brando, J. M. Anstee, M. E. Baird, L. A. Clementson, and M. A. Doblin, "Physical oceanographic processes influence bio-optical properties in the Tasman Sea," *Journal of Sea Research*, vol. 110, pp. 1–7, 2016.
- [33] N. Cherukuru, A. G. Dekker, N. J. Hardman-Mountford, L. A. Clementson, and P. A. Thompson, "Bio-optical variability in multiple water masses across a tropical shelf: Implications for ocean colour remote sensing models," *Estuarine, Coastal and Shelf Science*, vol. 219, pp. 223–230, 2019.
- [34] K. Oubelkheir, P. W. Ford, L. A. Clementson, N. Cherukuru, G. Fry, and A. D. Steven, "Impact of an extreme flood event on optical and biogeochemical properties in a subtropical coastal periurban embayment (Eastern Australia)," *Journal of Geophysical Research: Oceans*, vol. 119, no. 9, pp. 6024–6045, 2014.
- [35] K. Oubelkheir, P. W. Ford, N. Cherukuru, L. A. Clementson, C. Petus, M. Devlin, T. Schroeder, and A. D. Steven, "Impact of a tropical cyclone on terrestrial inputs and bio-optical properties in Princess Charlotte Bay (Great Barrier Reef lagoon)," *Remote Sensing*, vol. 15, no. 3, p. 652, 2023.
- [36] R. M. Pope and E. S. Fry, "Absorption spectrum (380–700 nm) of pure water. II. integrating cavity measurements," *Applied Optics*, vol. 36, no. 33, pp. 8710–8723, 1997.
- [37] R. Röttgers, D. McKee, and C. Utschig, "Temperature and salinity correction coefficients for light absorption by water in the visible to infrared spectral region," *Optics Express*, vol. 22, no. 21, pp. 25 093–25 108, 2014.
- [38] E. Boss and W. S. Pegau, "Relationship of light scattering at an angle in the backward direction to the backscattering coefficient," *Applied Optics*, vol. 40, no. 30, pp. 5503–5507, 2001.
- [39] C. D. Mobley, "A numerical model for the computation of radiance distributions in natural waters with wind-roughened surfaces," *Limnology and Oceanography*, vol. 34, no. 8, pp. 1473–1483, 1989.
- [40] B. N. Seegers, R. P. Stumpf, B. A. Schaeffer, K. A. Loftin, and P. J. Werdell, "Performance metrics for the assessment of satellite data products: an ocean color case study," *Optics Express*, vol. 26, no. 6, pp. 7404–7422, 2018.
- [41] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *International Conference on Machine Learning*. PMLR, 2017, pp. 1126–1135. [Online]. Available: <https://proceedings.mlr.press/v70/finn17a.html>
- [42] J. Choo, N. Cherukuru, E. Lehmann, M. Paget, A. Mujahid, P. Martin, and M. Müller, "Spatial and temporal dynamics of suspended sediment concentrations in coastal waters of the South China Sea, off Sarawak, Borneo: ocean colour remote sensing observations and analysis," *Biogeosciences*, vol. 19, no. 24, pp. 5837–5857, 2022.

- [43] N. Cherukuru, P. W. Ford, R. J. Matear, K. Oubelkheir, L. A. Clementson, K. Suber, and A. D. Steven, "Estimating dissolved organic carbon concentration in turbid coastal waters using optical remote sensing observations," *International Journal of Applied Earth Observation and Geoinformation*, vol. 52, pp. 149–154, 2016.
- [44] M. Wernand and H. Van der Woerd, "Spectral analysis of the Forel-Ule ocean colour comparator scale," *Journal of the European Optical Society-Rapid Publications*, vol. 5, p. 10014s, 2010.
- [45] S. Novoa, M. Wernand, and H. Van der Woerd, "The Forel-Ule scale revisited spectrally: preparation protocol, transmission measurements and chromaticity," *Journal of the European Optical Society-Rapid Publications*, vol. 8, p. 13057, 2013.
- [46] A. Saltelli, S. Tarantola, and K.-S. Chan, "A quantitative model-independent method for global sensitivity analysis of model output," *Technometrics*, vol. 41, no. 1, pp. 39–56, 1999.
- [47] C. Manzo, M. Bresciani, C. Giardino, F. Braga, and C. Bassani, "Sensitivity analysis of a bio-optical model for Italian lakes focused on Landsat-8, Sentinel-2 and Sentinel-3," *European Journal of Remote Sensing*, vol. 48, no. 1, pp. 17–32, 2015.



Yiqing Guo received the B.Eng. and M.Eng. degrees from Beihang University, Beijing, China, in 2012 and 2015, respectively, and the Ph.D. degree from the University of New South Wales (UNSW), Canberra, Australia, in 2019. He joined the Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia, in 2020, where he is currently a Research Scientist. His research interests include remote sensing and machine learning, and their applications to environmental problems.



Nagur R. C. Cherukuru received the M.Sc. degree in Geophysics from Andhra University, India, in 1998, and the Ph.D. degree in Marine Optics and Remote Sensing from the University of Plymouth, Plymouth, U.K., in 2005. He joined the Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia, in 2006, where he is currently a Principal Research Scientist. His research interests include bio-optical modelling and remote sensing of water quality.



interests include remote sensing and environmental monitoring.

Eric A. Lehmann received the Dipl. El.-Ing. degree from the Swiss Federal Institute of Technology (ETH), Zurich, Switzerland, in 1999, and the M.Phil. and Ph.D. degrees from the Australian National University, Canberra, ACT, Australia, in 2000 and 2004, respectively. He was a Research Fellow with Western Australian Telecommunications Research Institute from 2005 to 2008. He joined the Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia, in 2008, where he is currently a Senior Research Scientist. His research



interests include remote sensing, geospatial artificial intelligence, and water quality monitoring.

S. L. Kesav Unnithan received the B.Tech. degree from Cochin University of Science and Technology, Kochi, India, in 2016, the M.Tech. degree from the Indian Institute of Space Science and Technology, Thiruvananthapuram, India, in 2018, and the Ph.D. degree from the Indian Institute of Technology Bombay, Mumbai, India, and Monash University, Melbourne, VIC, Australia, in 2023. He joined the Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia, in 2023, where he is currently a Research Scientist. His research



Tim J. Malthus received the B.Sc., Dip.Sci., and Ph.D. degrees from the University of Otago, Dunedin, New Zealand, in 1978, 1980, and 1986, respectively. He was a Senior Lecturer with the University of Edinburgh, Edinburgh, U.K. from 1994 to 2009. In 2009, he joined the Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia, where he is currently a Senior Principal Research Scientist. His research interests include hyperspectral sensing and water quality monitoring.



Gemma Kerrisk received the B.Sc. degree in Genetics from Massey University, Palmerston North, New Zealand, in 2017, and the Postgraduate Diploma in Environmental Management from the University of Auckland, Auckland, New Zealand, in 2018. She joined the Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia, in 2019, where she is currently an Experimental Scientist. Her research interests include aquatic remote sensing, optical oceanography, and water quality monitoring.



Xiubin Qi received the B.Sc. and M.Sc. degrees from Nanjing University, Nanjing, China, and the Ph.D. degree from the University of California, Davis, CA, USA. She joined the Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia, in 2008, where she is currently a Senior Research Scientist. Her research interests include aquatic sensing technologies, satellite calibration and validation, and water quality monitoring.



tal and engineering applications.

Faisal Islam received the B.Tech. and M.Tech. degrees from the Indian Institute of Technology Delhi, Kanpur, India, in 2014 and 2015, respectively, and the Ph.D. degree from Mines Paris – PSL, Paris, France, in 2020. He worked as a Research Fellow with the University of New South Wales, Sydney, NSW, Australia, before joining the Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia, in 2023 as a Research Scientist. His research interests include statistical modelling and artificial intelligence for environmental



Tisham Dhar received the B.Eng. degree from the University of Adelaide, Adelaide, SA, Australia, in 2006. He held engineering and software development roles in government agencies, industry, and startup companies before joining the Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia, in 2023, where he is currently a Senior Engineer. His research interests include geospatial software engineering, Earth observation data systems, and open-source geospatial technologies.



Mark Doubell received the Ph.D. degree in Fluorescence Microstructure and Plankton Ecology from Flinders University, Adelaide, SA, Australia, in 2007. He was a Postdoctoral Fellow with Tokyo University of Marine Science and Technology, Tokyo, Japan, from 2007 to 2009. He joined South Australian Research and Development Institute (SARDI), Adelaide, SA, Australia, in 2009, where he is currently a Principal Scientist. His research interests include aquatic remote sensing, optical oceanography, and phytoplankton ecology.