Satellite remote sensing of global phytoplankton primary production by integrating machine learning algorithm and photophysiological model

August 2022

ZHAOXIN LI

Graduate School of

Science and Engineering

CHIBA UNIVERSITY

(千葉大学審査学位論文)

Satellite remote sensing of global phytoplankton primary production by integrating machine learning algorithm and photophysiological model

August 2022

ZHAOXIN LI

Graduate School of

Science and Engineering

CHIBA UNIVERSITY

Contents

List of Figures

List of Tables

List of Abbreviations and Symbols

$AbPM_{AM96}$	Absorption-based IPP model proposed by Antoine and Morel (1996)
ADD	Algorithm development dataset
ANTARES4	ANTArtic RESearch No. 4 cruise
AVD	Algorithm validation dataset
BATS	Bermuda Atlantic Time-series Study
BGC-Argo	Biogeochemical-Argo
Bias	Logarithmic bias (%)
BIOSOPE	BIogeochmistry and Optics South Pacific Experiment
CCI	Climate Change Initiative
CO_2	Carbon dioxide
CPC	Climate Prediction Center
CV	Coefficient of variation (%)
CZCS	Coastal Zone Color Scanner
DCM	Deep chlorophyll maximum
DI	Depth-integrated
DL	Day length (h)
DR	Depth-resolved
ECS	East China Sea
EM	Empirical
ERFR	Enhanced random forest regression
ERSST	Extended reconstructed SST
ESA	European Space Agency
EUMELI4	Eutrophic, MEsotrophic and oLIgotrophic program No. 4 cruise
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
FRRF	Fast repetition rate fluorometry
G2B	Gilerson 2-Band algorithm proposed by Gilerson et al. (2010)
GloboLakes	Global Observatory of Lake Responses to Environmental Change
GODAS	Global Ocean Data Assimilation System
HOT	Hawaii Ocean Time-Series
HPLC	High-performance liquid chromatography
HUD	Expeditions carried out on board the CCGS (Canadian Coast Guard Ship) *Hudson* (Canada)
HYCOM	Hybrid Coordinate Ocean Model
IOP	Inherent optical property
IPP	Depth-integrated primary production
JS	Japan Sea
KEOPS	KErguelen: compared study of the Ocean and the Plateau in Surface water
Lidar	Light detection and ranging
LKSM	Lake Kasumigaura
LSWT	Lake surface water temperature (°C)
LUT	Look-up table

(Continued.)

LUT_RK	LUT method built here for the Rhode River and Lake Kasumigaura
LUT20	LUT method proposed by Kulk et al. (2020)
M96	P-E algorithm proposed by Morel et al. (1996)
MALINA	Malina oceanographic expedition
MAPPS	MArine primary Production: model Parameters from Space
MDN	Mixture Density Networks
MERIS	MEdium Resolution Imaging Spectrometer
MLR	Multivariate linear regression
MKT	Mann-Kendall trend test
MODIS	MODerate resolution Imaging Spectroradiometer
MSI	MultiSpectral Instrument
NAB2008	North Atlantic Bloom Experiment 2008
NASA	National Aeronautics and Space Administration
NBP0103	Nathaniel B. Palmer Cruise 01-03
NCEP	National Centers for Environmental Prediction
NDGI	Normalized Difference Chlorophyll Index
NOBM	NASA Ocean Biogeochemical Model
NSCS	Northern South China Sea
OC	Ocean color
OLCI	Ocean and Land Color Imager
OLIPAC	OLIgotrophy in PACific
ONI	Oceanic Niño index
OP	Ocean Productivity
P14	P-E algorithm proposed by Picart et al. (2014)
PACE	Plankton, Aerosol, Cloud, ocean Ecosystem
PAR	Photosynthetically available radiation
POLYMER	POLYnomial-based algorithm applied to MERIS
POMME	Programme Océan Multidisciplinaire Méso Echelle
PP	Primary production
PROSOPE	PROductivity of Oceanic PElagic Systems
PSL	Physical Sciences Laboratory
PUR	Photosynthetic usable radiation
QA	Quality assurance
QAA	Quasi-analytical algorithm
QAA_Xue	Modified QAA proposed by Xue et al. (2019)
QAA_Yang	Modified QAA proposed by Yang et al. (2013)
RFR	Random forest regression
RMSD	Root mean square difference
RRCB	Rhode River of the Chesapeake Bay
RSS	Remote Sensing Systems
S16	P-E algorithm proposed by Silsbe et al. (2016)
S3A/B	Sentinel-3A/B
SeaWiFS	Sea-viewing Wide Field-of-View Sensor

(Continued.)

SST	Sea surface temperature (°C)
SMMR	Scanning Multichannel Microwave Radiometer
SNPP	Suomi National Polar-orbiting Partnership
SSM/I	Special Sensor Microwave/Imager
SSMIS	Special Sensor Microwave Imager/Sounder
SVM	Support vector machine
TI	Time-integrated
TPM	Theory-based primary production model
TPM_{ERFR}	Integration of the ERFR and the TPM
TPM_{LUT20}	Integration of the LUT20 and the TPM
TPM_{LUT_RK}	Integration of the LUT_RK and the TPM
TR	Time-resolved
TSD	Time series decomposition
UAPD	Unbiased absolute percentage difference (%)
VGPM	Vertically generalized production model
$VGPM_{BF97}$	Original VGPM proposed by Behrenfeld and Falkowski (1997b)
$VGPM_{H02}$	Modified VGPM proposed by Harding et al. (2002)
$VGPM_{KI05}$	Modified VGPM proposed by Kameda and Ishizaka (2005)
VIIRS	Visible Infrared Imaging Radiometer Suite
WTS	Western Tasman Sea
Y3B	Yang 3-Band algorithm proposed by Yang et al. (2010)
$^{13}C/^{14}C$	Carbon-13/14 (isotopes of carbon)
a	total absorption coefficient of water (m^{-1})
a_{ph}	Phytoplankton absorption coefficient (m^{-1})
$a_{ph}(443)$	a_{ph} at 443nm (m^{-1})
$\bar{a}_{ph}$	Spectrally averaged a_{ph} over 400–700 nm (m^{-1})
$\bar{a}^{*}_{ph}$	Chla-specific $\bar{a}_{ph}$ (m^2 mg^{-1})
b_b	total backscattering coefficient of water (m^{-1})
Chl*a*	Chlorophyll-*a* concentration (mg m^{-3})
*d*PAR	Daily PAR (mol photons m^{-2} d^{-1})
E_k	Light saturation parameter of the P-E curve (μmol photons m^{-2} s^{-1})
K_d(PAR)	Diffuse attenuation coefficient of PAR (m^{-1})
K_φ	half saturation parameter
$O_2/^{18}O_2$	Molecular oxygen-16/18 (isotopes of oxygen)
PAR(0^-)	PAR just beneath the water surface
P^B_{max}	Assimilation number of the P-E curve (mg C (mg Chl)$^{-1}$ h^{-1})
P^B_{opt}	Maximum carbon fixation rate of a water column (mg C (mg Chl)$^{-1}$ h^{-1})
P-E	Photosynthesis-irradiance (E)
r	Correlation coefficient
r_{PAR}	Light transmittance of downwelling PAR (%)
R^2	Determination coefficient
R_{rs}	Remote sensing reflectance (sr^{-1})
S1–S4	Four training scenarios for the development of ERFR

(Continued.)

T	Water temperature (°C)
z	Depth (m)
Z_{eu}	Euphotic zone depth (m)
Z_m	Mixed-layer depth (m)
α^B	Initial slope of the P-E curve (mg C (mg Chl)$^{-1}$ h^{-1} (μmol photons m^{-2} s^{-1})$^{-1}$)
μ	Phytoplankton growth rate
δ	Relative error (%)
θ	Carbon-to-chlorophyll ratio (g C (g Chl) $^{-1}$)
λ	Wavelength (nm)
φ_{max}	Maximum quantum yield (mol C (mol photons)$^{-1}$)

Abstract

Phytoplankton primary production (PP) is a vital indicator of the carbon uptake in aquatic ecosystems. Conventional field-based measurement of PP has made great progress in improving the accuracy of measured photosynthetic rate of phytoplankton in water samples. With the advent of the era of satellite remote sensing and the accumulation of reliable *in situ* PP measurements, remote sensing has long been recognized as the only feasible approach to obtain synoptic observations of PP. As a result, the modeling of PP based on remote sensing has received much attention from the scientific community since the 1980s, and it has been widely used to investigate the response of aquatic ecosystems to global climate change nowadays.

Although a variety of satellite-based primary production model has been designed for either regional or global scales with different concepts and model structures, the total amounts of annual PP at basin and global scales estimated from space are still controversial due to the variation of model performances. The uncertainty in satellite-based PP data could often lead to different spatiotemporal variabilities of PP for different regions of interest, and consequently, there remains a debate about characteristics of interannual changes in basin-scale and global PP and underlying mechanisms associated with relevant climate or environmental drivers. Therefore, it is necessary to improve the accuracy and applicability of remote estimation of PP for different types of water bodies. The theory-based primary production model (TPM), one of the earlier proposed photophysiological

models, is potentially applicable to a variety of water bodies because it well formulates the photophysiological property of phytoplankton with a semi-analytical nature. Its performance is highly dependent upon the reliability of two photosynthetic parameters, i.e., the assimilation number (P^{B}_{max}) and the light saturation parameter (E_k). Unfortunately, the remote assignment of P^{B}_{max} and E_k is acknowledged to be a challenging task, and the limited progress has impeded extensive use of the TPM.

To meet the need and to address the challenge stated above, first, a machine learning algorithm called the enhanced random forest regression (ERFR) was adapted to retrieve P^{B}_{max} and E_k from satellite observations. The ERFR was trained and validated using *in situ* datasets from a broad range of trophic and biogeographic conditions, covering oceanic, coastal, and inland water bodies. Evaluations with independent *in situ* data and matchup data showed that the ERFR outperformed conventional empirical and semi-analytical algorithms, and it could better capture the variability of P^{B}_{max} and E_k than look-up-table methods. Second, the ERFR were integrated with the TPM (together termed as TPM_{ERFR}), as one well generalized model, to estimate daily depth-integrated PP (termed as IPP) in clear to turbid waters. The root mean square difference (*RMSD*) of the satellite-based IPP estimates from the TPM_{ERFR} remained below 0.25 for various types of water bodies. In contrast, the benchmark models generally yielded IPP estimates with *RMSD*s of 0.27–0.62. The results substantiated the superior performance and applicability of the proposed TPM_{ERFR} at global scale. Third, the TPM_{ERFR} was

implemented to global monthly satellite products for the last 24 years (1998–2021) to generate long-term IPP product for the variability analysis. Bias-correction for multi-source input satellite products and gap-filling for the missing data in the generated IPP product were considered to reduce potential uncertainty and ensure data reliability. Additional IPP products generated by different methods were introduced for the inter-comparisons against the proposed TPM_{ERFR}. Our analyses suggested that reasonable spatial distributions and seasonal variations of IPP at basin and global scales were demonstrated based on the TPM_{ERFR}. The global climatological annual IPP was estimated to be 43.7 Pg C yr^{-1}. The interannual variabilities of IPP in the global, Pacific, and Indian oceans are greatly attributed to the transition between El Niño and La Niña events, while the reduction and expansion of sea ice cover have an impact on the changes of IPP in the Antarctic and Arctic oceans. A significant increasing trend of IPP is observed in the Atlantic Ocean, which could be partially explained by the interannual changes of Chl*a*.

The results of estimation assessments and variability analyses in this study indicated that the TPM_{ERFR} could be an encouraging alternative to seamlessly generate IPP products at global scale (from clear to turbid waters) and has the potential to provide valuable and reliable insights into spatiotemporal dynamics of IPP for over two decades.

Keywords: Depth-integrated primary production; Photosynthetic parameters; Phytoplankton photophysiology; Enhanced random forest regression; Ocean color remote sensing; Interannual variability.

Chapter 1. Introduction

1.1. Significance of phytoplankton primary production

Phytoplankton are responsible for primary production (PP) in worldwide aquatic ecosystems through the biochemical process of photosynthesis, and thus serve as the footstone of aquatic food webs and make the vast ocean an essential role in carbon sequestration. As a vital component of the global carbon cycle, the carbon fixation by oceanic phytoplankton is profoundly related to the envrionmental drivers of thermohaline circulation, climatic oscillations, surface wind force, and the solar cycle, etc (Behrenfeld et al., 2006; Di Lorenzo et al., 2008; Falkowski et al., 1998; Martinez et al., 2009). These physical processes regulate the water temperature, nutrient flux, and availability of incident light in the upper ocean where abundant phytoplankton conduct photosynthesis. The natural variation in ocean PP is implicitly coupled with climate-induced ocean changes (Bindoff et al., 2019), because phytoplankton photosynthesis enables the regulation of ocean-atmosphere carbon dioxide (CO_2) exchange and the biological carbon pump (Dunne et al., 2007; Falkowski et al., 2000; Palevsky et al., 2016; Prentice et al., 2001). In this context, synoptic and accurate monitoring of ocean PP plays a crucial role in understanding the responses of aquatic ecosystems to climate change and its side effects on water environments.

Plentiful previous works have stated a number of arresting evidences that underline the important role of inland freshwater ecosystems, such as lakes, rivers,

and reservoirs, in many aspects of the carbon cycle associated with human activites (Cole et al., 2007; Dean et al., 1998; Raymond et al., 2013). Although global inland waters only cover a small fraction of the planet's surface, some of these aquatic ecosystems can be very productive and be extremely active in regional carbon exchange (Likens, 1975; Tranvik et al., 2009). In contrast to open oceans, the emission and uptake of atmospheric CO_2 by inland freshwaters are fundamentally linked with anthropogenic factors, such as agricultural fertilizer, industrial wastes, atmospheric pollution, and changes in land use and land cover (Regnier et al., 2013). Therefore, a comprehensive understanding of PP variations in inland water environments is desired in order to gain further insight into regional carbon budgets, which will also give aid to the achievement of carbon neutrality in the coming decades, especially for the countries that are rich in freshwater resources.

Conventional *in situ* methods to obtain precise measurements of PP, such as the assimilation of inorganic carbon isotopes ($^{13}C/^{14}C$) (Hama et al., 1983; Nielsen, 1952) and the evolution of oxygen ($O_2/^{18}O_2$) (Bender et al., 1987; Marra, 2002), have been used for decades. Nevertheless, it is difficult to fully capture the heterogenity of spatial distributions of PP based on field surveys due to the limited sampling frequency and spatial coverage of *in situ* measurements. In consequence, most of open oceans are still undersampled, especially for the Pacific and Indian oceans, and the coverage of *in situ* measurements for inland water areas remains to be improved, leaving a huge data gap in those enclosed water areas. Fortunately,

with the advent of the satellite remote sensing era, it has become possible to acquire synoptic observations of ocean PP at different spatiotemporal scales from a series of ocean color sensors, such as the Coastal Zone Color Scanner (CZCS), Sea-viewing Wide Field-of-View Sensor (SeaWiFS), and MODerate resolution Imaging Spectroradiometer (MODIS) (Eppley et al., 1985; Hirawake et al., 2012; Longhurst et al., 1995; Platt and Sathyendranath, 1988; Son et al., 2005; Tilstone et al., 2014). Besides, given the significance of monitoring phytoplankton carbon fixation of freshwater ecosystems, previous attempts have also been dedicated to remotely estimate PP in inland freshwater areas based on the MEdium Resolution Imaging Spectrometer (MERIS) (Kauer et al., 2015; Son et al., 2014; Soomets et al., 2019).

Chavez et al. (2011) have systematically discussed the sensitivity of global ocean PP to climate changes and variability by using long-term *in situ* and satellite time series data. A close relationship between the interannual variability of ocean PP and climate-driven ocean-atmospheric oscillations. Brewin et al. (2021) have demonstrated the necessity and advantages of using satellite remote sensing to monitor the ocean biological carbon pump, which is vital to assess the changes in ocean PP and then in global carbon cycle. On the other side, Sayers et al. (2020) have investigated the contributions of carbon uptake from eleven of the world's largest lakes in the period 2003–2018 based on satellite data, and only three of them has proved to have significant trends associated with climate change. Their later work (Sayers et al., 2021) has provided a synoptic estimate of carbon fixation

in global freshwater lakes using remote sensing observations, offering new insights into the carbon budget in freshwater ecosystems. However, as they emphasized in the study, their method might not be able to accurately estimate lake-wide PP, implying that the reported results are promising but still need future validations.

Desipte the fact that a great number of works have been dedicated to monitor phytoplankton PP based on satellite remote sensing, including marine and freshwater ecosystems, the estimated total amounts of PP and relevant interannual trends are of low confidence (Gulev et al., 2021). It implies that a generalized approach is an urgent need to seamlessly estimate PP in global waters from space with the capability to accurately capture subtle patterns at different scales intead of a rough estimation only. Utimately, this approach is expected to promote more reasonable estimates of total PP so as to assure the rationality of the trends detected from the temporal changes of PP.

1.2. Progress in modeling primary production from space

Recognition of the role of the photophysiological traits of phytoplankton in the modeling of PP has led to a growing consensus that information on the concentration of chlorophyll-*a* (Chl*a*), photosynthetically available rdiation (PAR), and the parameters that indicate photosynthetic rates are required components for an accurate modeling (Brewin et al., 2021; Platt et al., 2017). Though models have been established based on diverse concepts and complexities (Behrenfeld and Falkowski, 1997a; Lee et al., 2015), it has been acknowledged

that these variants basically share a common theory of the photophysiological traits of phytoplankton. The theory has been parameterized as a semi-analytical model, known as the photosynthesis-irradiance (P-E) curve. The P-E curve explicitly depicts the photophysiological response of phytoplankton to available light (Sathyendranath et al., 2020; Sathyendranath and Platt, 2007), which implies that it is a well generalized concept that can be applicable to model phytoplankton growth in a variety of water bodies. In this study, this photophysiological model is named the theory-based primary production model (TPM) for simplicity. Mathematically, the TPM describes phytoplankton PP as a function of depth (z) and time (t), as follows:

$$\begin{aligned} \mathrm{PP}(z,\ t) &= P_{\max}(z) \times f\left[\mathrm{PAR}(z,\ t),\ E_{\mathrm{k}}(z)\right] \\ &= \mathrm{Chl}a(z) \times P^{\mathrm{B}}_{\max}(z) \tanh\left[\frac{\mathrm{PAR}(z,\ t)}{E_{\mathrm{k}}(z)}\right], \end{aligned} \tag{1.1}$$

where $P_{\max}$ refers to the maximum carbon fixation rate, and f is the light regulation function of the P-E curve, which is formulated by the hyperbolic tangent function because of its recognized applicability (Gallegos, 1992; Harrison and Platt, 1986; Platt and Jassby, 1976; Silsbe et al., 2016). $P^{\mathrm{B}}_{\max}$ is the assimilation number, and E_{k} is the light saturation parameter, both of which are known as photosynthetic parameters of the P-E curve (or P-E parameters for short).

Given the fact that the ocean color products of Chl$a$ and PAR have become increasingly more accurate over a wide range of water bodies (Frouin et al., 2018; Hu et al., 2012; Pahlevan et al., 2020), the TPM can be practically used to estimate

daily depth-integrated PP (IPP) from space if there is a sound strategy for the remote assignment of P^{B}_{max} and E_k. However, previous works have concluded that the TPM-like models do not always perform better than simpler models and their accuracy varies between regions (Campbell et al., 2002; Carr et al., 2006; Friedrichs et al., 2009; Saba et al., 2010, 2011); for example, it is common that most ocean-color models perform poorly in coastal and shallow (possibly Case-II) waters (Saba et al., 2011). One of the reasons for the variable performances of current IPP models is the imperfect estimation of photosynthetic parameters or their analogues depending on specific model structures. In other word, the uncertainty in satellite-based IPP data heavily depends on to what extent the photosynthetic parameters of a certain water region of interest (such as open oceans, coastal waters, and inland lakes) can be accurately retrieved. For this reason, the refinement of assigning these kinds of parameters from satellites has long been emphasized (Behrenfeld and Falkowski, 1997a; Campbell et al., 2002; Carr et al., 2006).

Fundamental parameters or variants for modeling IPP that are analogous to photosynthetic parameters can exist in different contexts with different definitions and notations, such as the maximum carbon fixation rate in a water column (P^{B}_{opt}) for the vertically generalized production model (VGPM) (Behrenfeld and Falkowski, 1997b), the phytoplankton growth rate (μ) (Behrenfeld et al., 2005), and the pair of the maximum quantum yield (φ_{max}) and the half saturation parameter (K_φ) used in the context of phytoplankton absorption-based models

(Antoine and Morel, 1996; Kiefer and Mitchell, 1983; Lee et al., 1996; Marra et al., 2003; Silsbe et al., 2016). Nowadays, unfortunately, the remote assignment of photosynthetic parameters (or their analogues/variants) in global waters remains a challenging and non-trivial goal.

On the one hand, remote estimation of some of analogues has a long history due to their convenience. For instance, the simplest approach to remotely estimate $P_{\mathrm{opt}}^{\mathrm{B}}$ was to bulid a linkage with sea surface temperature (SST) (Eppley, 1972; Behrenfeld and Falkowski, 1997b). Several modified versions have also been proposed, which consider the inclusion of Chl*a* (Kameda and Ishizaka, 2005) and phytoplankton absorption coefficient (Hirawake et al., 2012, 2011) for a better performance in different target water regions. It has also reported that the adjustment of the empirical funciton between $P_{\mathrm{opt}}^{\mathrm{B}}$ and SST using local field data could yield more accurate IPP estimates in the Chesapeake Bay (Harding et al., 2002; Son et al., 2014). In addition, as a practical strategy, a constant $P_{\mathrm{opt}}^{\mathrm{B}}$ determined by the local measurements for a small area may be representative enough to estimate realistic IPP values (Tripathy et al., 2012). In contrast, the remote assignment of μ is much more complex, which takes the growth irradiance into account and thus requires the input of Chl*a*, phytoplankton carbon, PAR, diffuse attenuation coefficient of PAR (K_{d}(PAR)), mixed-layer depth (Z_{m}), and nitracline depth of a water column (Westberry et al., 2008). As a result, it is difficult to be implemented in coastal and inland shallow waters. In terms of using the absorption based model to simulate phytoplankon carbon fixation (Lee et al.,

2015), φ_{max} and K_{φ} were usually assigned as regional or global constants based on the analysis of experimental data (Lee et al., 2011; Lobanova et al., 2018). Considering the inherent variability of these two parameters, Zoffoli et al. (2018) (Zoffoli et al., 2018) have tried to associate φ_{max} and K_{φ} with satellite-measured SST and PAR, respectively, based on the long-term observations from two stations: Hawaii Ocean Time-Series (HOT) and Bermuda Atlantic Time-series Study (BATS). Their results suggested that the relationship between φ_{max} and SST at HOT was inconsistent with that at BATS. Regional studies about retrieving the quantum yield using a piecewise function of Chl*a*, PAR, and K_d(PAR) for several inland lakes have also been reported (Arst et al., 2012, 2008; Kauer et al., 2015), but the proposed methods were highly localized and thus not easy to be directly adopted for other regions.

On the other hand, several earlier attempts have also been made to estimate photosynthetic parameters from other environmental parameters that can be routinely obtained from satellite data. For instance, simple regression analysis has been used to quantify the relationship between P^B_{max} and water temperature (T) or Chl*a* (Bouman et al., 2005; Eppley, 1972; Morel, 1991). Picart et al. (2014) have combined inversion algorithms of the carbon-to-chlorophyll ratio (θ) and maximum growth rate of phytoplankton to calculate P^B_{max} using Chl*a*- and nutrient-based schemes. Behrenfeld et al. (2016) have delineated the change of E_k in global oceans using satellite-measured daily PAR and K_d(PAR). An alternative approach is to use *in situ* information to directly assign photosynthetic parameters

on a pixel-by-pixel basis with the aid of historical databases (Bouman et al., 2018; Kulk et al., 2020). This kind of approach assumes that averaged P^{B}_{max} and E_k are representative enough in a distinct region and a specific season, which has been operationally used to estimate IPP in the global ocean (Kulk et al., 2020; Longhurst et al., 1995; Platt and Sathyendranath, 1999) and in Lake Tanganyika (Africa) (Bergamino et al., 2010). In addition, a more dynamic strategy for the assignment has been proposed to use a nearest-neighbor method that considers the ambient characteristics of Chl*a* and SST (Platt et al., 2008).

No study has yet demonstrated generic effectiveness for the remote assignment of P^{B}_{max} and E_k, especially for turbid waters where even knowledge of *in situ* information on these parameters is rarely understood. Therefore, a broadly applicable method and a comprehensive assessment of existing algorithms are still required. To capture the natural variability of P^{B}_{max} and E_k in different aquatic ecosystems, machine learning algorithm is a promising tool because of its efficiency in data interrogation. For example, Tang et al. (2008) have indicated that the support vector machine (SVM) performs better in oceanic waters than the original VGPM with the same input parameters. In addition, a recent effort has been dedicated to estimate global ocean gross PP from satellite observations based on the random forest regression (RFR) (Huang et al., 2021), which demonstrates the encouraging utility of machine learning algorithms. More importantly, a novel machine learning algorithm, the enhanced random forest regression (ERFR), has recently been designed to improve the extrapolation

capability of traditional RFRs (Zhang et al., 2019). Applications of this state-of-the-art machine learning algorithm may provide a new opportunity to address the challenge of remotely assigning photosynthetic parameters by leveraging data from clear to turbid waters. Eventually, satellite-based estimation of global IPP is expected to be improved by supplying refined retrievals of photosynthetic parameters, promoting the monitoring and investigation of spatiotemporal variability of IPP in global waters.

1.3. Research objectives and structures

Given the current challenge of remote estimation of phytoplankton IPP in various types of water bodies and the demand of reducing the uncertainty in current satellite-based IPP data for a better understanding of interannual or climatological variabilities of global IPP, we need a new strategy and the power of machine learning to enhance the modeling of IPP from space. The objective of this study is consequently to first refine the remote assignment of $P^{\mathrm{B}}_{\mathrm{max}}$ and E_{k} based on the ERFR and subsequently to improve the remote estimation of IPP in clear to turbid waters by integrating the ERFR and the TPM (i.e., the TPM$_{ERFR}$). The performances of the ERFR and TPM$_{ERFR}$ for different water types (clear to turbid) are evaluated by the comparisons with other existing algorithms based on independent *in situ* datasets and satellite matchup data. The sensitivity of the TPM$_{ERFR}$ to satellite-measured input parameters is also analyzed to quantify the impact of the uncertainty in model input on the retrievals of $P^{\mathrm{B}}_{\mathrm{max}}$, E_{k}, and IPP. Before the implementation of the TPM$_{ERFR}$ for product generation, the potential

bias in multi-source satellite products is investigated and then corrected to reduce the uncertainty in the generated IPP product. In addition, the procedure of gap-filling is employed to improve the data coverage of our IPP product. Global monthly IPP product generated by the TPM_{ERFR} in the period 1998–2021 is used to reassess annual IPP at basin and global scales, and the characteristics of spatiotemporal variability of basin-scale and global IPP are discussed to illustrate the underlying mechanisms that explain the interannual changes of IPP.

The overall workflow of this study is presented in Figure 1.1. In summary, the concise research objectives of this study are as follows:

(1). To improve the remote estimation of phytoplankton IPP in clear to turbid waters by the integration of a novel machine leanring algorithm, the ERFR, and a theory-based photophysiological model, the TPM (together termed as the TPM_{ERFR});

(2). To generate high-quality and gap-free monthly IPP product from long-term and multi-source satellite data based on the proposed TPM_{ERFR} for the last 24 years between 1998 and 2021;

(3). To comprehensively analyze the spatial and temporal (including seasonal and interannual changes) variabilities of IPP at both basin and global scales, and subsequently to unravel the underlying mechanisms associated with climate- and environment-related drivers.

This study is organized as follows:

In Chapter 2. Materials and data processing, the acquisition of algorithm development and validation datasets are introduced, including the data sources, *in situ* measured parameters and related measurement methods. Multi-source satellite and auxiliary products used for model evaluation and product generation as well as additional data used for subsequent inter-comparisons and analyses are also introduced. The detailed description of the procedures of data processing is provided, including retrievals of required ocean color parameters and bias-correction for multi-source products.

In Chapter 3. Methodology, the flow of photosynthetic parameter estimation based on the ERFR and IPP computation based on the TPM are presented, along with the information of existing algorithms for inter-comparisons, and then the procedures of long-term product generation are provided. The methods used to investigate model sensitivity and to analyze time series data are also introduced.

In Chapter 4. Model evaluation, the results of the systematical evaluations of the proposed TPM_{ERFR} are presented, including its performance in oceanic and inland waters in the cases of taking *in situ* and satellite data as input, and the comparisons against other existing methods. The application of the TPM_{ERFR} on mapping IPP distributions in inland areas is demonstrated. The sensitivity, source of error, and future perspectives of the TPM_{ERFR} are also discussed.

In Chapter 5. Spatiotemporal avariability analysis, the rationality of the TPM_{ERFR}-based IPP distributions and estimates of annual IPP at basin and global scales is fisrt assessed and then verified. The comprehensive analysis of seasonal

and interannual variations of IPP in global oceans and individual ocean basins is presented, followed by the discussion of the underlying mechanisms for the revealed changes in IPP associated with climate and envrionmental variables.

In Chapter 6. Conclusions, the main results and conclusions are summarized, including the results of model evaluation and product assessment, the typical spatial and seasonal patterns of IPP, and the interannual variability of IPP at basin and global scales along with the interpretation of mechanisms.

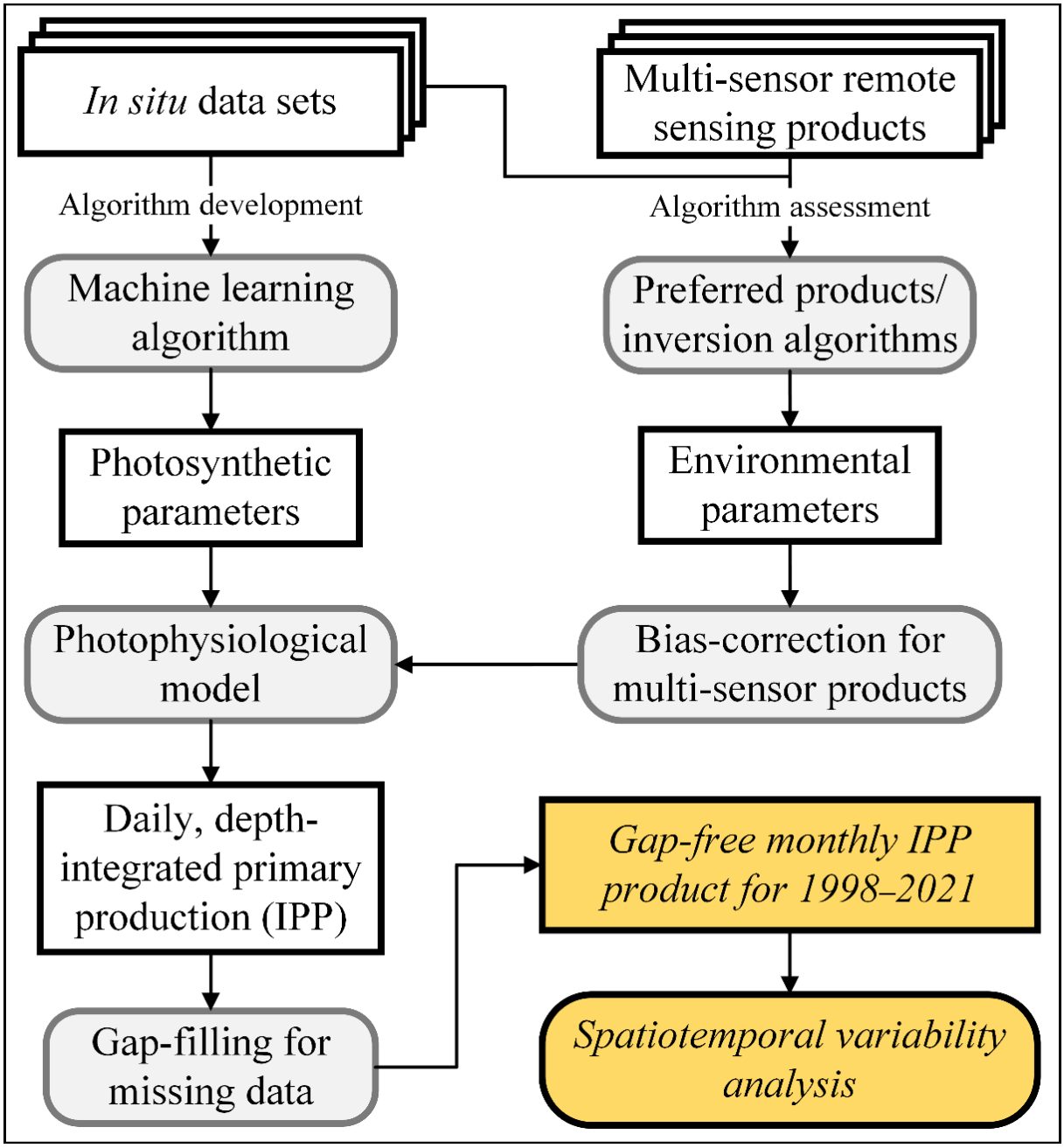

Figure 1.1. Overall workflow of this study, including the development of machine learning algorithm, the assessment and selection of preferred ocean color inversion algorithms, the model evaluation and application to generate gap-free monthly IPP product in the period 1998−2021, and the spatiotemporal variability analysis.

Chapter 2. Materials and data processing

2.1. Field and auxiliary data

2.1.1. Dataset acquisition

A total of 18 datasets from various representative water bodies around the world (see Figure 2.1) were compiled from several data sources, including LEFE-CYBER (Les Enveloppes Fluides et l'Environnement-CYcle Biogéochimiques, Environnement et Resources), PANGAEA (Data Publisher for Earth & Environmental Science), BCO-DMO (Biological & Chemical Oceanography Data Management Office), and NIES (National Institute for Environmental Studies) (Table 2.1 and Table 2.2). Fourteen of these datasets that have more complete measurements of photosynthetic and environmental parameters were used for the development of the ERFR, and are hereafter denoted as algorithm development datasets (ADDs). These ADDs were collected from a variety of environments with varying physical and biogeochemical charateristics, including ocean gyres, polar and marginal seas, and an estuary (the Rhode River on the western shore of the Chesapeake Bay). The other four datasets were used for independent algorithm validation and hereafter referred to as algorithm validation datasets (AVDs). The majority of the AVDs came from a synthesized global database (MAPPS) (Bouman et al., 2018) and a long-term monitoring database from Lake Kasumigaura, Japan (LKSM) (Takamura and Nakagawa, 2016). Most of the AVDs originated from water bodies different from those in the ADDs.

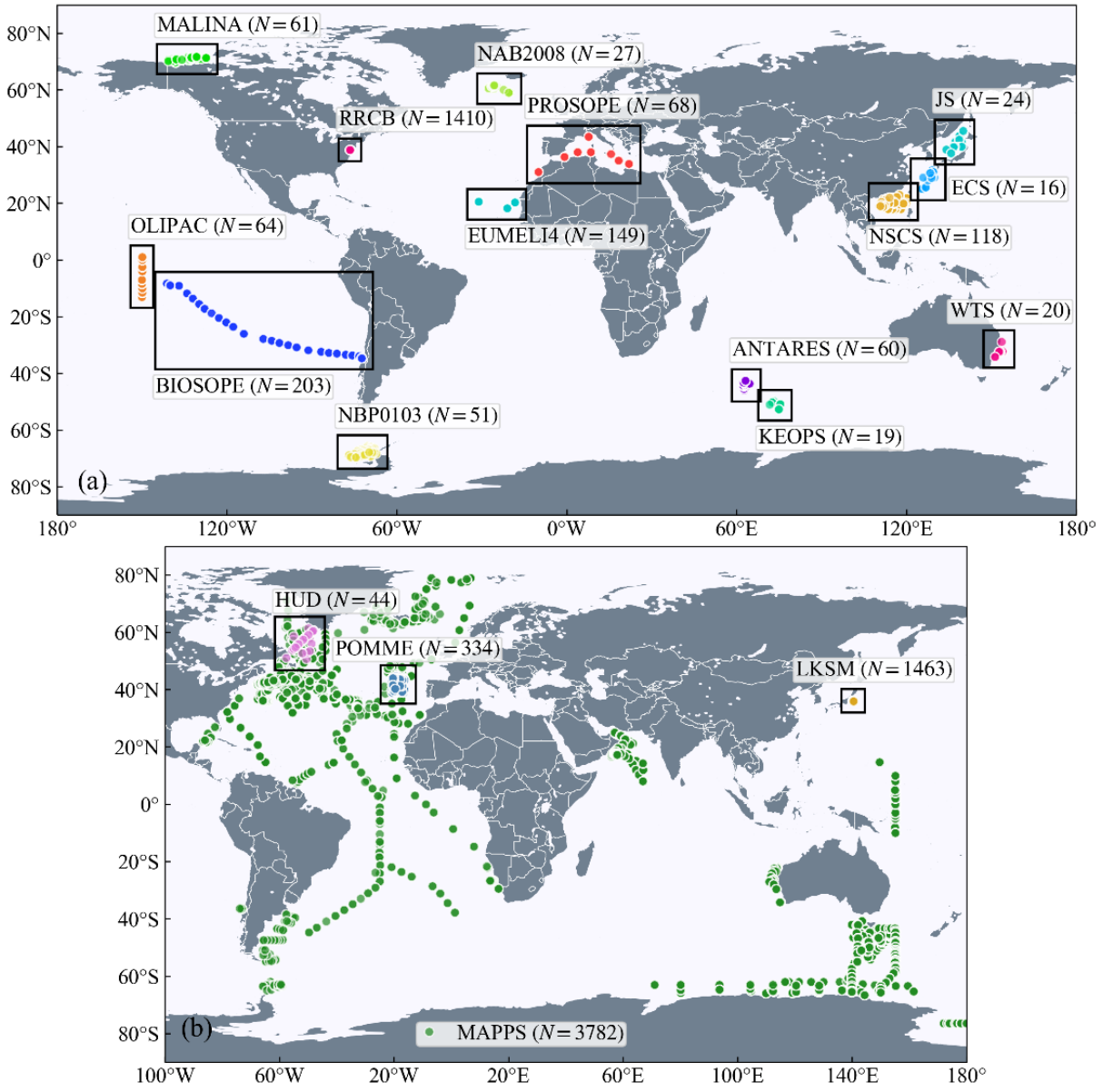

Figure 2.1. Spatial distributions of the compiled ADDs (a) and AVDs (b). The sampling sites of individual datasets are shown by colored circles. The corresponding sampling regions are outlined by black boxes, except for the MAPPS dataset. *N* represents the number of discrete data samples. Black dashed boxes indicate the locations of valid IPP matchups.

2.1.2. Measurement description

The compiled datasets contain *in situ* measured vertical profiles of photosynthetic, bio-optical and physical parameters with numbers of sampling depths that varied between individual projects/cruises. The exception were the RRCB and LKSM, where the data were sampled at only one depth at each site because the waters were assumed to be well mixed. Table 2.1 shows sampling regions, years, ranges of sampling depth, and respective references for each dataset. Table 2.2 gives the detailed information on data sources and corresponding links of database website or the plublished references for the used supplementary data. Table 2.3 provides the main available parameters (measured, recovered, or matched with satellite data) and measurement methods.

Typically, P^{B}_{max}, E_k, and the initial slope of the P-E curve (α^B) defined as the ratio of P^{B}_{max} to E_k were determined by short-term $^{13}C/^{14}C$ incubations, known as P-E experiments (Hama et al., 1983; Nielsen, 1952). During the experiments, each water sample was dispensed equally into several incubation chambers, and then were incubated at a given range of irradiances while water temperature was maintained as *in situ* temperature, which was regulated by water circulation (Lewis and Smith, 1983; Mackey et al., 1995; Marcel et al., 1994). The incubation times varied from 0.5 to 4 h, which depended on the corresponding project or cruise. One of the following three equations: the photoinhibition function (Platt et al., 1980), the hyperbolic tangent (Jassby and Platt, 1976), or the equation of Smith (1936), were used for the fitting of P-E curves.

Table 2.1. Summary of the complied datasets (ADDs and AVDs) used in this study. Abbreviation, sampling region, year, and depth for each of projects/cruises are given.

No.	Project/Cruise	Region (year)	Sampling depth	Reference
Algorithm Development Datasets (ADDs, N = 2,290)				
1	EUMELI4	Northeast Tropical Atlantic (1992)	2.2–159.0 m	Babin et al., 1996
2	OLIPAC	Equatorial Pacific (1994)	5.0–140.0 m	Uitz et al., 2008
3	ANTARES4	Southern Indian Ocean (1999)	5.2–124.5 m	Sedwick et al., 2002
4	PROSOPE	Mediterranean Sea (1999)	1.7–114.5 m	Claustre et al., 2020
5	JS	Japan Sea (1999)	10.0–40.0 m	Yoshikawa T. and Furuya K., 2008
6	ECS	East China Sea (2000)	10.0–75.0 m	Yoshikawa T. and Furuya K., 2008
7	NBP0103	Western Antarctic Peninsula (2001)	4.7–20.0 m	Perry et al., 2018
8	BIOSOPE	Subtropical South Pacific (2004)	2.4–221.6 m	Claustre et al., 2008
9	KEOPS	Southern Ocean (2005)	8.0–123.0 m	Uitz et al., 2009
10	NAB2008	North Atlantic (2008)	9.7–30.9 m	Vernet et al., 2004
11	MALINA	Beaufort Sea (2009)	2.9–80.1 m	Massicotte et al., 2009
12	WTS	Western Tasman Sea (2010)	3.0–52.0 m	Van Dongen-Vogels et al., 2015
13	NSCS	Northern South China Sea (2010–2012)	0.0–85.0 m	Xie et al., 2015
14	RRCB	Rhode River on the Chesapeake Bay (1990–2009)	0.1–3.0 m	Gallegos et al., 2012 and 2013
Algorithm Validation Datasets (AVDs, N = 5,623)				
1	POMME	North Atlantic (2001)	3.9–198.5 m	Claustre et al., 2005
2	HUD	Labrador Sea (2006–2008)	2.3–3.2 m	Fragoso et al., 2017
3	MAPPS	Multiple water areas (1997–2007, 2010, 2012, 2013)	0.0–142.0 m	Bouman et al., 2018
4	LKSM	Lake Kasumigaura (1981–2016)	0.0–2.0 m	Takamura et al., 2016

Table 2.2. Detailed information about data sources, links of database website, and references for the supplementary data used in this study.

Dataset	Source	Link/Reference
Algorithm development datasets (ADDs)		
EUMELI4	LEFE-CYBER	http://www.obs-vlfr.fr/proof/
OLIPAC	LEFE-CYBER	https://doi.org/10.17600/94010090
ANTARES4	LEFE-CYBER	https://doi.org/10.17600/99200010
PROSOPE	LEFE-CYBER	https://doi.org/10.17882/71723
JS	Personal communication	Yoshikawa T. and Furuya K., 2008
ECS	Personal communication	
NBP0103	BCO-DMO	https://www.bco-dmo.org/dataset/2375
BIOSOPE	LEFE-CYBER	http://www.obs-vlfr.fr/proof/
KEOPS	LEFE-CYBER	http://www.obs-vlfr.fr/proof/
NAB2008	BCO-DMO	https://www.bco-dmo.org/dataset/746215
MALINA	LEFE-CYBER	https://doi.org/10.17882/75345
WTS	PANGAEA	https://doi.pangaea.de/10.1594/PANGAEA.843554
NSCS	Supplementary data (see reference)	Xie et al., 2015
RRCB	PANGAEA	https://doi.pangaea.de/10.1594/PANGAEA.816494; https://doi.pangaea.de/10.1594/PANGAEA.819825
Algorithm validation datasets (AVDs)		
POMME	LEFE-CYBER	http://www.obs-vlfr.fr/proof/
HUD	PANGAEA	https://doi.org/10.1594/PANGAEA.871872
MAPPS	PANGAEA	https://doi.org/10.1594/PANGAEA.874087
LKSM	NIES	https://db.cger.nies.go.jp/gem/moni-e/inter/GEMS/database/kasumi/index.html

Chl*a* measurements were determined by one of the three methods: high-performance liquid chromatography (HPLC) (Ras et al., 2008), fluorometry (Holm-Hansen et al., 1965), and spectrophotometry. Spectra of phytoplankton absorption coefficient, $a_{ph}(\lambda)$, were obtained by the spectrophotometric method (Bricaud and Stramski, 1990; Kishino et al., 1985). Vertical profiles of PAR were measured by integrating the spectral downwelling irradiance, $E_d(\lambda)$, in the visible domain (400–700 nm), as follows:

$$\mathrm{PAR}(z) = \int_{400\ \mathrm{nm}}^{700\ \mathrm{nm}} E_d(\lambda,\ z)\mathrm{d}\lambda, \tag{2.1}$$

where z is water depth. K_d(PAR) was equated to the slope of a linear regression between the natural logarithm of PAR(z) and depth:

$$\mathrm{PAR}(z) = \mathrm{PAR}(0^-)\exp\left[-K_d(\mathrm{PAR})\times z\right], \tag{2.2}$$

where PAR(0^-) is the PAR just beneath the water surface. The light transmittance of downwelling PAR, r_{PAR}, at a given depth was equated to PAR(z)/PAR(0^-). The euphotic zone depth (Z_{eu}) is defined to be the depth where r_{PAR} is 1% and can be estimated as 4.605/K_d(PAR).

Measurements of Z_m were determined from *in situ T* profiles based on the criterion that the water density at Z_m was 0.125 kg m^{-3} greater than the density at 10 m (Levitus, 1982). Though the Z_m data obtained under this definition would commonly depict a thicker mixed-layer than the one determined by the criterion of temperature gradient, this kind of difference in Z_m data has negligible impact on the calculation of IPP.

Table 2.3. Summary of the main available parameters, methods to measured Chl*a* and a_{ph} data, incubation times of P-E experiments, and P-E functions used to fit P-E curves for each dataset. The parameters are also listed with different marks, which represent whether the parameter was measured *in situ*, recovered by post-processing, or matched with satellite data.

Dataset	Main parameters [1]							Measurement methods		Incubation length (h)	P-E function [4]
	P^B_{max}	E_k	Chl*a*	Z_{eu}	a_{ph}	T	*d*PAR	Chl*a* [2]	a_{ph} [3]		
Algorithm development datasets (ADDs)											
EUMELI4	○	○	○	○	◊	○●	◊	HPLC	/	1–2	H
OLIPAC	○	○	○	○	○	○●	◊	HPLC	BS90	1–2	H
ANTARES4	○	○	○●	○●	◊●	○●	◊●	FLUO	/	1	P
PROSOPE	○	○	○●	○●	◊●	○●	◊●	HPLC	/	2	H
JS	○	○	○●	○●	○●	○●	○●	FLUO	K85	0.5	H
ECS	○	○	○●	○●	○●	○●	○●	FLUO	K85	2	H
NBP0103	○	○	○	○	◊	○●	◊●	FLUO	/	1	P
BIOSOPE	○	○	○●	○●	○●	○●	◊●	HPLC	BS90	2–3.5	H
KEOPS	○	○	○	○	◊	○●	◊●	HPLC	/	1	P
NAB2008	○	○	○●	○●	◊●	○●	◊●	FLUO	/	2	E
MALINA	○	○	○●	○●	○●	○●	◊●	HPLC	BS90	2–4	P
WTS	○	○	○●	○●	◊●	○●	◊●	HPLC	/	1	P
NSCS	○	○	○●	○●	◊●	○●	◊●	FLUO	/	4	P
RRCB	○	○	○●	○●	◊●	○●	○●	SPEC	/	1	H
Algorithm validation datasets (AVDs)											
POMME	○	○	○●	◊●	○●	○●	◊●	HPLC	BS90	2	H
HUD	○	○	○●	◊●	◊●	○●	◊●	HPLC	/	2–3	P/H
MAPPS	○	○	○●	◊●	◊●	◊●	◊●	HPLC/FLUO	/	1.5–4	P/H
LKSM	○	○	○●	○●	◊●	○●	○●	SPEC	/	1	S

[1] ○: *In situ* measured; ◊: Recovered by post-processing; ●: Matched with satellite data.

[2] HPLC: High-performance liquid chromatograph; FLUO: Fluorometric method; SPEC: Spectrophotometric method.

[3] "/": The measurements of a_{ph} spectra were unavailable and thus were recovered by wavelength-specific coefficients and measurements of Chl*a*; BS90: Bricaud and Stramski, 1990; K85: Kishino et al., 1985.

[4] H: Hyperbolic tangent function; P: Photoinhibition function; E: Exponential function without photoinhibition; S: Smith's function.

2.1.3. Post-processing and quality control

As shown in Table 2.4, the unavailable parameters for each of the ADDs and AVDs were recovered by using a variety of solutions. Missing Z_{eu} and $a_{ph}(\lambda)$ data were inferred from Chl*a* measurements by using the model of Morel et al. (2007) and the wavelength-specific coefficients of Bricaud et al. (1998), repsectively. Another set of coefficients was fitted using Chl*a*-$a_{ph}(\lambda)$ pairs sampled from Lake Kasumigaura (Yang et al., 2011) and used to retrieve $a_{ph}(\lambda)$ data specifically for the LKSM and RRCB datasets. The auxiliary products from the National Centers for Environmental Prediction (NCEP) (Kalnay et al., 1996), Global Ocean Data Assimilation System (GODAS) (Behringer and Xue, 2004), and Hybrid Coordinate Ocean Model (HYCOM) were obtained and used to fill in the missing daily PAR (*d*PAR in mol photons m^{-2} d^{-1}), *T*, and Z_m data, respectively. For each profile, suspicious data identified by either visual inpsection or attached quality flags were discarded. Only photosynthetic parameters that fell within realistic ranges ($\varphi_{max} < 0.125$, $0.2 < P^{B}_{max} < 25$, and $0.002 < \alpha^{B} < 0.2$) (Bouman et al., 2018; Brewin et al., 2017; Platt and Jassby, 1976) and sampled between 0 and $1.5Z_{eu}$ were used. φ_{max} was derived following Uitz et al. (2008) (Uitz et al., 2008) using α^{B} and Chl*a*-specific $a_{ph}(\lambda)$ averaged in the visible domain ($\bar{a}^{*}_{ph}$), as follows:

$$\varphi_{max} = 0.0231\alpha^{B} / \bar{a}^{*}_{ph}, \quad (2.3)$$

where 0.0231 is a factor that converts units of carbon, photon energy, and time. After quality-control, approximately 4% of the data were excluded.

Table 2.4. Summary of the solutions used to recover unavailable parameters. The required measurements or auxiliary products as input (with corresponding temporal and spatial resolutions), descriptions of approaches, and references are shown.

Parameter	Input	Approach description	Reference
Z_{eu}	*In situ* Chl*a* at surface	Empirical function	Morel et al., 2007
a_{ph}	*In situ* Chl*a* profile and wavelength-specific coefficients	Empirical function	Bricaud et al., 1998/This study
T	GODAS assimilated ocean temperature (monthly, 1/3° latitude × 1° longitude × 40 vertical levels)	Spatiotemporal matchup (the data were linearly interpolated into sampling depths)	Behringer and Xue, 2004
dPAR	NCEP downward solar radiation flux (daily, ~2°)	Spatiotemporal matchup (a factor of 0.178 was used for unit conversion)	Kalnay et al., 1996
Z_m	HYCOM (monthly, 9 km)	Spatiotemporal matchup	Ocean productivity website (http://sites.science.oregonstate.edu/ocean.productivity/)

2.1.4. Characteristics of photosynthesis parameters and IPP data

The P^{B}_{max} data ranged from 0.20 to 24.4 mg C (mg Chl)$^{-1}$ h^{-1} (Figure 2.2a), with medians of 2.89 and 2.42 mg C (mg Chl)$^{-1}$ h^{-1} and coefficients of variation (CVs) of 74.5 and 69.1% for the ADDs and AVDs, respectively. In contrast, the dynamic range of E_k varied by three orders of magnitude (Figure 2.2b), in the range 5.75−1150.00 μmol photons m^{-2} s^{-1} for the ADDs and 4.26−1261.00 μmol photons m^{-2} s^{-1} for the AVDs. The corresponding medians (CVs) were 123.93 (68.1%) and 93.24 μmol photons m^{-2} s^{-1} (82.5%), showing a high degree of the natural variability of P^{B}_{max} and E_k in various aquatic environments.

Table 2.3 and Table 2.5 provide the information of *in situ*-satellite matchup for environmental parameters from ocean color products and for calculated IPP data (datasets without valid IPP data are not shown), respectively. The detailed procedure of matchup extraction can be found in Section 2.3. A total of 601 and 2873 valid *in situ* IPP data were obtained for oceanic and inland waters, respectively. Of these, the respective numbers of IPP matchups for oceanic and inland waters were 57 and 70 (see Figure 2.1 for locations). To conduct comprehensive comparative analyses, both matchup and independent *in situ* data (AVDs) were used. In general, IPP data varied by three orders of magnitude between 14.18 and 12,196.02 mg C m^{-2} d^{-1} (Figure 2.3a), with medians of 510.87 mg C m^{-2} d^{-1} and 888.92 mg C m^{-2} d^{-1} for oceanic and inland waters, respectively (Table 2.5).

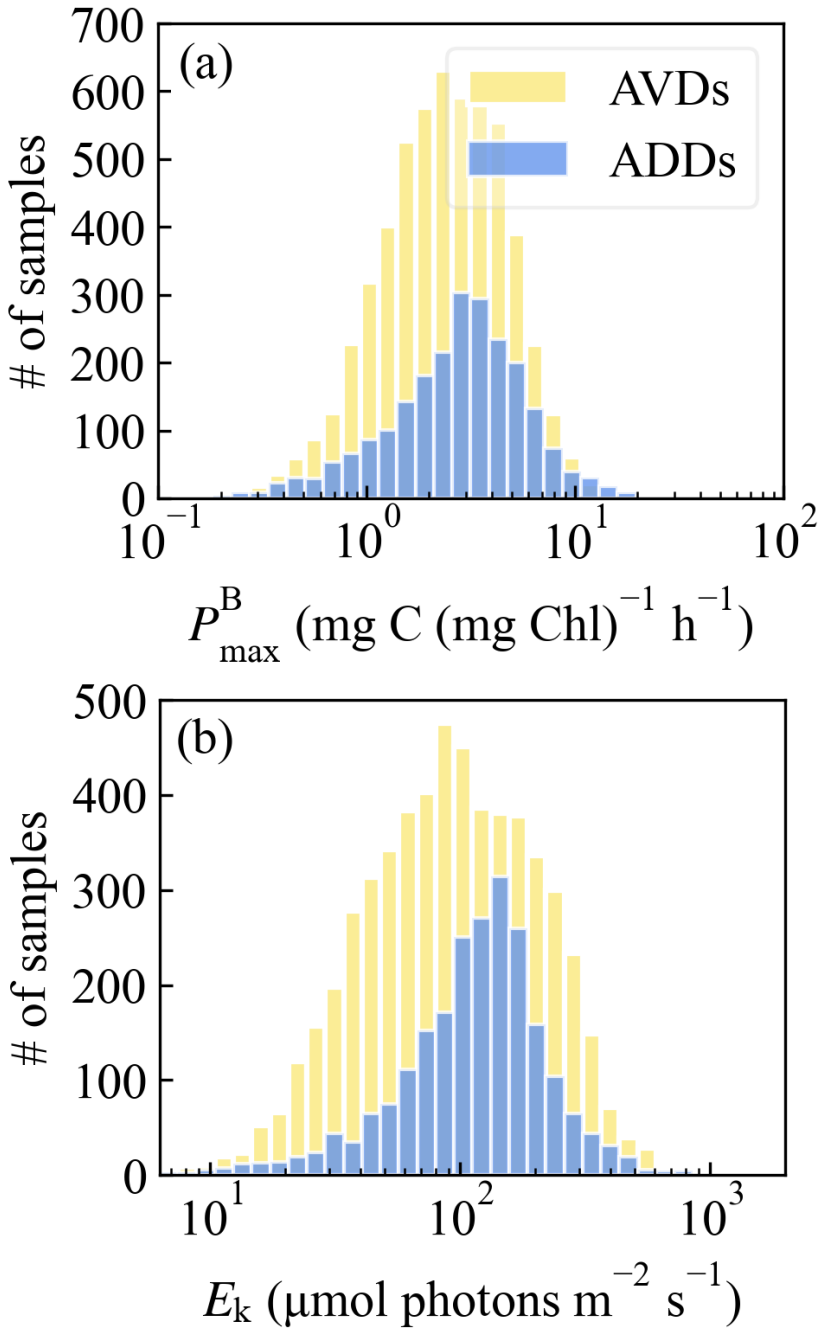

Figure 2.2. Histograms of the frequency distributions of P^{B}_{max} (a) and E_k (b) for the ADDs and AVDs.

Table 2.5. Information (data range and number of data) for valid *in situ* and matched IPP data available for ADDs and AVDs from different water types.

Water type	Dataset	Range of IPP (mg C m^{-2} d^{-1}) (median value)	Number of IPP data	
			In situ	Matchup
Oceanic waters	ADD 1 (EUMELI4)	53.22–3,272.13 (143.69)	18	/
	ADD 2 (OLIPAC)	71.57–295.98 (119.69)	8	/
	ADD 3 (ANTARES4)	493.10–1,144.84 (868.52)	9	1
	ADD 4 (PROSOPE)	71.32–1,922.59 (143.20)	10	8
	ADD 8 (BIOSOPE)	35.78–1,432.68 (226.24)	27	10
	ADD 9 (KEOPS)	531.86–1,232.35 (654.72)	4	/
	ADD 11 (MALINA)	109.53–195.98 (152.75)	2	1
	ADD 13 (NSCS)	169.21–1,639.92 (684.10)	23	5
	AVD 1 (POMME)	104.29–1,486.65 (308.05)	45	9
	AVD 3 (MAPPS)	14.18–5,787.45 (600.96)	455	23
	All	14.18–5,787.45 (510.87)	601	57
Inland waters	ADD 14 (RRCB)	24.98–12,196.02 (1235.90)	1410	29
	AVD 4 (LKSM)	47.09–5602.33 (641.59)	1463	41
	All	24.98–12,196.02 (888.92)	2873	70

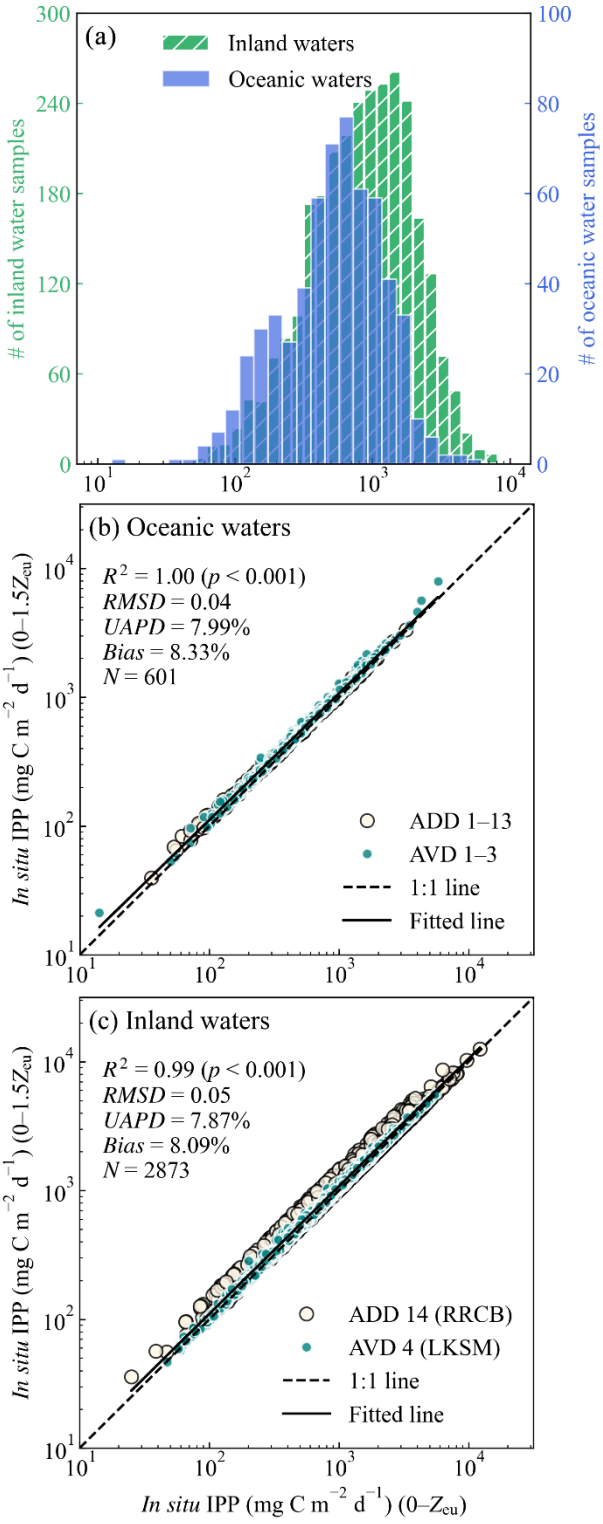

Figure 2.3. (a) Frequency distribution histograms of *in situ* IPP data from oceanic and inland waters. Comparison between *in situ* IPP data calculated by the vertical integration over the euphotic zone (0−Z_{eu}) and those over the productive zone (0−1.5Z_{eu}) from oceanic waters (b) and inland waters (c).

The difference between vertically integrating PP within the euphotic zone ($0-Z_{eu}$) and the productive zone ($0-1.5Z_{eu}$) were examined for different water types. As shown in Figure 2.3b and Figure 2.3c, the calculated IPP data over the euphotic and productive zones was nearly identical for each water type, though extending the vertical integration to $1.5Z_{eu}$ tended to slightly overestimate IPP data for the RRCB dataset due to the shallow bathymetry of the Rhode River. Nevertheless, this result suggests that, in theory, extending the vertical integration to $1.5Z_{eu}$ for depth-resolved models to compute satellite-based IPP would not cause a discrepancy in comparisons with IPP data of the euphotic zone that are measured *in situ* or estimated by DI models.

2.2. Satellite data for different water types

All *in situ* datasets were simply categorized into two water types according to the sampling location: (1) ADD 1–13 and AVD 1–3 were categorized as oceanic waters, and (2) ADD 14 and AVD 4 were categorized as inland waters. Multiple satellite products were used specifically for different water types to assess model applicability. Several environmental parameters measured by satellites are required for remote sensing application of our model, including ocean color parameters, water temperautre, and *d*PAR. The satellite products used for respective purposes and water types were summarized in Table 2.6. The procedure of retrieving ocean color parameters by preferred algorithms was described, and the accuracy of satellite-measured environmental parameters was assessed.

Table 2.6. Satellite products used for different water types with corresponding purposes, time spans of usage, temporal resolutions, and spatial resolutions given in parentheses.

Water type & Dataset	Purpose		Satellite-measured parameters				
	Matchup extraction	IPP product generation	T	Chl$a$	Z_{eu}	a_{ph}	dPAR
Oceanic waters: ADD 1–13 & AVD 1–3	○		SST-CCI (1992–2013, daily, 0.05°)	OC-CCI (1997–2013, daily, 4 km)			SeaWiFS/MODIS (1997–2013, daily, 9/4 km)
Inland waters: ADD 14 (RRCB)	○		SST-CCI (1990–2009, daily, 0.05°)	MERIS (2002–2012, 300 m)			SeaWiFS/MODIS (1997–2016, daily, 9/4 km)
		○ (showcase)	MODIS (2020, daily, 4 km)	MSI (2020, 60 m)		/	MODIS (2020, daily, 4 km)
AVD 4 (LKSM)	○		GloboLakes (1995–2016, daily, 0.05°)	MERIS/MSI (2002–2012/2015–2016, 300/60 m)			SeaWiFS/MODIS (1997–2016, daily, 9/4 km)
	○	○ (showcase)	CGLOPS (2018–2019, daily, 1 km)	MSI (2018–2019, 60 m)		/	MODIS (2018–2019, daily, 4 km)
Global waters		○ (1998–2021)	SST-CCI/MODIS (1998–2021, monthly, 0.05°/4 km)	OC-CCI (1998–2021, monthly, 4 km)		/	SeaWiFS/MODIS (1998–2021, monthly, 4 km)

2.2.1. Satellite products for oceanic waters

The daily products of remote sensing reflectance (R_{rs}) and Chl*a* from the Ocean Color-Climate Change Initiative (OC-CCI, version 5.0) project (Sathyendranath et al., 2019) and the daily SST product from the SST-CCI (version 2.1) project (Merchant et al., 2019) were used. These products are available from the ESA's CCI website (https://climate.esa.int/en/). Because these products provide merged observations from multiple sensors, they have improved data coverage. The Level-3 daily (monthly) product of *d*PAR (SST) from SeaWiFS or MODIS-Aqua/Terra were obtained from the NASA's Ocean Color website (https://oceancolor.gsfc.nasa.gov/). If *d*PAR (or SST) estimates from both MODIS-Aqua and -Terra were available, the arithmetic mean of them was used.

2.2.2. Satellite products for inland waters

MERIS Level-1 and Sentinel-2A/B MultiSpectral Instrument (MSI) Level-1C images were used for inland waters because their fine spatial resolution was best suited to the small size of the target waters. The images were downloaded from the ESA's Earth Online data archive (https://earth.esa.int/eogateway/) and Copernicus Open Access Hub (https://scihub.copernicus.eu/dhus/), respectively. The lake surface water temperature (LSWT) product from the Global Observatory of Lake Responses to Environmental Change (GloboLakes, version 4.0) project (Carrea and Merchant, 2019) was acquired from the Center for Environmental Data Analysis (CEDA) archive (http://archive.ceda.ac.uk/) and was specifically used to extract satellite-filed matchup data for Lake Kasumigaura. In addition,

another LSWT product, provided by the Copernicus Global Land Operations (CGLOPS) project (reprocessed version 1.0.2) that is available from the Copernicus LSWT Data Portal (https://land.copernicus.eu/global/products/lswt), was used for generating IPP product of Lake Kasumigaura as a showcase. Because the CGLOPS project provides more up-to-date observations of LSWT. Whereas the SST-CCI and MODIS SST products was used for matchup extraction and generation of IPP product in the Rhode River and the upper Chesapeake Bay, respectively. The *d*PAR product from SeaWiFS/MODIS was used for inland waters in the same manner as oceanic waters.

2.2.3. Field-satellite matchup extraction

To extract matchups from oceanic waters, we first identified pairs of daily satellite products and *in situ* measurements that were acquired on the same day, and then the product data within a 3 × 3 pixel box centered on the matched *in situ* sites was extracted. The standard data-screening protocol (Bailey and Werdell, 2006) was applied to minimize systematic uncertainties in data matching. In summary, (1) the protocol required that at least 50% of pixels within the 3 × 3 pixel box contain valid data; (2) the median value and standard deviation of the data in all pixels containing valid data were calculated, and pixels containing data that lay outside the range defined by the median ±1.5 × standard deviations were excluded; (3) the mean value of the data in the remaining pixels was taken as a valid matchup if its CV was less than 15%. The same procedure and protocol were applied to extract matchups of *T* and *d*PAR from the daily products that covered

the Rhode River and Lake Kasumigaura. Wider time windows of ±24 h and ±3 d were used to extract matchups from MERIS and MSI data, respectively.

To evaluate the accuracy of the matchup data, inculding satellite-estimated *d*PAR, *T*, Chl*a*, and K_d(PAR), the statistical measures of root mean square error (*RMSE*), mean absolute percentage error (*MAPE*), and *bias* were used. As definitions in Eq. (2.4) to Eq. (2.6), the units of *RMSE* and *bias* are the same with corresponding input parameters (y_i' and y_i represent estimated and measured value, and N is the number of data points). The use of these metrics can put our results in the context of other literatures to gauge whether the accuracy is satisfactory.

$$RMSE = \sqrt{\frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}}(y_i' - y_i)^2} \tag{2.4}$$

$$MAPE = \frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}}\left|\frac{y_i' - y_i}{y_i}\right| \times 100\% \tag{2.5}$$

$$bias = \frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}}(y_i' - y_i) \tag{2.6}$$

Assessments of the Chl*a* and *T* matchups extracted from oceanic waters were shown in Figure 2.4a and Figure 2.4b, respectively, where the depth-dependent Chl*a* and *T* were estimated from corresponding surface values by the method of Uitz et al. (2006) and the GODAS reanalysis data, respectively (see Section 3.2.2). Vertical retrievals of Chl*a* and *T* from satellites are important to depth-resolved IPP models, and thus the matchups for different depths should also be assessed.

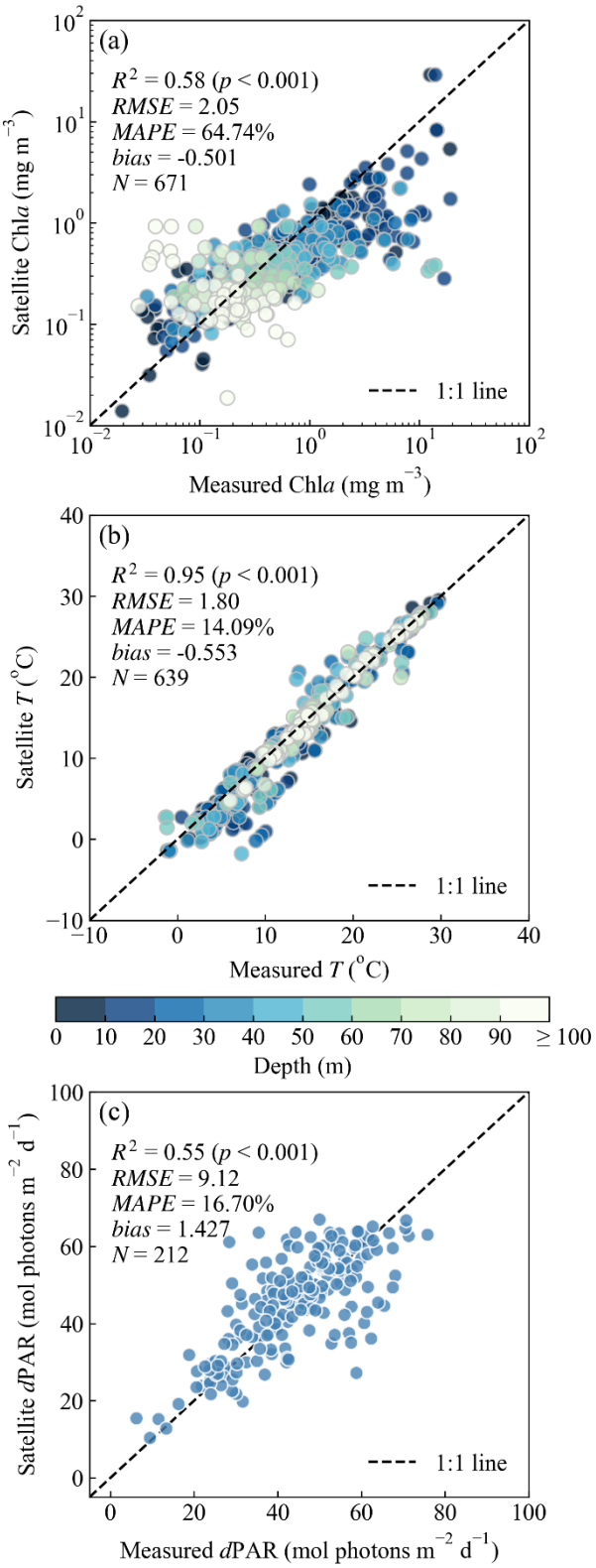

Figure 2.4. Evaluation of satellite-estimated Chl*a* (a), *T* (b), and *d*PAR (c) for oceanic waters. The vertical profiles of Chl*a* and *T* from satellites were estimated by the method of Uitz et al. (2006) and by the GODAS reanalysis data, respectively. The sampling depths are indicated as different colors. *N* is the number of matchups.

The matchups of Chl*a* showed an overestimation for small values (< 0.1 mg m^{-3}) and an underestimation for large Chl*a* values (> 1 mg m^{-3}), with R^2 of 0.58 and *RMSE* of 2.1 mg m^{-3}. The Chl*a* estimates at surface layer were generally more accurate than those at deeper layer. While the matchups of *T* exhibited a satisfactory accuracy for all depths, with high R^2 of 0.95 and low *RMSE* of 1.8 °C. Figure 2.4c shows the matchups of *d*PAR from oceanic waters, where the number of valid matchups was 212 mainly because this parameter is not depth-dependent. In general, part of the satellite-measured *d*PAR was well consistent with the *in situ* measured values, but some scatters were found to be away from the 1:1 line, which might be caused by the uncertainty of the recovered *d*PAR values from the NCEP dataset. Nevertheless, the overall accuracy of satellite-measured *d*PAR for oceanic waters is acceptable, with a *MAPE* of 16.7 % and a small *bias* of 1.4 mol photons m^{-2} d^{-1}.

For inland waters, assessments of the extracted matchups of *d*PAR and *T* from the RRCB and LKSM datasets were shown in Figure 2.5. In general, the machups of *d*PAR from both datasets had a reasonable accuracy with the values of *MAPE* lower than ~11.5% and the values of R^2 around 0.90. Whereas for the matchups of *T*, though the accuracy of LKSM dataset was clearly higher than that of RRCB dataset, mainly because the sampling sites of RRCB were closer to land area, the satellite-estimated *T* in these two inland water areas were considered to be acceptable. The retrievals and evaluations of the satellite-estimated Chl*a* and K_d(PAR) in inland waters can be found in Section 2.2.4.

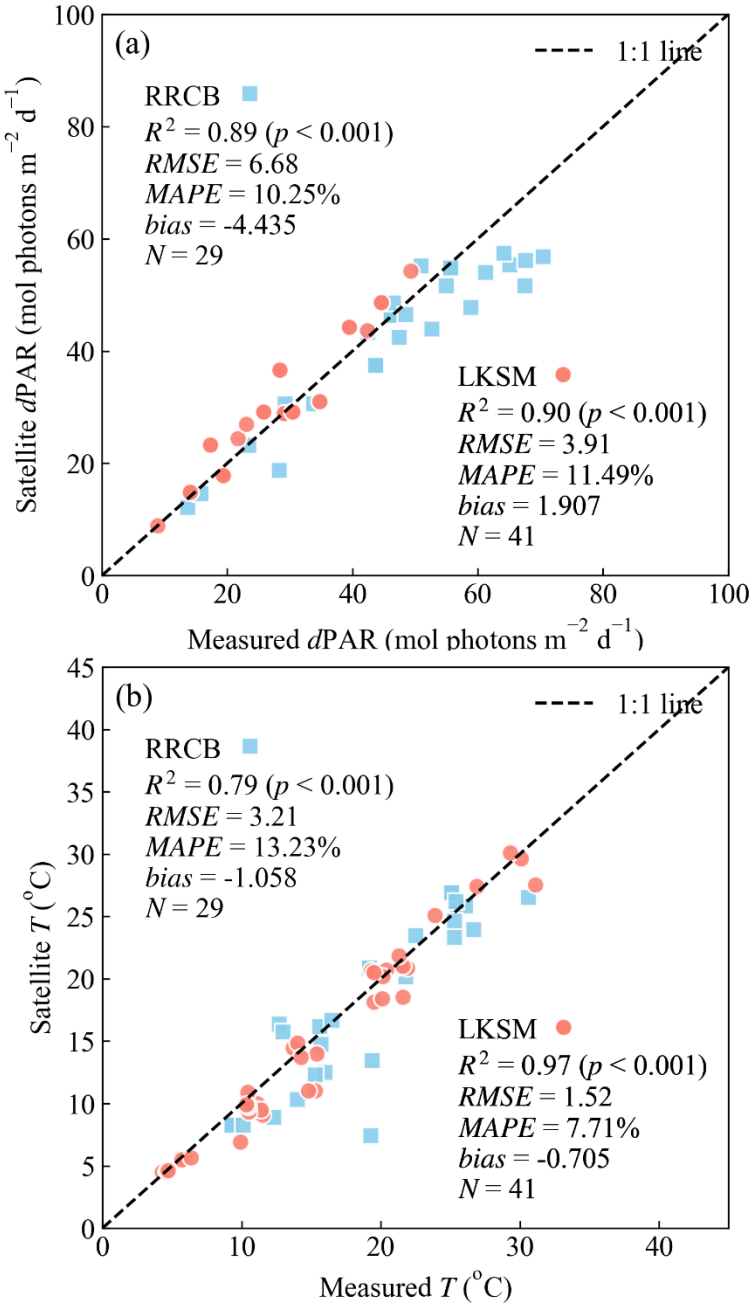

Figure 2.5. Evaluation of satellite-estimated *d*PAR (a) and *T* (b) from the RRCB and LKSM datasets (inland waters).

2.2.4. Bio-optical parameter retrievals from ocean color

In addition to the OC-CCI Chl*a* product that was directly available for oceanic waters, the bio-optical parameters of Z_{eu}, equivalent to K_d(PAR), and a_{ph} were retrieved from the OC-CCI $R_{rs}(\lambda)$ data. Similarly, the Chl*a*, Z_{eu}, and a_{ph} data for inland waters were retrieved from MERIS/MSI $R_{rs}(\lambda)$ data obtained after atmospheric correction. Since current ocean color algorithms for retrieving bio-optical parameters have limited performance and applicability corresponding to the target water types, such as Case-I and Case-II waters, several popular algorithms were selected and tested to determine the most proper one for each bio-optical parameter of each water types.

Figure 2.6 illustrates the overall workflow of generating bio-optical parameters for different water types. Atmospheric correction of MERIS and MSI images was performed using the POLYnomial-based algorithm applied to MERIS (POLYMER, version 4.13) (Steinmetz et al., 2011) because of its proven performance and wide applicability in different water environments (Qin et al., 2017; Tilstone et al., 2021; Warren et al., 2019). Then, a quality assurance (QA) model (Wei et al., 2016) was adopted to exclude questionable MERIS/MSI $R_{rs}(\lambda)$ with a low QA score (< 0.5) returned by the POLYMER by assessing the rationality of R_{rs} spectral shapes. Furthermore, a virtual-band estimator (Wei et al., 2019) was applied to the quality-controlled MSI $R_{rs}(\lambda)$ data so that the spectra could be extended to 412 nm to retrieve a_{ph} at 412 nm, which is vital for $a_{ph}(\lambda)$ estimation.

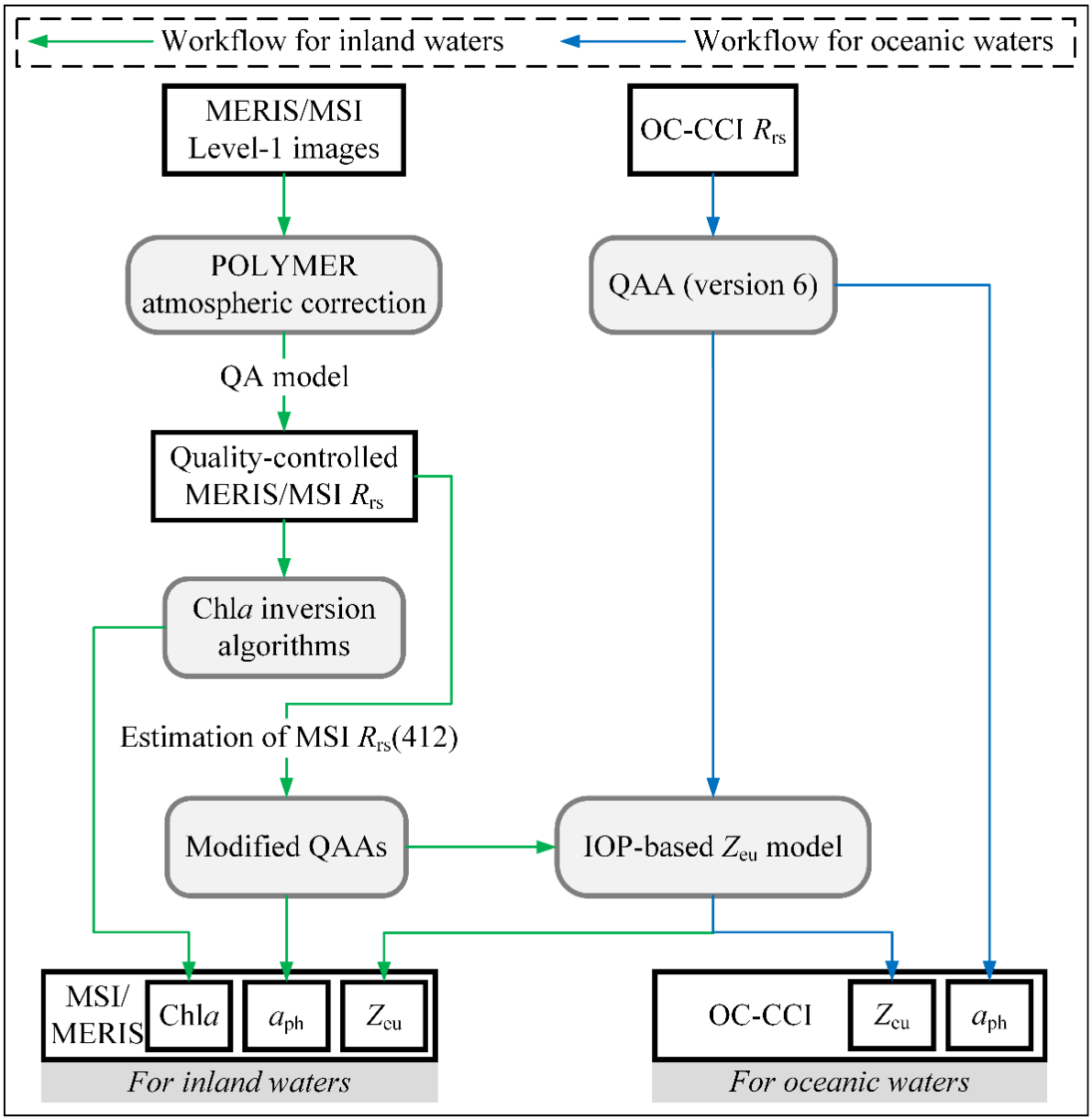

Figure 2.6. The workflows used to generate bio-optical parameters from MERIS/MSI images for inland waters (Chl*a*, a_{ph}, and Z_{eu}) and from the OC-CCI R_{rs} products for oceanic waters (a_{ph} and Z_{eu}) using preferred algorithms determined by performance assessment.

The quasi-analytical algorithm (QAA, version 6) (IOCCG, 2014; Lee et al., 2002) was used to estimate the inherent optical properties (IOPs) from $R_{rs}(\lambda)$ in oceanic waters. The IOP retrievals included the total absorption coefficient (a), total backscattering coefficient (b_b), and a_{ph}. The configuration of QAA was set following Yu et al. (2019) to enhance the model performance in oligotrophic waters, such as the Pacific gyre. The IOP-based Z_{eu} model (Lee et al., 2007, 2005) was employed to estimate Z_{eu} (or K_d(PAR)) by using the QAA-derived a and b_b at 490 nm, termed as $a(490)$ and $b_b(490)$, respectively. The model was also used for inland waters to estimate Z_{eu} because its semi-analytical nature can well deal with clear and turbid waters. The OC-CCI-based $a_{ph}(\lambda)$ data were retrieved by the QAA, and hence the maximum value of $a_{ph}(\lambda)$ and the spectrally averaged $a_{ph}(\lambda)$ over the visible domain (termed as $\bar{a}_{ph}$) could also be obtained.

As shown in Figure 2.7a, the satellite-measured K_d(PAR) data had an obvious correlationship with the measured ones, with a high R^2 of 0.63. The scatters generally distributed along the 1:1 lines except for some outliers, which might be caused by the uncertainty in the recovered K_d(PAR) values. Given a small *bias* and a *MAPE* of 23.7 %, we can still conclude that the K_d(PAR) data estimated by the QAA and the IOP-based model are satisfactory. The assessment of satellite-measured $\bar{a}_{ph}$ was given in Figure 2.7b. The depth-dependence of estimated $\bar{a}_{ph}$ was not considered because there was not a practically usable method to address this issue. Overall, the $\bar{a}_{ph}$ retrieval had a reasonable accuracy ($R^2 = 0.46$) for all depths with small values of *bias* and *RMSE*.

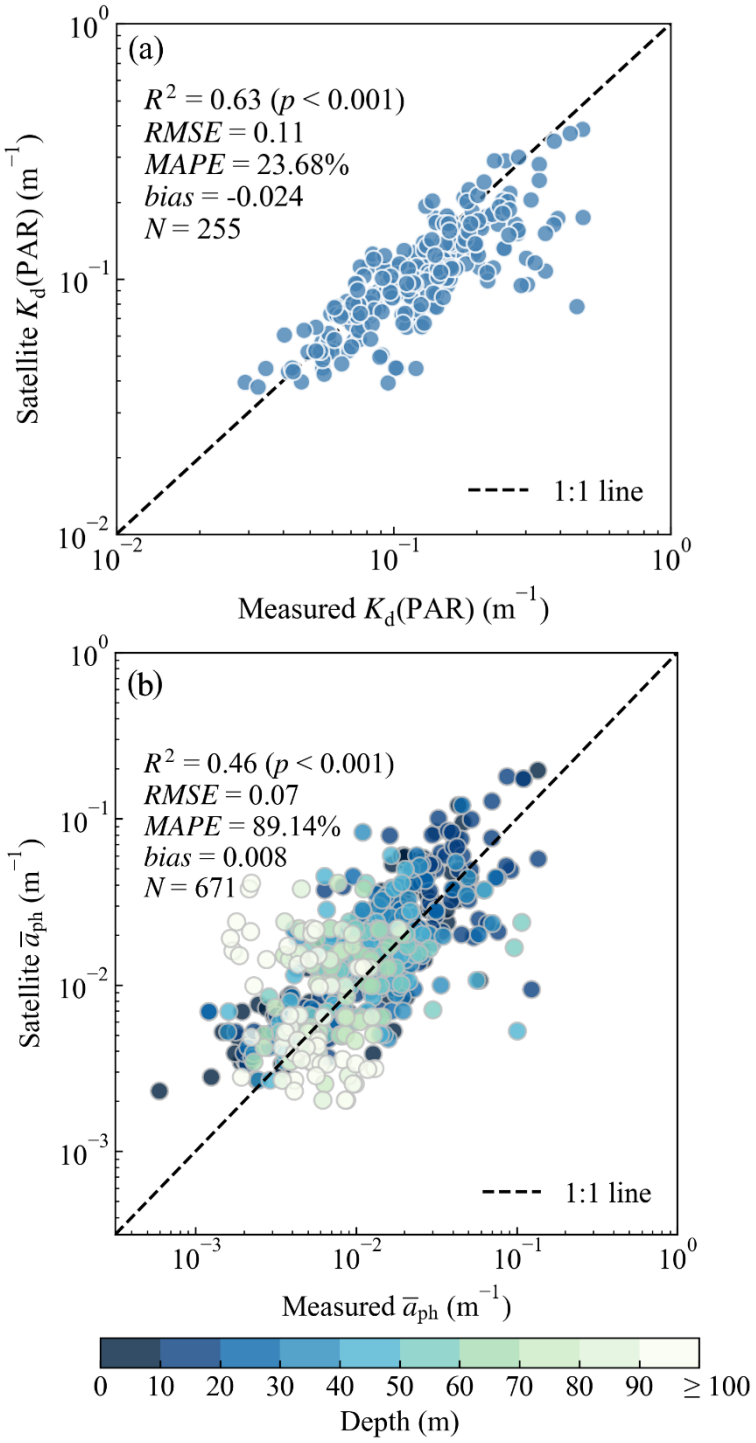

Figure 2.7. Evaluation of satellite-estimated K_d(PAR) (a) and $\bar{a}_{ph}$ (b) for oceanic waters. *N* is the number of matchups.

The QAA-based K_d(PAR) estimation from MERIS and MSI showed a significant underestimation in the Rhode River and Lake Kasumigaura as shown in Figure 2.8a and Figure 2.9a, suggesting that the QAA (version 6) is not directly applicable to turbid waters. Therefore, two modified versions of the QAA that had been verified to work better in optically complex (Case-II) waters were selected and evaluated, namely the QAA_Xue (Xue et al., 2019) and QAA_Yang (Yang et al., 2013). Evaluations showed that the QAA_Yang-based model yielded some outliers of MERIS K_d(PAR) estimates (Figure 2.8b) and underestimated K_d(PAR) for MSI (Figure 2.9b) in Lake Kasumigaura. Whereas the QAA_Xue-based model returned more realistic K_d(PAR) estimates for MERIS (Figure 2.8c) and MSI in Lake Kasumigaura (Figure 2.9c) with lower *RMSE* and *MAPE*. As for the K_d(PAR) retrieval in the Rhode River, the QAA_Xue-based model also exhibited a better performance than the QAA_Yang, with a higher R^2 of 0.77 and a lower *MAPE* of 30.6 %. Hence, the QAA_Xue was used to estimate K_d(PAR) because it provided more accurate a(490) and b_b(490) from MERIS and MSI. Moreover, MERIS and MSI-based a_{ph} data at 443 nm (a_{ph}(443)) were derived using the QAA_Yang, which were further transformed to $\bar{a}_{ph}$ by a linear function with a slope of 0.435 and an intercept of 0.015. These two coefficients were fitted using the forementioned *in situ* $a_{ph}(\lambda)$ data measured in Lake Kasumigaura. We assumed that this linear function could be applicable to the Rhode River to some extent since the bio-optical characteristics of these two turbid water areas could be similar. The accuracy of estimated $\bar{a}_{ph}$ was proved to be acceptable (Figure 2.10).

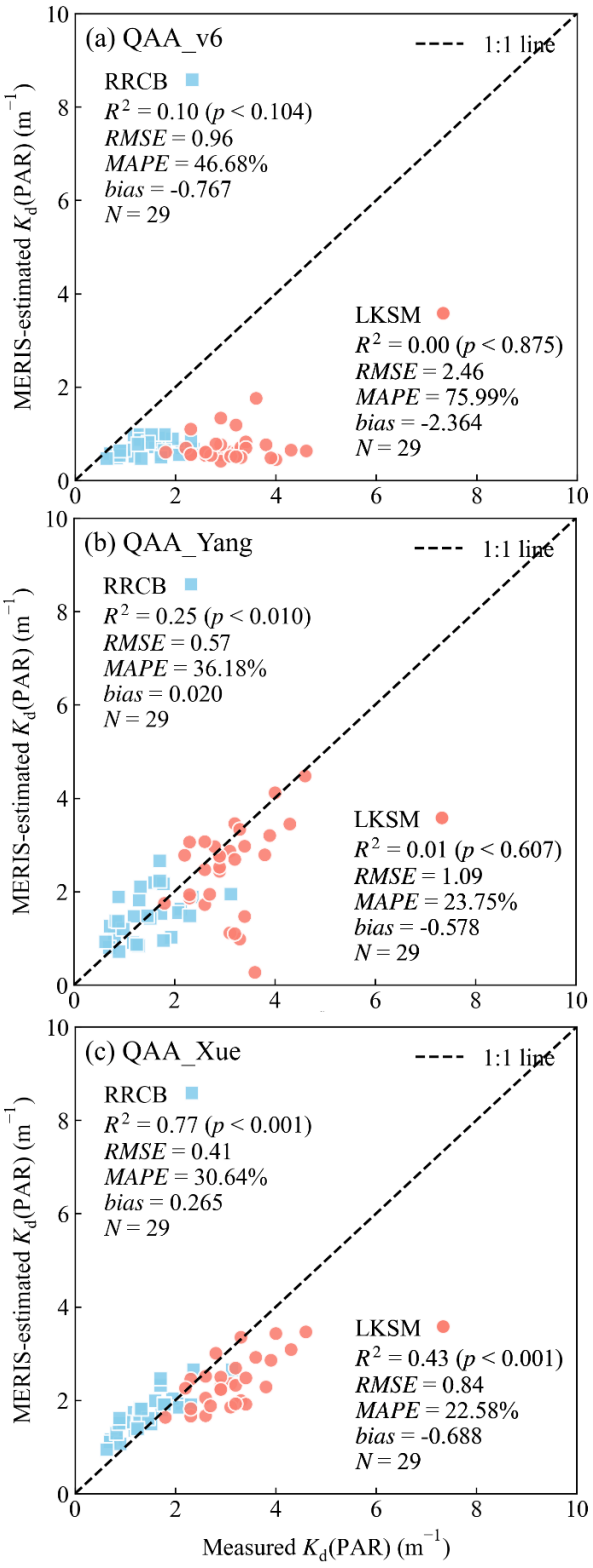

Figure 2.8. Comparison of satellite-estimated K_d(PAR) based on MERIS matchups using different versions of QAA: QAA_v6 (a), QAA_Yang (b), and QAA_Xue (c) for the RRCB and LKSM datasets (inland waters).

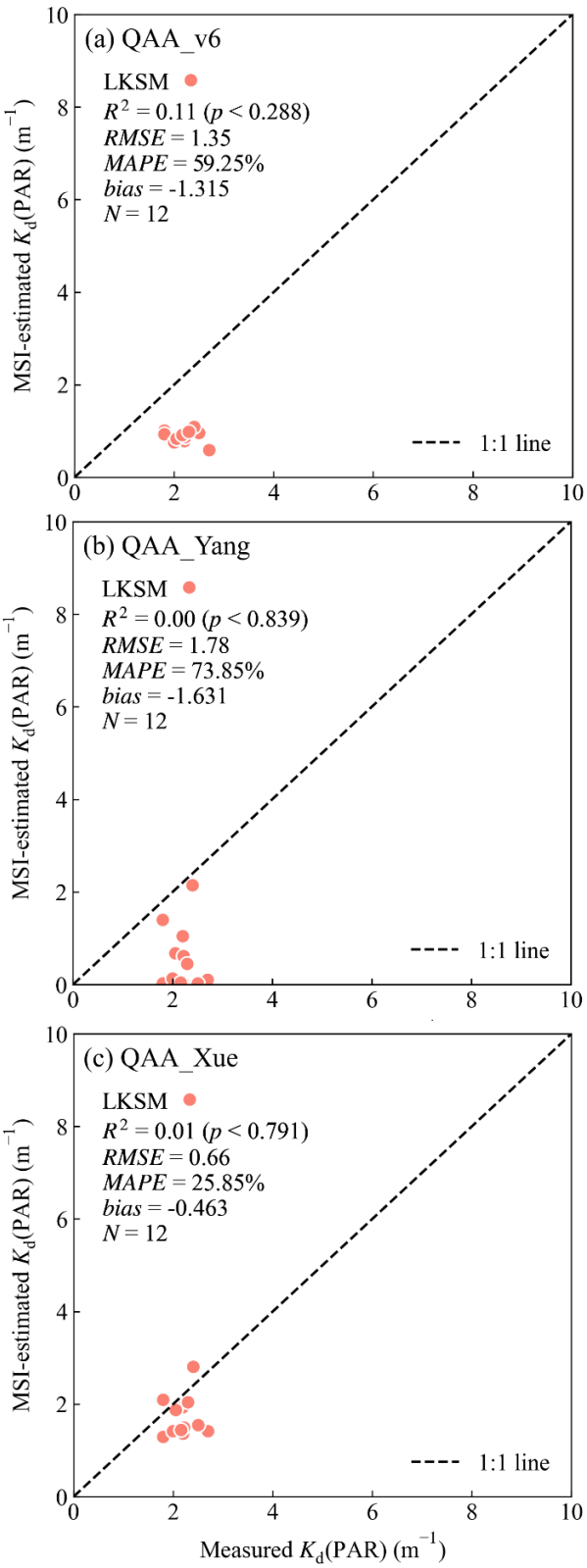

Figure 2.9. Same as Figure 2.8, but for MSI matchups of the LKSM dataset (inland waters).

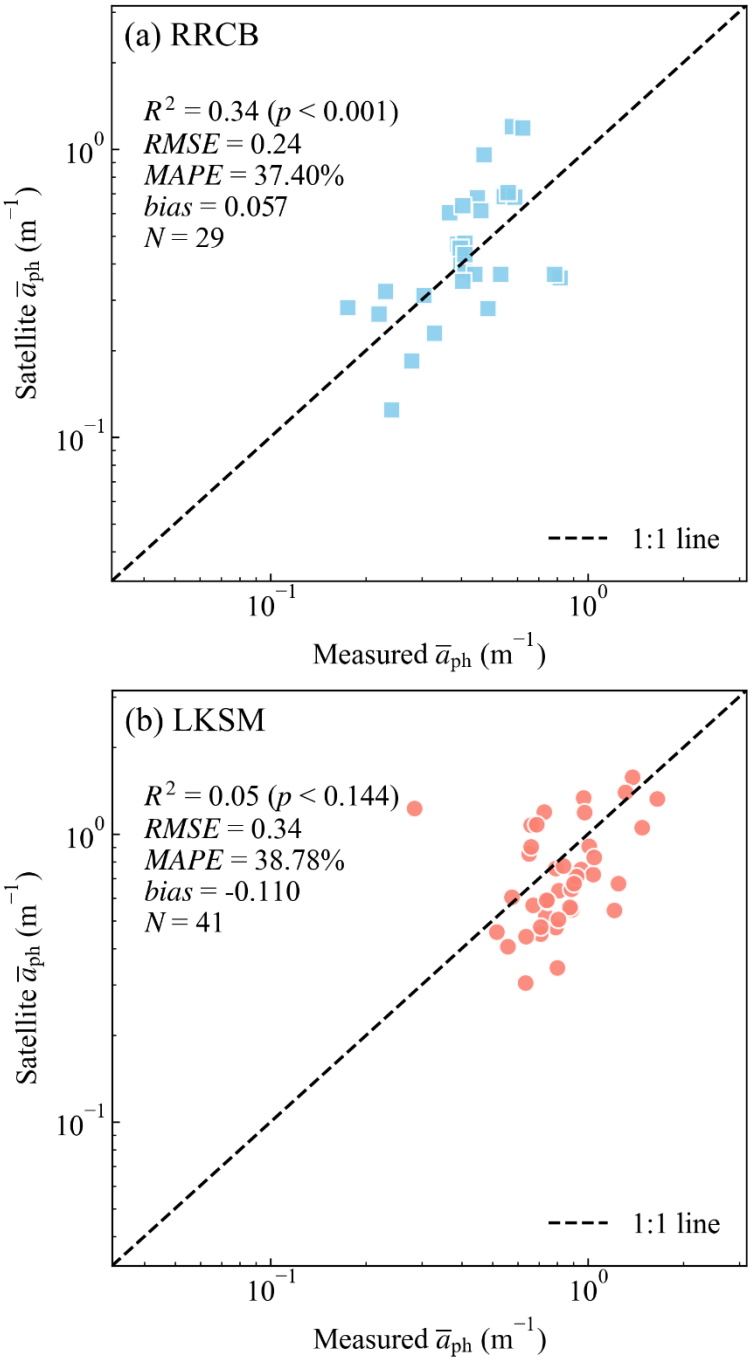

Figure 2.10. Comparison of MERIS/MSI-estimated $\bar{a}_{ph}$ from the estimated $a_{ph}(443)$ based on QAA_Yang for the RRCB (a) and LKSM (b) datasets (inland waters).

Four Chl*a* inversion algorithms were tested to determine the preferred algorithms for inland waters, namely Gilerson 2-Band (G2B) (Gilerson et al., 2010), Yang 3-Band (Y3B) (Yang et al., 2010), the Normalized Difference Chlorophyll Index (NDCI) (Mishra and Mishra, 2012), and the mixture density network (MDN) (Pahlevan et al., 2020). Note that the MDN was designed for the Sentinel-3A/B (S3A/B) Ocean and Land Color Imager (OLCI) and Sentinel-2 MSI, and hence the POLYMER-corrected R_{rs} spectra of MERIS were linearly shifted into the OLCI bands before the implementation of MDN. The identical or very close bands between MERIS and OLCI were unnecessary to be adjusted. Table 2.7 provides the fitted coefficients and corresponding errors of band shifting. As shown in Figure 2.11, the MDN performed best on the MERIS $R_{rs}(\lambda)$ in the Rhode River with the lowest *MAPE* of 45.1% and more valid matchups ($N = 29$), while other three algorithms had *MAPE* values over 51.2%. As for Lake Kasumigaura, the NDCI yielded the most accurate MERIS Chl*a* estimates with the highest R^2 of 0.84 and the lowest *MAPE* of 19.7 %. Moreover, Figure 2.12 shows that the NDCI and G2B significantly underestimated MSI Chl*a* in Lake Kasumigaura (bias < –8.0 mg m^{-3}), and the Y3B had lower *RMSE* and *MAPE* than the MDN, even though the R^2 of MDN was slightly higher. Overall, the scatters of the Y3B distributed closer to 1:1 line than those of the MDN. In conclusion, the assessment suggested that the NDCI and Y3B were preferred to estimate Chl*a* in Lake Kasumigaura from MERIS and MSI data, respectively. Whereas the MDN was preferred to estimate MERIS-based Chl*a* in the Rhode River.

Table 2.7. The band configurations of OLCI and MERIS for the implementation of MDN. A spectral reconstruction model was established based on linear regression using the *in situ* R_{rs} spectra collected in Lake Kasumigaura in order to convert the MERIS spectra to the OLCI-like spectra. In detail, the missing $R_{rs}(\lambda)$ (the target output) in MERIS spectra was estimated using the spectral information from two neighboring bands (i.e., $R_{rs}(x_1)$ and $R_{rs}(x_2)$), e.g., $R_{rs}(673) = a \times R_{rs}(665) + b \times R_{rs}(681) + c$.

Wavelength (nm)		Input bands		Regression coefficients			R^2	*MAPE* (%)
OLCI	MERIS	x_1	x_2	*a*	*b*	*c*		
411	412							
442	443							
490	490							
510	510				/		/	/
560	560							
619	620							
664	665							
673	/	665	681	0.47326	0.47642	-0.00020	0.99	2.83
681	681							
708	709				/		/	/
753	754							
761	/	754	779	1.87736	-0.93472	-0.00013	0.88	18.46
764	/	754	779	1.98605	-1.05270	-0.00010	0.90	11.78
767	/	754	779	1.26497	-0.31616	-0.00003	0.98	9.33
778	779				/		/	/

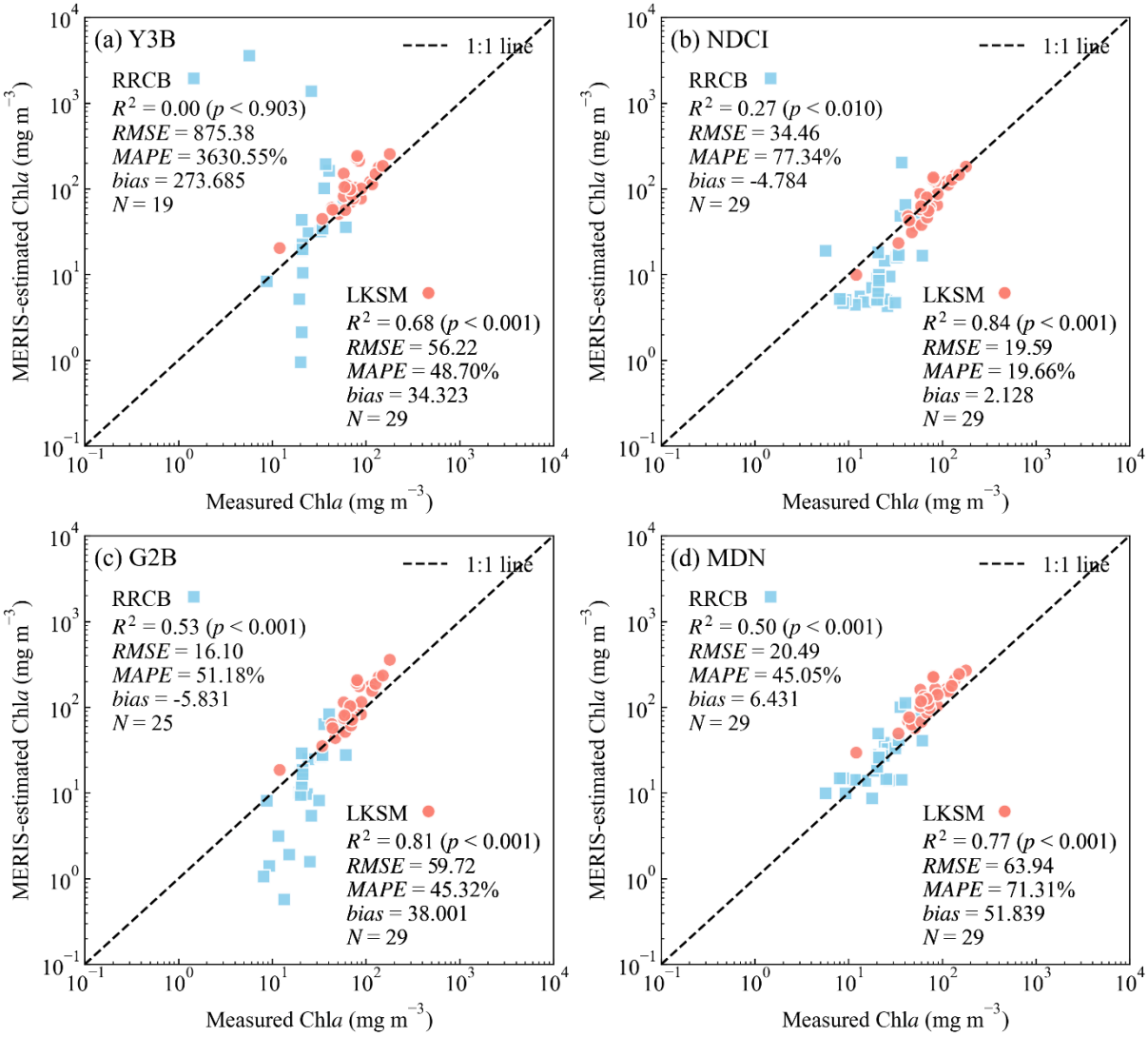

Figure 2.11. Comparison between MERIS-estimated Chl*a* using different algorithms and *in situ*-measured Chl*a*: Y3B (a), NDCI (b), G2B (c), and MDN (d) for the RRCB and LKSM datasets (inland waters).

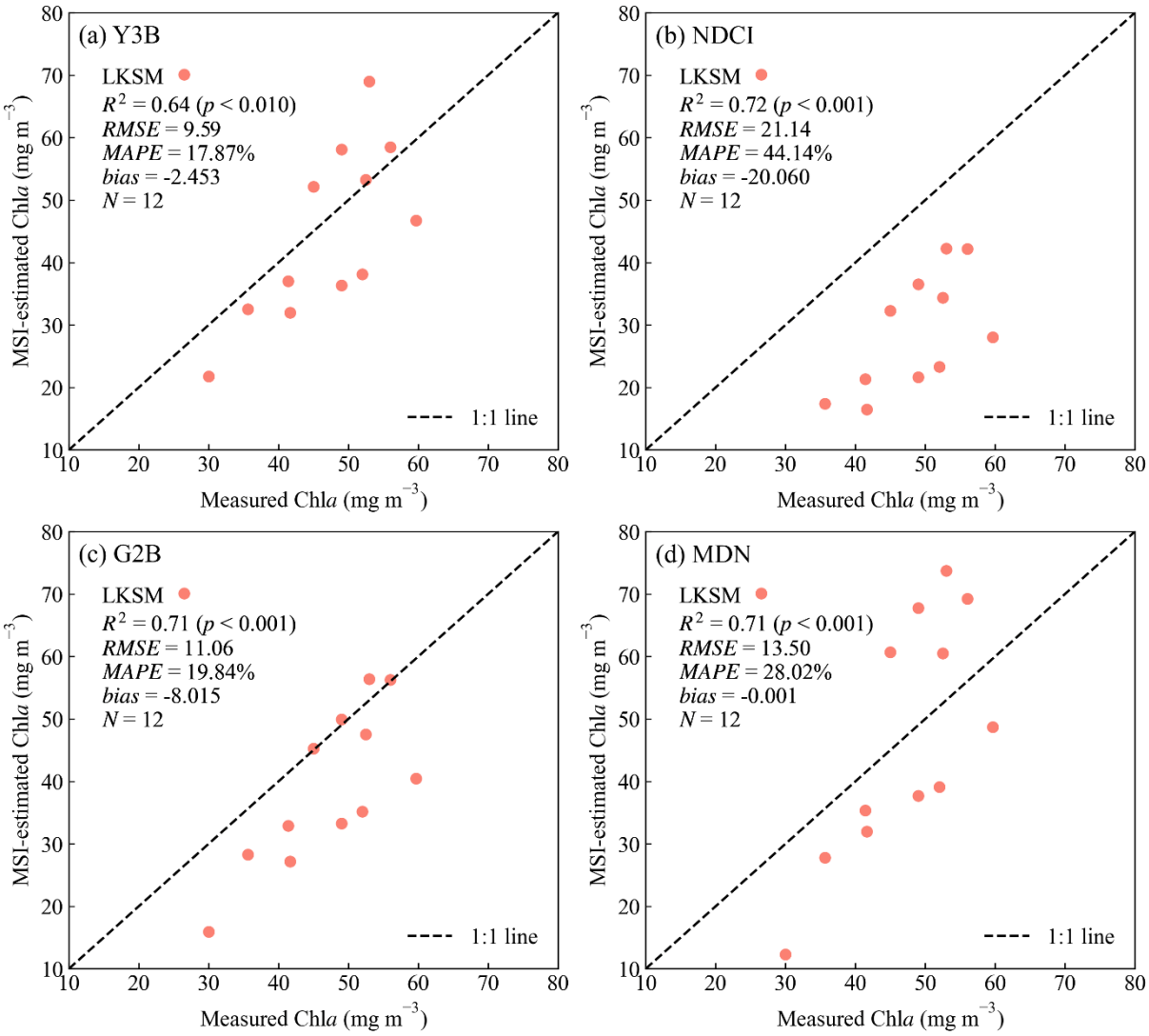

Figure 2.12. Same as Figure 2.11, but for MSI matchups of the LKSM dataset (inland waters).

2.3. Bias-correction for satellite products

Multiple satellite products were used to generate long-term IPP product in this study, and given the potential uncertainty of ocean color signals, algorithms, and products, the systematic bias in the input products should be first investigated and then be corrected, if necessary, to ensure data consistency and continuity. The principal environmental parameters that we considered were *d*PAR and SST.

2.3.1. Multi-sensor *d*PAR products

To quantify the systematic bias between monthly mean *d*PAR from SeaWiFS and MODIS (average of MODIS-Aqua and MODIS-Terra), linear regression analysis was used for the overlapped time periods (2000–2007). MODIS *d*PAR was taken as reference, and only the pixels available in both products were used. Monthly changes of the linear slopes and intercepts were exhibited in Figure 2.13a and Figure 2.13b, respectively, where apparent seasonal patterns can be observed. The slopes typically ranged between 0.96 and 1, while the intercepts ranged between –1.2 and 0.1 mol photons m^{-2} d^{-1} approximately, implying that SeaWiFS *d*PAR data were prone to be larger than MODIS. Thus, multi-year monthly mean slopes and intercepts were used to correct the bias in SeaWiFS *d*PAR. Figure 2.13c displays the time series of global monthly mean *d*PAR from MODIS and SeaWiFS. As expected, the bias-corrected time series of SeaWiFS *d*PAR is more consistent with those of MODIS-Aqua/Terra, as the seasonal amplitudes were nearly identical, which means that our bias-correction approach was effective.

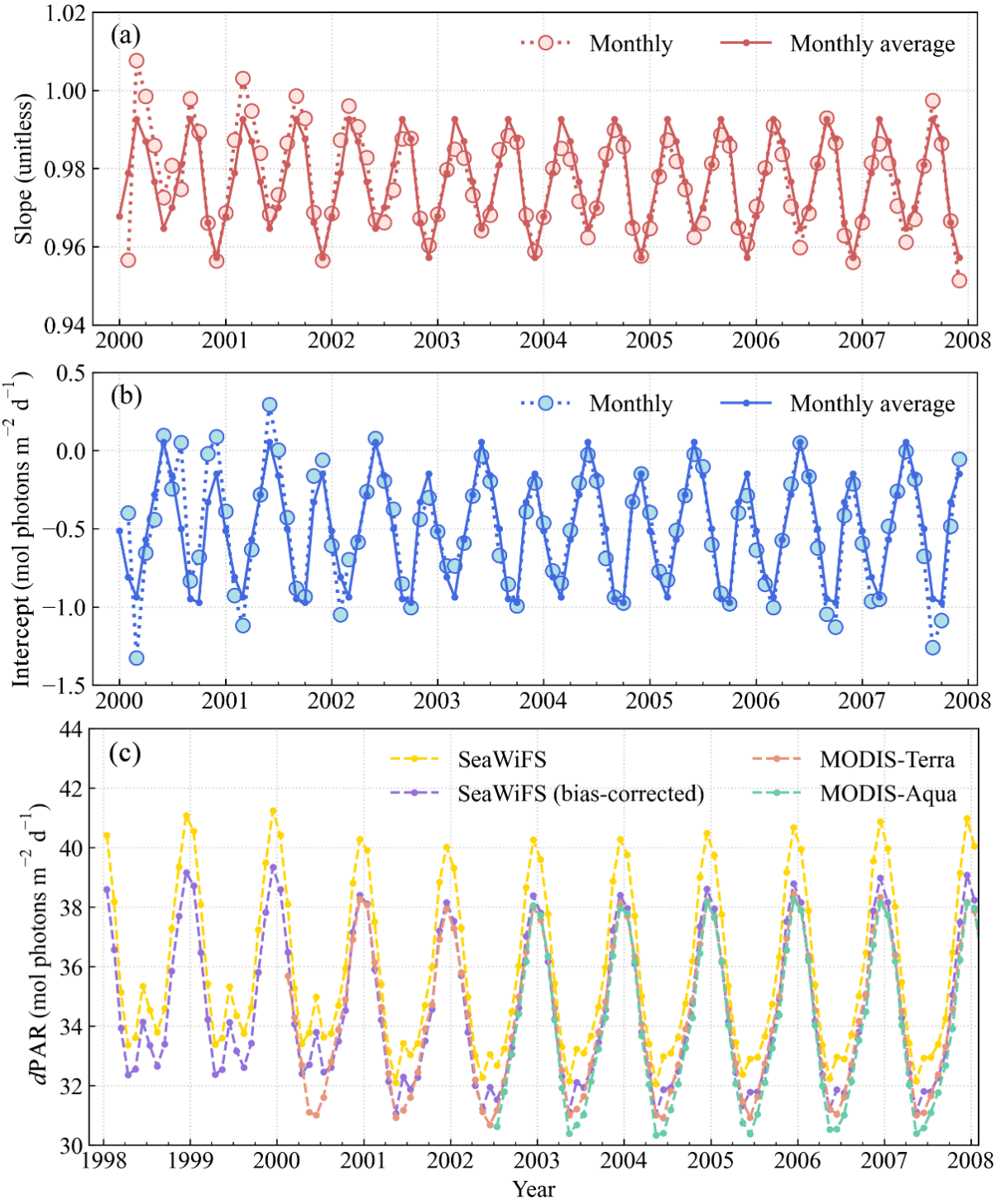

Figure 2.13. Monthly changes of the linear slopes (a) and intercepts (b) obtained by the linear regression between SeaWiFS and MODIS *d*PAR in the period 2000–2007. (c) Time series of global monthly mean *d*PAR from MODIS and SeaWiFS before and after bias-correction in the period 1998–2008.

Figure 2.14a–c demonstrates the spatial distributions of monthly mean *d*PAR from MODIS and SeaWiFS, before and after bias-correction, in Jul. 2002, while Figure 2.15a–c is the same but in Dec. 2002. The spatial characteristics of global *d*PAR correspond to latitude and season. Tropical to temperate oceans are typically dominated by high *d*PAR, and the changes of *d*PAR in polar oceans correspond to the cycle of polar day and polar night. Three spatial distributions of *d*PAR shared a very similar spatial pattern, without discernable biases. To inspect the consistency of *d*PAR distribution more explicitly, the differences of *d*PAR between SeaWiFS and MODIS (i.e., SeaWiFS *d*PAR minus MODIS *d*PAR) before and after bias-correction for SeaWiFS were exhibited in Figure 2.14d and Figure 2.14e (same as Figure 2.15d and Figure 2.15e), respectively. As expected, before bias-correction, the values of SeaWiFS *d*PAR were larger than those of MODIS for nearly all pixels. After bias-correction, the discrepancy of *d*PAR in medium to high latitude oceans (40°N to 80°N, 40°S to 80°S) was smaller, but there was still a trend of overestimation especially for polar oceans. Whereas the original overestimation observed in low to medium latitude oceans (30°S to 30°N) was generally turned to underestimation, because the linear regression was performed to all available pixels but not on a pixel-by-pixel basis. Although our bias-correction approach is simple, it was proved to be overall effective. The linear coefficients fitted for each month can be used to deal with the seasonal change of *d*PAR bias between SeaWiFS and MODIS, without distorting the spatiotemporal variability of global *d*PAR.

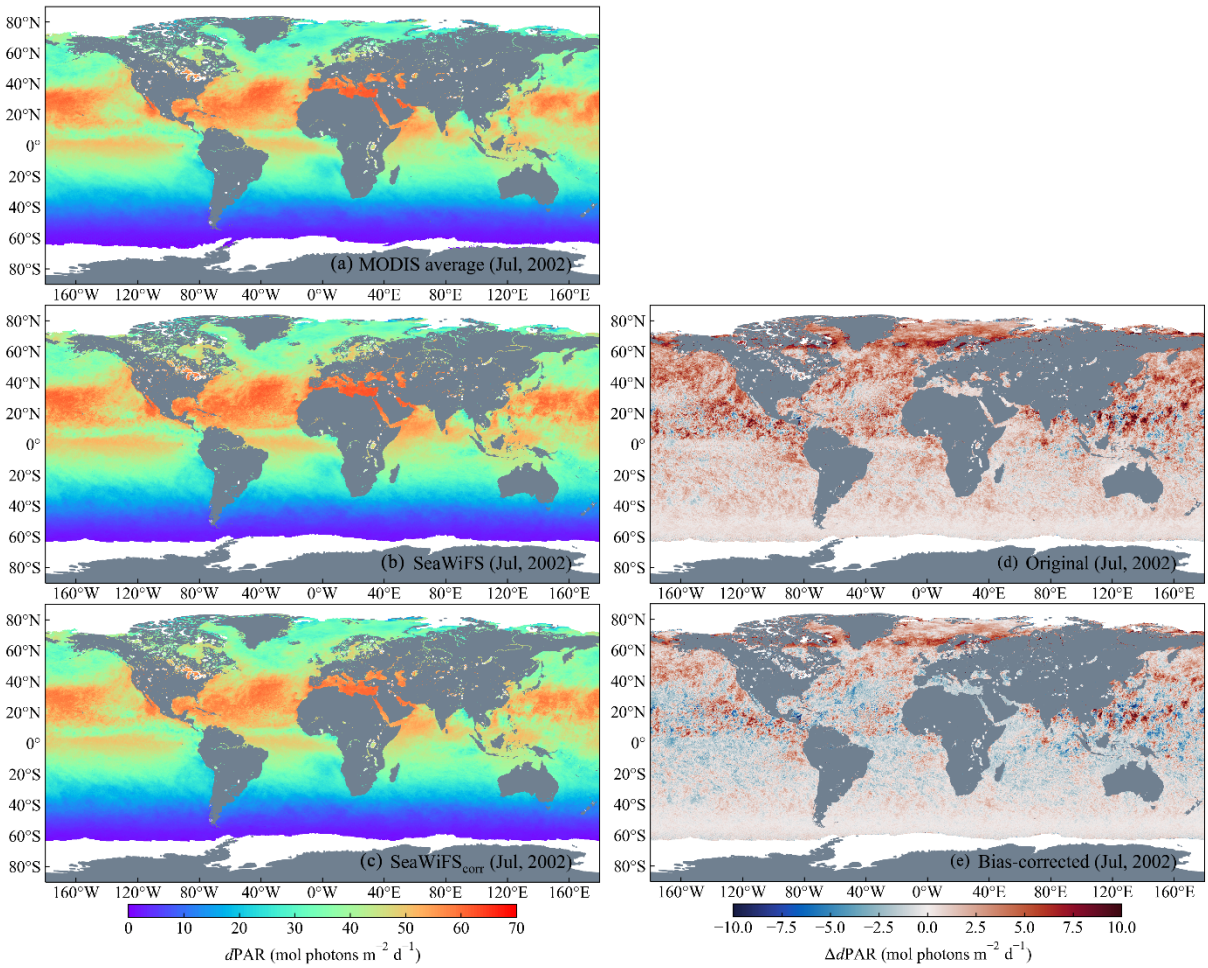

Figure 2.14. Spatial distributions of monthly mean *d*PAR from MODIS (a) and from SeaWiFS before (b) and after bias-correction (c) in Jul. 2002. The corresponding spatial distributions of *d*PAR difference between SeaWiFS and MODIS before (d) and after bias-correction (e) are also shown.

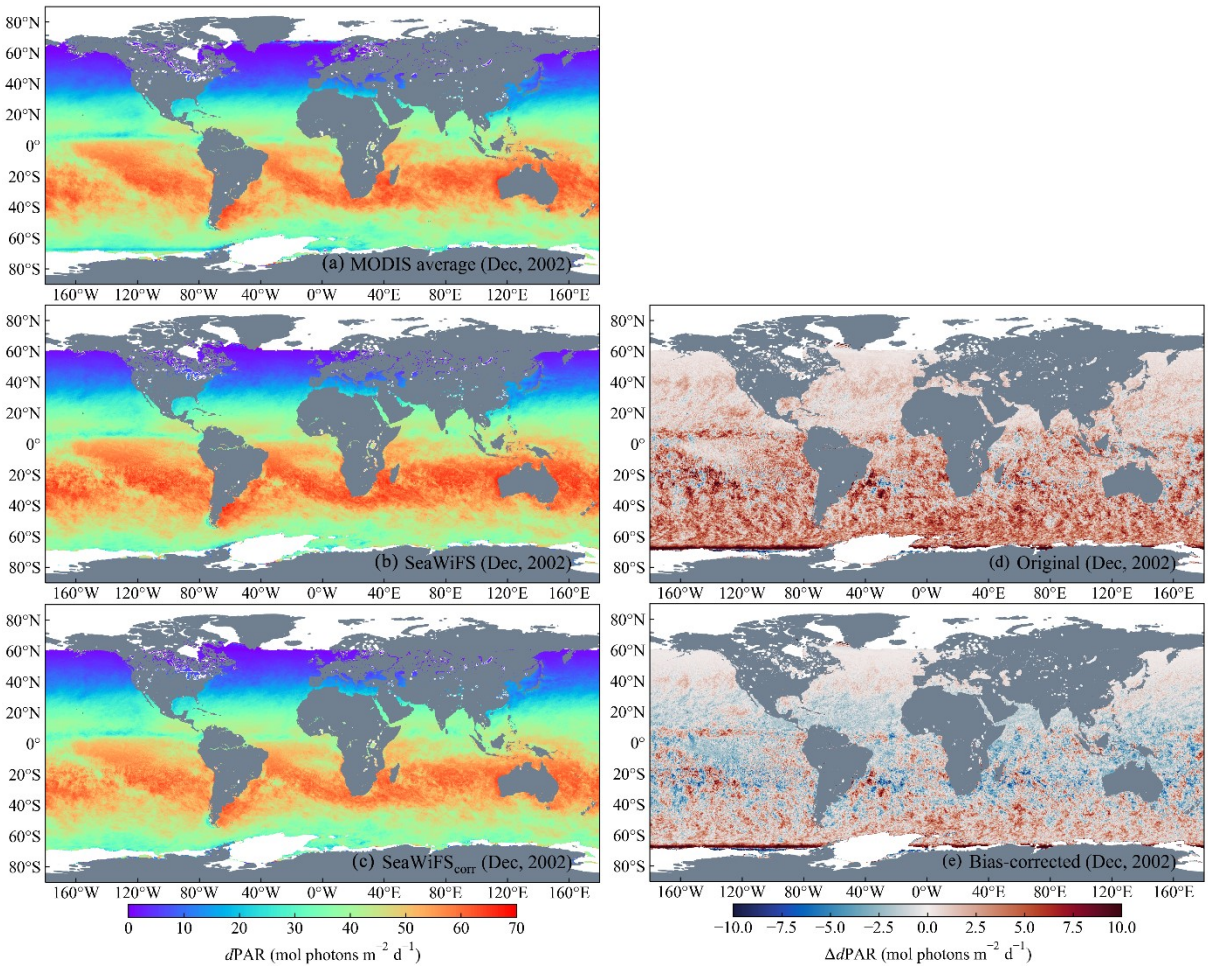

Figure 2.15. Same as Figure 2.14, but in Dec. 2002.

2.3.2. Multi-sensor SST products

The linear regression analysis was also performed on global monthly mean SST from the SST-CCI project and MODIS to investigate the systematic bias between these two products. Note that the SST-CCI project provides global merged SST observations spanning 1981–2016, aggregated from two continuous series of radiometers (Merchant et al., 2019). The SST-CCI product is of high quality, but unfortunately data beyond 2016 has not been released yet. Therefore, to cover the target time period of this study, 1998 to 2021, MODIS-Aqua/Terra SST products were used for the period 2000–2021, and SST-CCI product was used for the period 1998–2000. For this reason, MODIS SST was taken as a reference during the linear regression analysis.

The monthly changes of linear slopes and intercepts were shown in Figure 2.16a and Figure 2.16b, respectively. Seasonal fluctuation was also observed, but with very small amplitudes. The slopes ranged between 0.99 and 1.01, while the intercepts ranged between –0.4 and 0.2°C, implying that there was no significant bias between SST-CCI and MODIS SST products. Figure 2.16c displays the time series of global monthly mean SST from two products in the period 1998–2005. It was found that the SST trend of SST-CCI well agreed with those of MODIS, with or without bias-correction, and the systematic bias was generally lower than ~0.4°C. Although the SST records had a certain degree of bias even for MODIS series, the bias could be neglectable (< 0.125°C). The results suggest that the SST products from SST-CCI and MODIS can be directly used together.

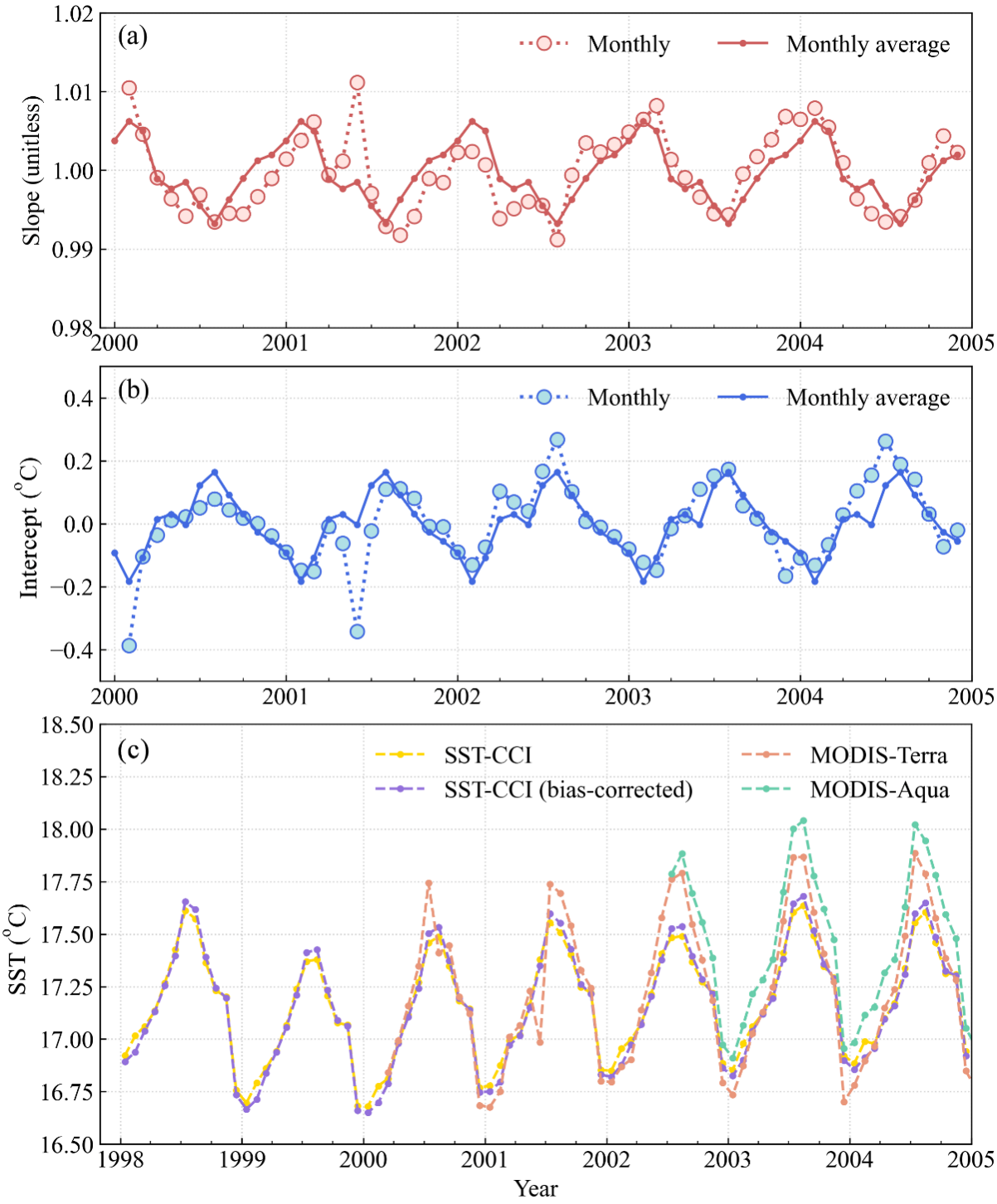

Figure 2.16. Monthly changes of the linear slopes (a) and intercepts (b) obtained by the linear regression between SST-CCI and MODIS SST in the period 2000–2004. (c) Time series of global monthly mean SST from MODIS and SST-CCI before and after bias-correction in the period 1998–2004.

Figure 2.17a and Figure 2.17b demonstrate the spatial distributions of monthly mean SST from MODIS and SST-CCI in Jul. 2002, respectively, while Figure 2.18a and Figure 2.18b are the same but in Dec. 2002. The corresponding distributions of SST difference between SST-CCI and MODIS (the former minus the latter) are shown in Figure 2.17c and Figure 2.18c. The spatial characteristics of global SST mainly correspond to the meridional distribution of solar energy (i.e., *d*PAR) and the change of sea ice coverage in high latitude water areas (such as Hudson Bay in the Arctic Ocean). High SST values (> 25°C) were typically observed in low to medium latitude oceans (30°S to 30°N), while SST values were usually lower than 10°C in high latitude oceans (poleward of 40°S or 40°N). The spatial pattern of two SST products were nearly identical. Still, underestimation of SST could be found mainly in the Northern hemisphere in boreal summer, while in the Southern hemisphere in austral summer. Different overpass times of satellite instruments and variable cloud conditions could be the reasons for the occurrence of these discrepancies. Since the bias of SST between SST-CCI and MODIS is generally within 2°C, and SST is not a direct input for IPP calculation (see Eq. (1.1)) but is required for the remote estimation of photosynthetic parameters (see Section 3.1), we can conclude that the successive use of SST-CCI and MODIS SST products without bias-correction is feasible. The potential influence of satellite-measured SST on the estimation of P^{B}_{max} and E_k based on the ERFR and the subsequent computation of IPP using the photophysiolocial model, i.e., the TPM, will still be discussed in subsequent sections.

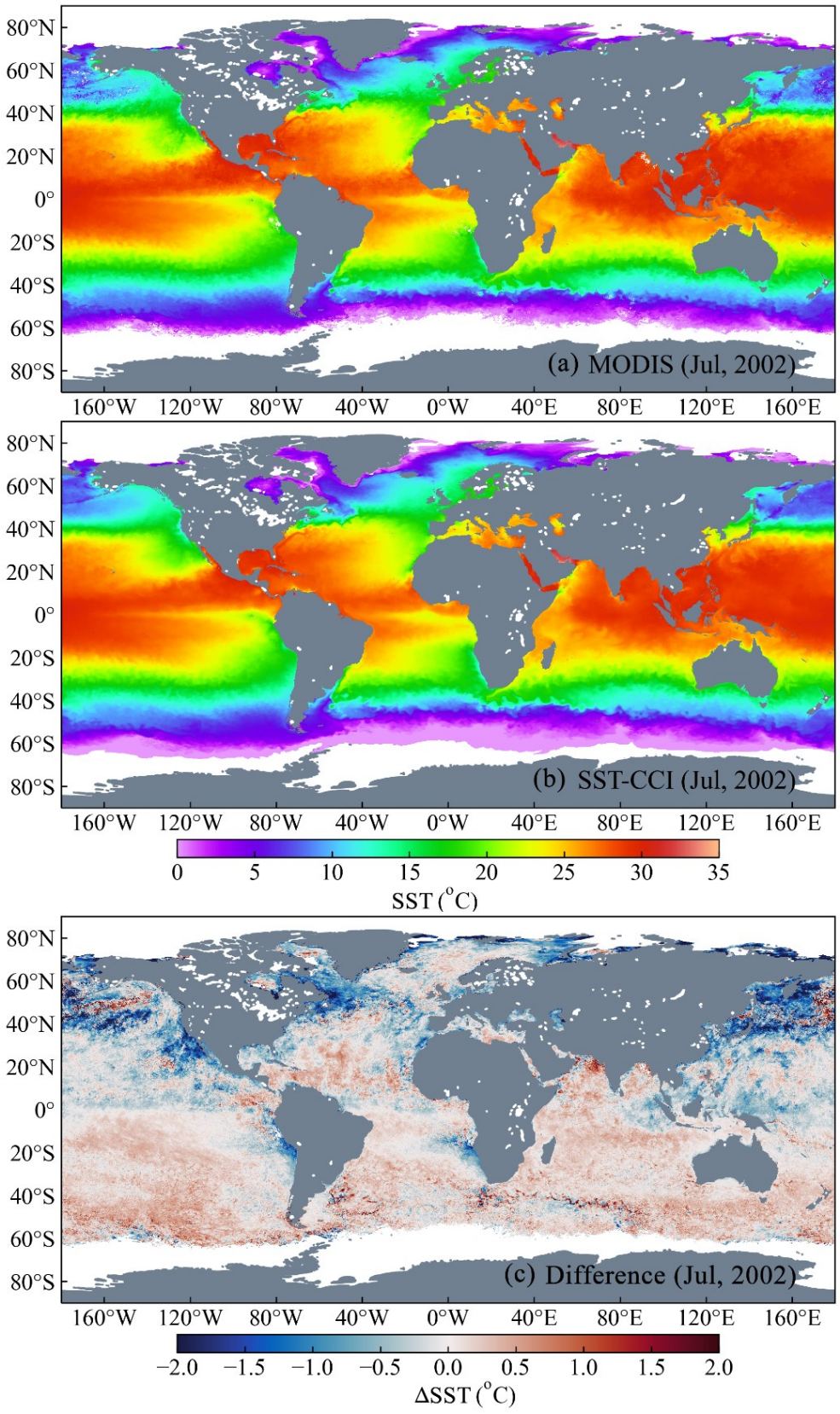

Figure 2.17. Spatial distributions of monthly mean SST (a) from MODIS and SST-CCI (b) in Jul. 2002. The corresponding distribution of SST difference are also shown (c).

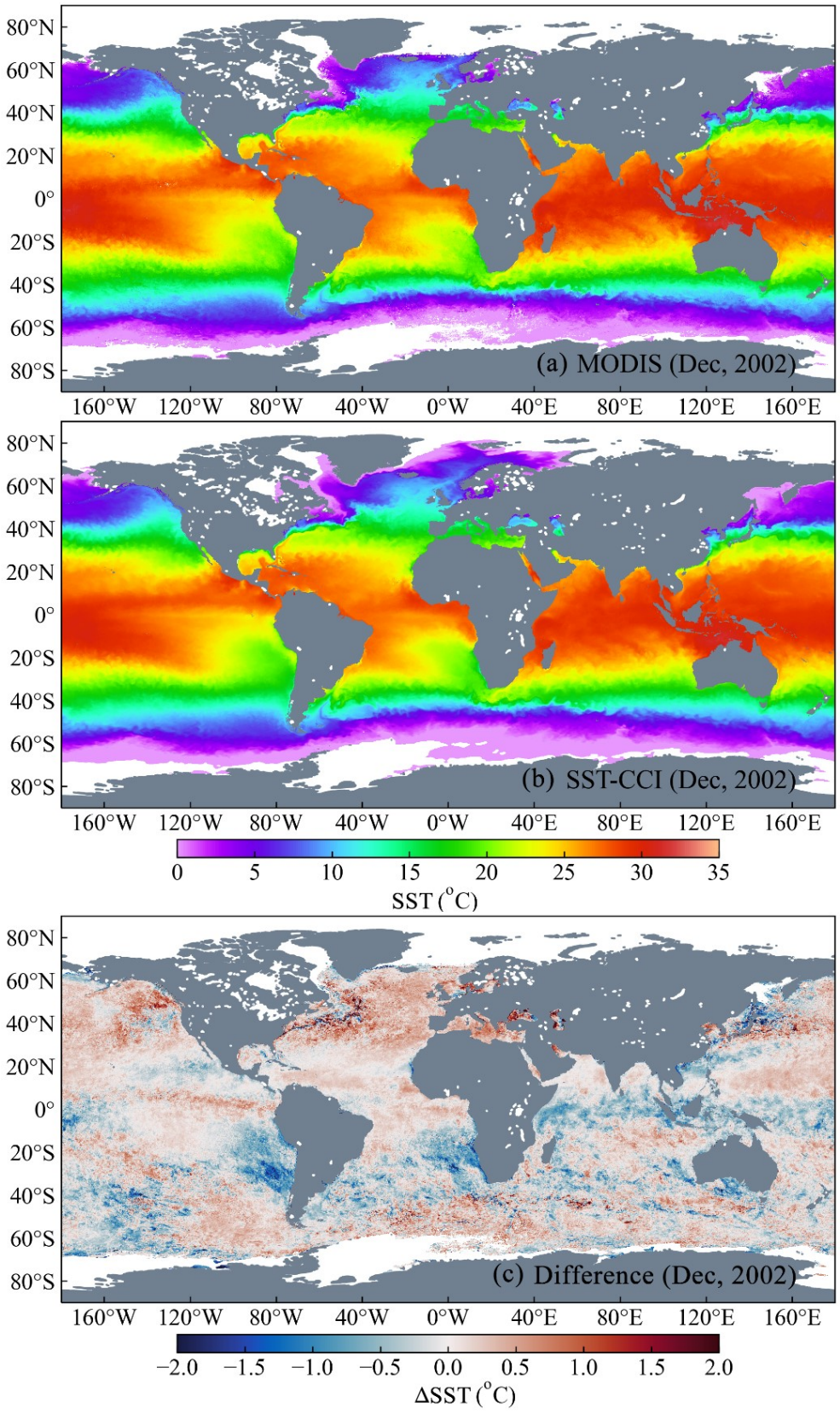

Figure 2.18. Same as Figure 2.17, but in Dec. 2002.

2.4. Additional data for inter-comparison and analysis

In addition to the forementioned *in situ*, remote sensing, and auxiliary data used for algorithm development, validation, and product generation, several additional data were obtained as supplementary for the subsequent inter-comparison and variability analysis. The summarized description of each data and corresponding usage was tabulated in Table 2.8.

2.4.1. Existing IPP products

2.4.1.1. Ocean Productivity project

Two alternative IPP products were obtained from the Ocean Productivity (OP) website (http://sites.science.oregonstate.edu/ocean.productivity/index.php). The products are operationally supported by NASA, and the input satellite parameters were gap-filled by considering the similarity of data at both spatial and temporal scales. More detailed procedure of the OP-based gap-filling can be found on its official website (http://orca.science.oregonstate.edu/gap_fill.php). The multi-year monthly IPP products generated based on the original VGPM (Behrenfeld and Falkowski, 1997b) and the carbon, absorption, and fluorescence euphotic-resolving production model (CAFE) (Silsbe et al., 2016) were used here, as two representative IPP models that have received much attention from the scientific community. Specifically, to highlight the source of products, these models were referred to as the $VGPM_{OP}$ and $CAFE_{OP}$, respectively. All IPP products span the time period from 1998 to 2021 with a spatial resolution of 9 km. From 1998 to

June 2002, the monthly gap-filled products of SeaWiFS were used to generate IPP estimates, while from July 2002 to the end of 2021, the MODIS products were used instead. It is noteworthy that, given the absence of the monthly products between Aug. and Sep. in 2020 from the OP website, the amount of monthly IPP estimates were linearly interpolated along time dimension before performing the subsequent analyses. The seasonal and interannual changes of IPP estimated by these three models were compared with the changes estimated by our model to evaluate the confidence level of satellite-revealed trends.

2.4.1.2. Biological Pump and Carbon Exchange Processes project

The Biological Pump and Carbon Exchange Processes (BICEP) project (https://bicep-project.org/) has published a new IPP product, which was generated based on a depth- and spectral-resolved model using the OC-CCI v4.2 ocean color products. The BICEP IPP product contains global monthly IPP data for the period 1998 to 2020 at 9 km spatial resolution (Kulk et al., 2021a). As a recently released product, the BICEP utilized the LUT-retrieved P^{B}_{max} and E_k together with the OC-CCI-based environmental parameters to produce global IPP data, which has certain aspects in common with the TPM$_{ERFR}$, and thereby it is appropriate to include the BICEP IPP product for inter-comparison. The detailed description of model development and application can be found in Kulk et al., 2020. Similar to the OP products, the BICEP product was also used to systematically investigate the interannual trends of IPP at basin and global scales and to diagnose whether the trends revealed by other data show a similar pattern.

Table 2.8. Summary of additional data used for inter-comparison and variability analysis, including IPP, SIE, wind speed, and climate indices.

Product	Source of input data	Resolution	Available period	Provider
Existing IPP products				
$CAFE_{OP}$	SeaWiFS/MODIS	Monthly, 9 km	1998–2021	OP (NASA)
$VGPM_{OP}$	SeaWiFS/MODIS	Monthly, 9 km	1998–2021	OP (NASA)
BICEP	OC-CCI v4.2	Monthly, 4 km	1998–2020	BICEP (ESA)
$AbPM_{GC}$	GlobColour	Monthly, 4 km	1998–2021	GlobColour (ESA)
Sea ice extent product				
SIE	SMMR, SSM/I, and SSMIS	Monthly, 25 km	1998–2021	OSISAF (EUMETSAT)
Global wind speed product				
Wind speed	SSM/I, SSMIS, WindSat, and AMSR2	Monthly, 1°	1998–2021	RSS (NASA)
Climate indices				
ONI	ERSST, version 5	Monthly	1998–2021	CPC (NOAA)
AMO	Kaplan SST, version 2	Monthly	1998–2021	PSL (NOAA)

2.4.1.3. GlobColour project

The GlobColour project (https://hermes.acri.fr/index.php) supported by ESA (Maritorena et al., 2010), which is very similar to the OC-CCI project, provides continuous multi-sensor merged ocean color products from 1997 to present, covering the mainstream parameters, such as Chl*a*, R_{rs}, suspended particulate matter (SPM), Secchi disk depth, particulate backscattering coefficient (b_{bp}), etc. In particular, the GlobColour project distributes primary production products generated by using third-party algorithms, which can be used as an independent data source for comparison against our IPP product. The GlobColour IPP product was computed based on the phytoplankton-absorption-based algorithm proposed by Antoine and Morel (1996), which was referred to as the AbPM$_{GC}$ hereafter. The Level-3 monthly IPP product with a spatial resolution of 4 km from 1998 to 2021 was downloaded on the Copernicus Marine Environment Monitoring Service (CMEMS) website (https://resources.marine.copernicus.eu/products).

2.4.2. Sea ice extent product

As one of important climate data records, the time series of sea ice extent (SIE) for the Arctic and Antarctic in the period from 1998 to 2021 produced by the EUMETSAT Ocean and Sea Ice Satellite Application Facility (OSISAF) were used to study the relationship between the changes of SIE and IPP in polar waters. The monthly SIE product is acquired from the OSISAF sea-ice index (version 2.1) data (https://osisaf-hl.met.no/v2p1-sea-ice-index), which were calculated from the climate data record of sea ice concentration (SIC, defined as the fraction of a

grid cell that is covered by sea ice, ranging between 0 and 100%) from three sensors, including the Scanning Multi-channel Microwave Radiometer (SMMR), Special Sensor Microwave/Imager (SSM/I), and Special Sensor Microwave Imager/Sounder (SSMIS) (Lavergne et al., 2019). SIE is usually defined as the total area of the grid cells where SIC exceed 15%, which is a commonly used threshold for SIC data at different spatial resolutions. The calculated climate records of SIE in each hemishpere were used to investigate the potential linkage between IPP and SIE in the Antarctic and Arctic oceans.

2.4.3. Global wind speed product

Global ocean surface wind speed product was acquired from the Remote Sensing Systems (RSS) (https://www.remss.com/measurements/wind/wspd-1-deg-product/) funded by NASA. The product was constructed using the version 7 of radiometer wind speed data processed by the RRS from the microwave sensors including SSM/I, SSMIS, WindSat polarimetric radiometer, and Advanced Microwave Scanning Radiometer 2 (AMSR2). The multi-sensor data have been well intercalibrated, and a unified method is adopted to estimate wind speeds at a height of 10 meters above the ocean surface (Wentz, 2015). The merged product contains monthly mean wind speed values at 1° spatial resolution for global oceans over the 24-year period 1998–2021 (data before 1998 are available but not used here). The wind speed product is intended for analyzing the change of IPP modulated by the physical process of air-sea interaction.

2.4.4. Climate Indices

The oceanic Niño index (ONI) is one of the most commonly-used climate indices for monitoring the climate patterns of SST annomalies in the tropical Pacific, more specifically, for tracking the occurrence of the El Niño-Southern Oscillation (ENSO). The ONI is available from the NOAA Climate Prediction Center (CPC) (https://origin.cpc.ncep.noaa.gov/). It is defined as the 3-month running average of SST annomalies calculated using the extended reconstructed SST (ERSST, version 5) in the the east-central tropical Pacific, i.e., the Niño 3.4 region (5^{o}N–5^{o}S, 120^{o}–70^{o}W). Oceanic full-fledged El Niño (warm episode) or La Niña (cold episode) events are considered to exist when the ONI exceeds 0.5^{o}C or –0.5^{o}C for at least five consecutive overlapping seasons. The ONI was used to investigate the respone of IPP (mainly in the Pacific and Indian oceans) to the seasonal shifts between warm and cold episodes due to the present of El Niño and La Niña events over the last 24 years.

In addition, the time series of Atlantic Multidecadal Oscillation (AMO) index, defined as the anomalies of SST in the Northern Atlantic Ocean after linear detrending, was available from the NOAA Physical Sciences Laboratory (PSL) (https://psl.noaa.gov/data/timeseries/AMO/), and the calculation is based on the dataset of Kaplan SST (version 2). Warm and cool phases of the AMO indicate positive and negative SST anomalies, respectively, which may last for over 20 years at a time. The 24-year record of AMO was used to explore its potential effect on the trend of IPP especially in the Atlantic Ocean.

Chapter 3. Methodology

3.1. Enhanced random forest regression (ERFR)

RFR is a classical ensemble machine-learning algorithm that takes advantage of multiple regression trees using the bootstrap aggregating (bagging) method (Breiman, 2001). The RFR has been extensively applied to remote estimation for various kinds of biogeophysical parameters (Chen et al., 2019; Shen et al., 2020). However, in spite of its high performance, the RFR is also known to be incapable of extrapolating, and consequently it cannot be used to predict results with input features that fall outside the range of training data (Zhang et al., 2019).

Regarding the retrieval of photosynthetic parameters (P^{B}_{max} and E_k), it is an overwhelmingly difficult task to collect adequte training datasets for global waters. Therefore, a modified RFR algorithm, ERFR, was adapted in this study to overcome the extrapolation limitation of the original RFR. The concept of ERFR is to enhance extrapolation capability by using multivariate linear regression (MLR) as an intermediate predictor and then to improve the performance of prediction by combining the MLR with a traditional RFR. The final ouput of the ERFR can be expressed as follows (Figure 3.1):

$$\mathbf{Y}_{fin} = \mathbf{Y}_{inter} + \mathbf{U}' = \mathbf{XB} + f_{RFR}(\mathbf{X}), \tag{3.1}$$

where the matrix $\mathbf{Y}_{fin}$ contains column vectors of estimated P^{B}_{max} and E_k, and the matrix $\mathbf{X}$ contains column vectors of input features with constants. The matrix product of $\mathbf{X}$ and $\mathbf{B}$ that contains the regression coefficients gives the intermediate

prediction matrix $\mathbf{Y}_{inter}$ returned by the MLR. The estimated residual matrix $\mathbf{U}'$ is yielded by the RFR (i.e., $\mathbf{U}' = f_{RFR}(\mathbf{X})$), which is trained using the matrix $\mathbf{X}$ and the residuals between the measured values and intermediate predictions.

The input features were selected from four candidates—T, Chl$a$, K_d(PAR), and r_{PAR}—for enabling the ERFR to mine vital information of P^B_{max} and E_k from both biotic (phytoplankton biomass) and abiotic factors (temperature and available light) that influence phytoplankton growth (Eppley, 1972; Kameda and Ishizaka, 2005; Picart et al., 2014; Sathyendranath et al., 2009). To determine the optimal ERFR, four training scenarios (denoted as S1–S4 hereafter) with different combinations of input features (S1: T; S2: T and Chl$a$; S3: T, Chl$a$, and K_d(PAR); S4: T, Chl$a$, K_d(PAR) and r_{PAR}) were tested via 10-fold cross-validations. During each round of cross-validations, the training data consisted of random samples from each of the ADDs via stratified sampling to ensure that all ADDs contributed to the training of ERFR. The vectors of P^B_{max}, E_k, Chl$a$, and K_d(PAR) were log-transformed. The matrix $\mathbf{X}$ was standardized by subtracting mean values and dividing by standard deviations to facilitate algorithm convergence. To clarify the improved extrapolation capability of the ERFR, a traditional RFR was trained via the optimal scenario for comparison. The optimal number of regression trees in the RFR was determined to be 300 through trial and error, and other hyperparameters were set to default values. The issues of overfitting and chance correlation were found to be insignificant in this study.

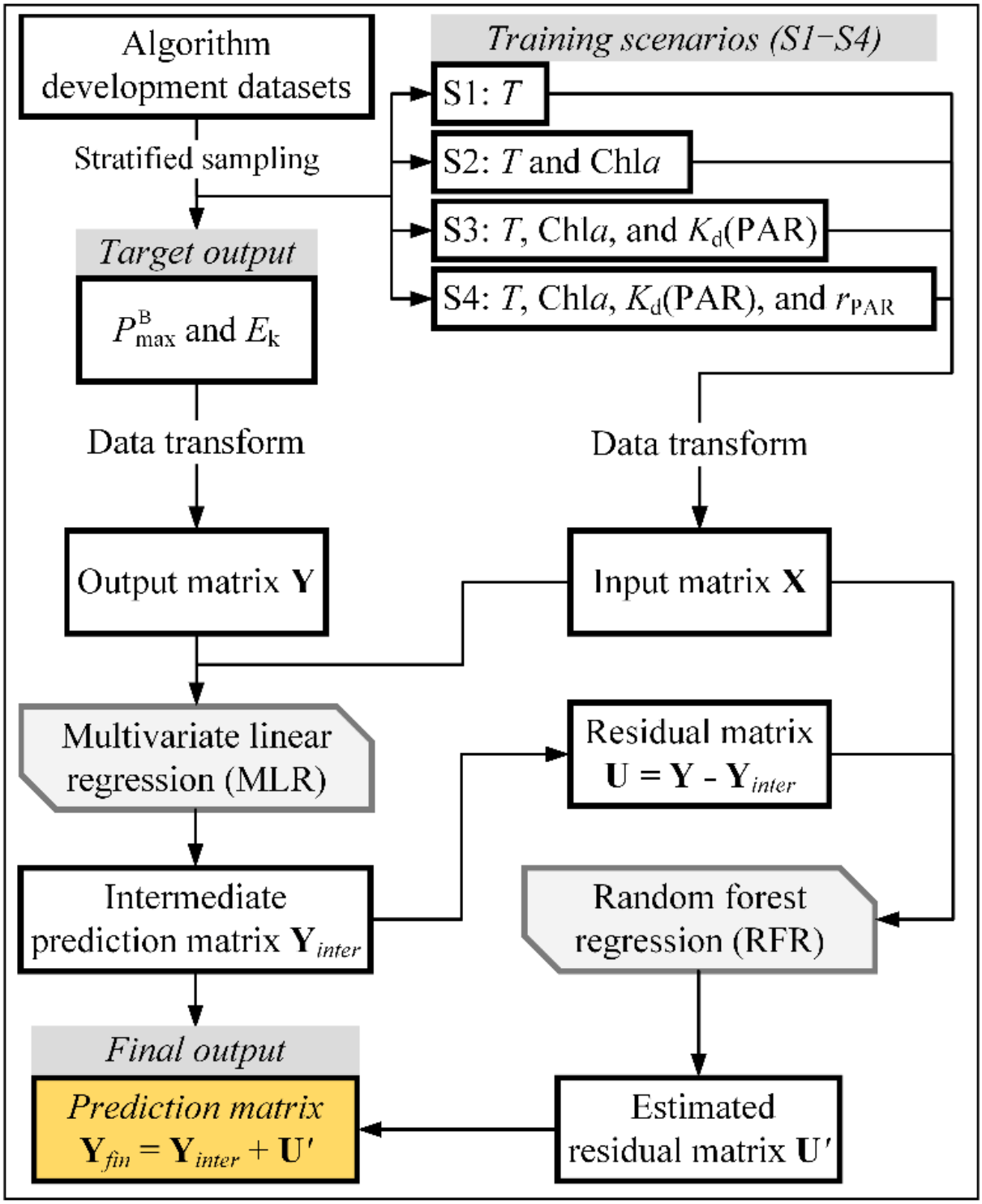

Figure 3.1. Schematic diagram of the ERFR designed in this study for the estimation of P^{B}_{max} and E_k. Four training scenarios (referred to as S1–S4) with different combinations of input candidates (*T*, Chl*a*, K_d(PAR), and r_{PAR}) were tested to determine the optimal ERFR.

3.2. IPP computation: basic and remote sensing schemes

3.2.1. Basic scheme for *in situ* data

Computation of *in situ* IPP (in mg C m^{-2} d^{-1}) based on the TPM can be expressed as a double integration of PP(z, t) (Eq. (1.1)) over the day length (DL, in hours) and the water depth from the surface to Z_{eu}:

$$\mathrm{IPP} = \int_0^{\mathrm{DL}} \int_0^{Z_{eu}} \mathrm{Chl}a(z) \times P^{\mathrm{B}}_{\max}(z) \tanh\left[\frac{\mathrm{PAR}(z,\ t)}{E_{\mathrm{k}}(z)}\right] \mathrm{d}z\mathrm{d}t, \tag{3.2}$$

where DL is estimated as a function of latitude and day of the year. The time-dependent PAR(z, t) (in μmol photons m^{-2} s^{-1}) was represented by using dPAR and a sine function that described the diurnal change of incident light (Platt et al., 1990) and by substitution into Eq. (2.2):

$$\mathrm{PAR}(z,\ t) = \frac{\pi d\mathrm{PAR}(0^-)}{0.0072\mathrm{DL}} \sin\left(\frac{\pi t}{\mathrm{DL}}\right) \exp\left[-K_{\mathrm{d}}(\mathrm{PAR}) \times z\right], \tag{3.3}$$

where the amplitude of the sine function is the maximum PAR at local noon and was derived from dPAR(0^-) (equals 0.95×dPAR) and DL. Note that the potential time dependences of Chl$a$, Z_{eu}, $P^{\mathrm{B}}_{\max}$, and E_{k} were ignored in this model. The double integration in Eq. (3.2) was computed using the trapezoidal rule. In practice, we set 13 increments of time from 0 to DL, and 21 increments of depth from 0 to Z_{eu}. Because the depth dependence of Chl$a$, $P^{\mathrm{B}}_{\max}$, and E_{k} were assumed to be negligible in inland waters, these three parameters remained constant at all depths during the computaion. For oceanic waters, in contrast, *in situ* IPP data

were calculated using measured profiles of Chl*a*, P^{B}_{max}, and E_k. The actual characteristics throughout the euphotic zone could not be well captured for some profiles because there were not enough valid samples. The IPP data from those profiles were therefore not included in subsequent comparative analyses.

3.2.2. Remote sensing scheme

Figure 3.2 shows the general workflow for producing satellite-based P^{B}_{max}, E_k, and IPP for different water types. For inland waters, the retrievals were obtained by directly substituting satellite input parameters into the TPM$_{ERFR}$ following the basic computation scheme. In contrast, the remote sensing scheme for oceanic waters is dependent upon the stratification of water columns. A profile with a ratio of Z_m to Z_{eu} that exceeded 1 was recognized to be vertically homogeneous, and satellite Chl*a* and SST data were used for all depths. Otherwise, the profile was presumed to be non-uniform, and the method of Uitz et al. (2006) was adopted to estimate the vertical profile of Chl*a* based on a shifted Gaussian function. The vertical integration was made within the productive zone between 0 and 1.5Z_{eu} instead of the euphotic zone to better account for the deep chlorophyll maximum (DCM) (Antoine et al., 1996; Brewin et al., 2017; Uitz et al., 2010). The satellite-based vertical profile of *T* was estimated based on GODAS assimilated data. For each site of oceanic waters, the GODAS monthly *T* profile was extracted and normalized by subtracting the surface value and interpolated into 21 depth intervals within the productive zone. The daily *T* profile could then be restored by adding the satellite-measured SST to the normalized profile.

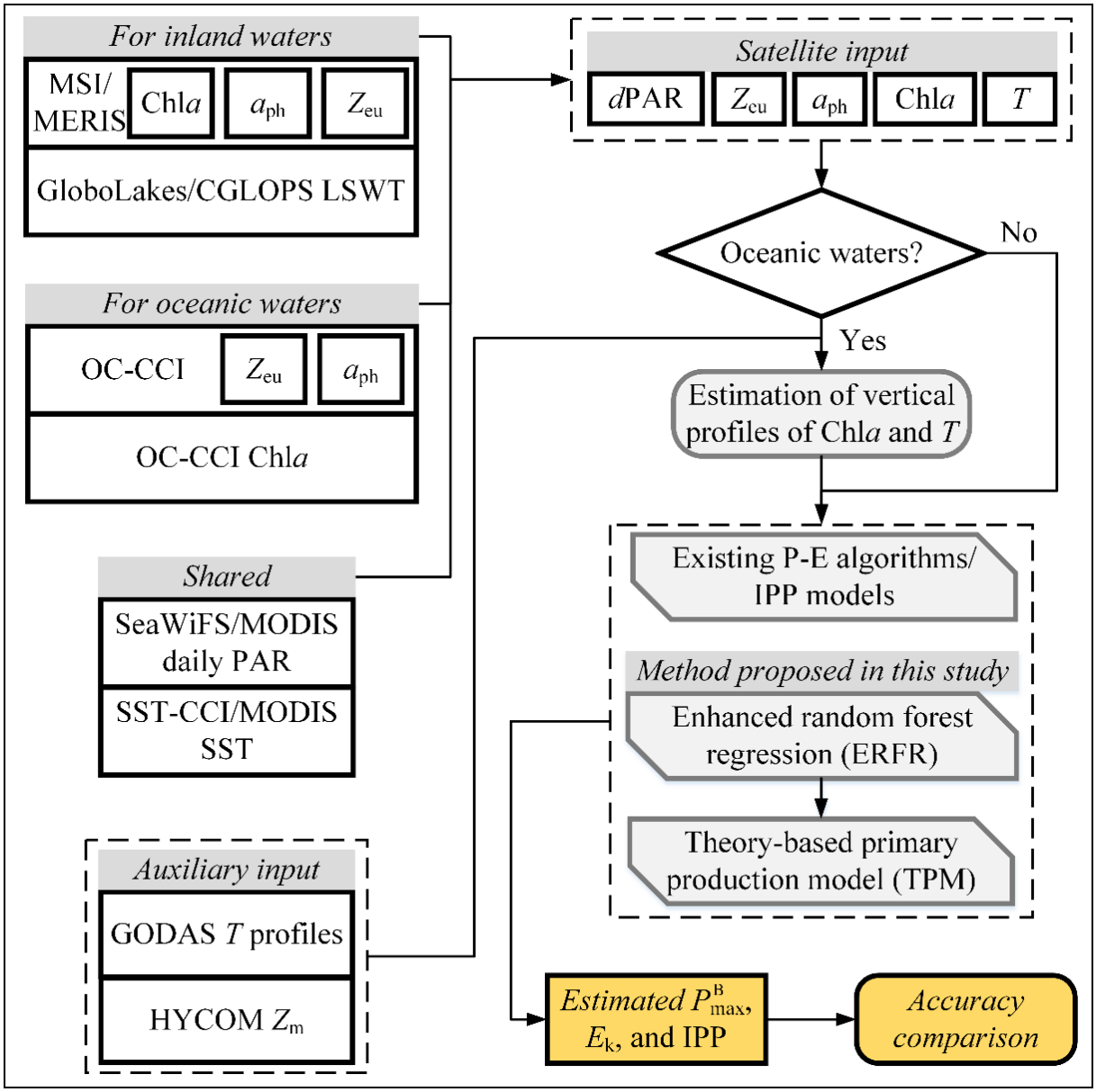

Figure 3.2. The diagram for estimating P^B_{max}, E_k, and IPP in cases of oceanic and inland waters using the methods proposed in this study and the ones published in previous studies.

3.3. Existing methods for comparative analysis

Several existing methods were used for comparative analysis to evaluate the performances of our methods and the existing ones for P^{B}_{max}, E_k, and IPP estimation in different water types. Table 3.1 provides the detailed information of the existing methods selected here (our methods were also included for reference), and note that the method selection is not exhaustive but representative. In detail, four existing methods were introduced to retrieve photosynthetic parameters (referred to as P-E algorithms) for the comparison with the ERFR, and five IPP models that were designed on basis of different concepts and model structures were employed for the comparison with the TPM_{ERFR}.

3.3.1. Existing P-E algorithms

The M96 describes P^{B}_{max} and E_k as power functions of T; the P14 and S16 are more complex and consider the available light within the mixed layer of the ocean. To apply the P14 and S16 to the RRCB and LKSM data, Z_m was assumed to be equated to Z_{eu}. The LUT20 is a look-up-table (LUT) method that assigns P^{B}_{max} and E_k to a geographic position based on the biogeographic province of Longhurst (Longhurst, 2007) (as shown in Figure 3.3) and season (i.e., Mar. to May for spring (autumn), Jun. to Aug. for summer (winter), Sep. to Nov. for autumn (spring), and Dec. to Feb. for winter (summer) in the Northern (Southern) hemisphere) to which it belongs. Because the LUT20 is restricted to the ocean, a similar LUT named the LUT_RK, was created using the long-term data from the

Rhode River and Lake Kasumigaura for each site and season. The details of each introduced P-E algorithms were provided as follows.

3.3.1.1. M96 (Morel et al., 1996)

$$P_{\max}^{\mathrm{B}} = 4.6 \times 1.065^{T-20}, \tag{3.4.1}$$

$$\mathrm{KPUR} = 80 \times 1.065^{T-20}, \tag{3.4.2}$$

$$E_{\mathrm{k}} = \mathrm{KPUR}\ \frac{a_{\mathrm{ph,max}}^{*}}{\overline{a}_{\mathrm{ph}}^{*}}, \tag{3.4.3}$$

where KPUR is an analogous parameter to E_{k}, which is used when available irradiance is described by photosynthetic usable radiation (PUR) instead of PAR. KPUR can be converted to E_{k} according to Eq. (3.4.3) using the maximum value of Chl*a*-specific $a_{\mathrm{ph}}(\lambda)$ ($a_{\mathrm{ph}}^{*}(\lambda)$), which is usually equal to $a_{\mathrm{ph}}^{*}(443)$, and the spectrally-averaged $a_{\mathrm{ph}}^{*}(\lambda)$ in visible domain ($\overline{a}_{\mathrm{ph}}^{*}$).

3.3.1.2. P14 (Picart et al., 2014)

$$P_{\max}^{\mathrm{B}} = \left[\frac{1}{\mathrm{Chl}a} 10^{[1.81+0.63 log_{10}(\mathrm{Chl}a)]}(1-0.023T)(1+0.037E_{\mathrm{ML}})\right]\mu_{\max}, \tag{3.5.1}$$

$$\mu_{\max} = 0.851(1.066^{T})\frac{ln2}{24}\left(\frac{\mathrm{Chl}a}{\mathrm{Chl}a+0.094}\right), \tag{3.5.2}$$

$$E_{\mathrm{ML}} = \frac{d\mathrm{PAR}}{K_{\mathrm{d}}(\mathrm{PAR})\times Z_{\mathrm{m}}}\left[1-\mathrm{e}^{-K_{\mathrm{d}}(\mathrm{PAR})\times Z_{\mathrm{m}}}\right], \tag{3.5.3}$$

where E_{ML} is the average light in the mixed layer, and $\mu_{\max}$ is the maximum growth rate.

Table 3.1. Summary of existing methods for estimating photosynthetic parameters (P-E algorithms) and IPP (IPP models) used for comparative analysis. The proposed ERFR and TPM*ERFR* are also included for reference. The output, required input, method types, and corresponding reference are given.

No.	Method	Output		Input parameters						Type*	Reference
		P^{B}_{max}	E_k	T	Chl*a*	Z_{eu}	a_{ph}	Z_m	dPAR		
P-E algorithms											
/	ERFR	○	○	○	○	○				ML	This study
1	M96	○	○	○						EM	Morel et al., 1996
2	P14	○		○	○	○		○	○	SA	Picart et al., 2014
3	S16		○			○		○	○	SA	Silsbe et al., 2016
4	LUT20/ LUT_RK	○	○	(Location & season)						LUT	Kulk et al., 2020/This study
IPP models											
/	TPM*ERFR*	IPP		○	○	○		○	○	DR, TR	This study
1	TPM*LUT20*/ TPM*LUT_RK*				○	○		○	○	DR, TR	Kulk et al., 2020/This study
2	AbPM*AM96*			○	○	○	○	○	○	DR, TR	Antoine and Morel, 1996
3	VGPM*BF97*			○	○	○			○	DI, TI	Behrenfeld and Falkowski, 1997b
4	VGPM*KI05*			○	○	○			○	DI, TI	Kameda and Ishizaka, 2005
5	VGPM*H02*			○	○	○			○	DI, TI	Harding et al., 2002

* ML: Machine learning, EM: Empirical, SA: Semi-analytical, LUT: Look-up table, DR: Depth-resolved, DI: Depth-integrated, TR: Time-resolved, TI: Time-integrated.

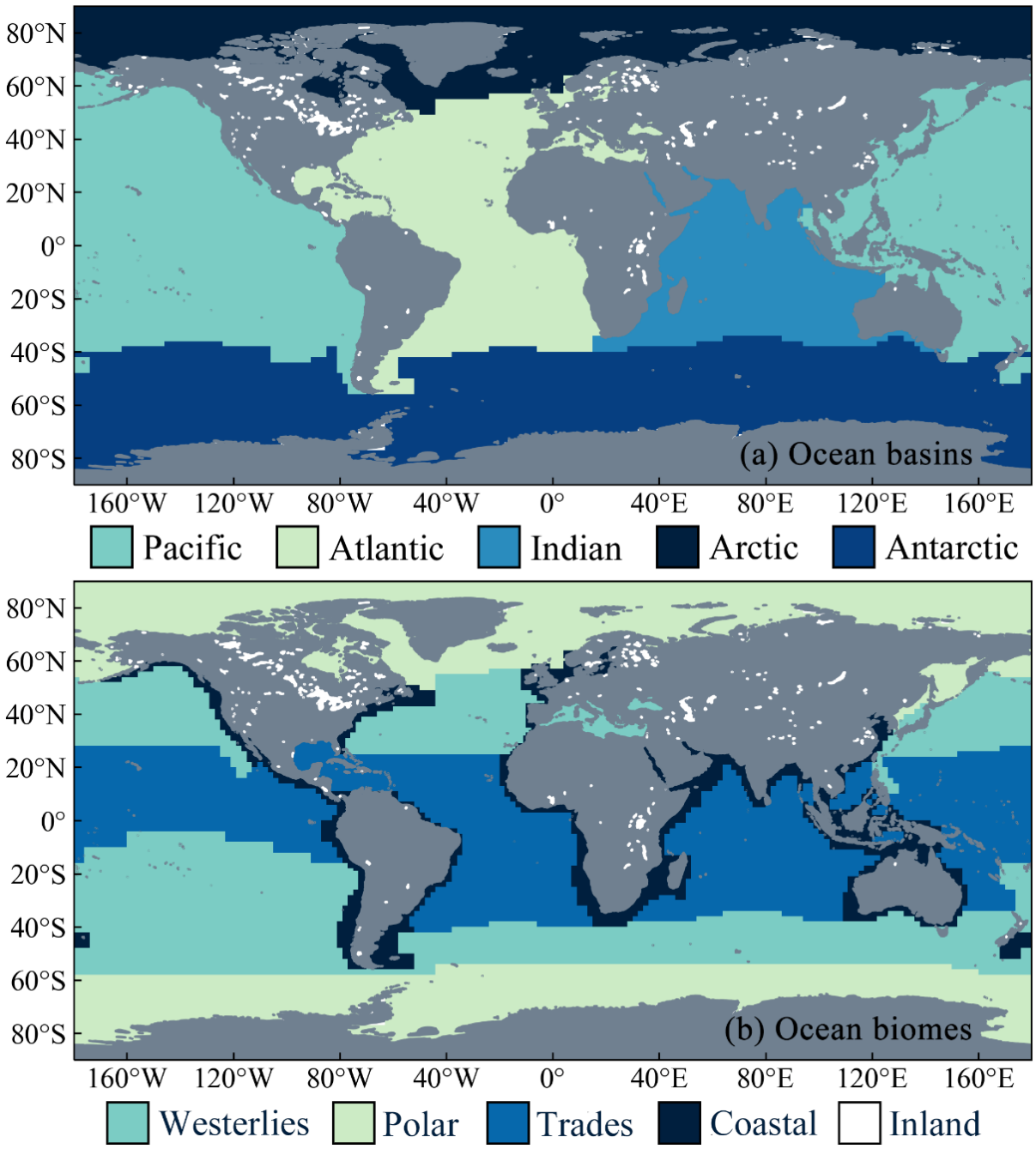

Figure 3.3. The distributions of ocean basins (a) and biomes (b) defined by Longhurst (2007).

3.3.1.3. S16 (Behrenfeld et al., 2016; Silsbe et al., 2016)

$$E_{\mathrm{k}}(Z_{\mathrm{m}} > Z_{\mathrm{eu}}) = 0.0864 \times 19\mathrm{e}^{0.038(d\mathrm{PAR}/\mathrm{DL})^{0.45}/K_{\mathrm{d}}(\mathrm{PAR})}, \tag{3.6.1}$$

$$E_{\mathrm{k}}(Z_{\mathrm{m}} \le Z_{\mathrm{eu}}) = 0.0864 \times 19\mathrm{e}^{0.038(d\mathrm{PAR}/\mathrm{DL})^{0.45}/K_{\mathrm{d}}(\mathrm{PAR})} \times \frac{1+\mathrm{e}^{-0.15d\,\mathrm{PAR}/\mathrm{DL}}}{1+\mathrm{e}^{-3I_{\mathrm{ML}}}}, \tag{3.6.2}$$

$$E_{\mathrm{k}}(z > Z_{\mathrm{m}}) = 0.0864 + \frac{E_{\mathrm{k}}(Z_{\mathrm{m}}) - 0.0864}{d\mathrm{PAR}(Z_{\mathrm{m}}) - 0.1} \times \left[d\mathrm{PAR}(z) - 0.1\right], \tag{3.6.3}$$

$$I_{\mathrm{ML}} = d\mathrm{PAR}/\mathrm{DL}\,\mathrm{e}^{-0.5K_{\mathrm{d}}(\mathrm{PAR}) \times Z_{\mathrm{m}}}, \tag{3.6.4}$$

where I_{ML} is the median light in the mixed layer (Behrenfeld et al., 2005). E_{k} was converted to μmol photon m^{-2} s^{-1} from mol photon m^{-2} d^{-1} by dividing by 0.0864. The lower bound of E_{k} is set as 10 μmol photon m^{-2} s^{-1} (Silsbe et al., 2016).

3.3.1.4. LUT_20/LUT_RK (Kulk et al., 2020; This study)

The LUT published in Kulk et al. (2020) was directly adopted as LUT20, while LUT_RK was built using the RRCB and LKSM datasets in this study.

3.3.2. Existing IPP models

The LUT20 and LUT_RK, like the ERFR, can also serve as upstream P-E estimators for the TPM, therefore two IPP models, namely the TPM_{LUT20} and TPM_{LUT_RK}, were added for comparisons against the TPM_{ERFR}. An a_{ph}-based IPP model (AbPM_{AM96}) and the original VGPM (VGPM_{BF97}) as well as its two variants (i.e., VGPM_{KI05} and VGPM_{H02}) were also introduced. Model complexity ranged from depth- (DI) and time-integrated (TI) models (i.e., the VGPM-like models) to depth- (DR) and time-resolved (TR) models (i.e., the TPM_{LUT20}, TPM_{LUT_RK},

and AbPM$_{AM96}$) that were run in the same way as the TPM$_{ERFR}$ (Section 3.2.2). For simplicity, the AbPM$_{AM96}$ was run in a wavelength-integrated manner. It should be noted that no attempt was made to re-calibrate these P-E algorithms and IPP models in this study. The details of each introduced IPP models were provided as follows.

3.3.2.1. TPM$_{LUT20}$/TPM$_{LUT_RK}$ (Kulk et al., 2020; This study)

IPP computation using TPM$_{LUT20}$ or TPM$_{LUT_RK}$ is in the same manner as the TPM$_{ERFR}$ (Eq. (3.2)), however the depth-dependence of $P^{\mathrm{B}}_{\max}$ and E_{k} were ignored because the LUT20 and LUT_RK only contain the information on surface waters.

3.3.2.2. AbPM$_{AM96}$ (Antoine and Morel, 1996)

AbPM$_{AM96}$ is originally a wavelength-resolved model, but it was run in wavelength-integrated manner as follows:

$$\mathrm{IPP} = 12\overline{a}^{*}_{\mathrm{ph,max}}\varphi_{\mu,\max}\int_{0}^{\mathrm{DL}}\int_{0}^{Z_{\mathrm{eu}}}\mathrm{Chl}a(z)\times\mathrm{PUR}(z,\,t)\tanh\left[\frac{\mathrm{KPUR}(z)}{\mathrm{PUR}(z,\,t)}\right]\mathrm{d}z\mathrm{d}t, \tag{3.7.1}$$

$$\alpha'^{\mathrm{B}} = 12\overline{a}^{*}_{\mathrm{ph,max}}\varphi_{\mu,\max} = 16\ \mathrm{mg\ C\ (mg\ Chl)^{-1}\ (mol\ photons\ m^{-2})^{-1}}, \tag{3.7.2}$$

$$\mathrm{PUR}(z,\,t) = \mathrm{PAR}(z,\,t)\overline{\mathrm{A}}^{*}, \tag{3.7.3}$$

$$\overline{\mathrm{A}}^{*} = \frac{\overline{a}^{*}_{\mathrm{ph}}}{a^{*}_{\mathrm{ph,max}}}, \tag{3.7.4}$$

where α'^{B} is the initial slope of the photosynthesis-irradiance curve when available irradiance is described by PUR instead of PAR, which was set as a

global constant that is valid for Chl*a* ranging between 0.03–3 mg m^{-3}, and $\overline{\mathrm{A}}^{*}$ is the dimensionless normalized $\bar{a}^{*}_{\mathrm{ph}}$ ranging between 0 and 1. KPUR was estimated by Eq. (3.4.2).

3.3.2.3. VGPM$_{BF97}$ (Behrenfeld and Falkowski, 1997)

$$\mathrm{IPP} = 0.66125 \times P^{\mathrm{B}}_{\mathrm{opt}} \times \frac{d\mathrm{PAR}}{d\mathrm{PAR} + 4.1} \times Z_{\mathrm{eu}} \times \mathrm{Chl}a \times \mathrm{DL}, \tag{3.8.1}$$

$$P^{\mathrm{B}}_{\mathrm{opt}} = \begin{cases} 0, \ T < -10\,^{\circ}\mathrm{C} \\ 1.13, \ -10 \le T < -1\,^{\circ}\mathrm{C} \\ \sum_{i=0}^{7} a_i T^i, \ -1 \le T \le 28.5\,^{\circ}\mathrm{C} \\ 4, \ T > 28.5\,^{\circ}\mathrm{C} \end{cases}, \tag{3.8.2}$$

where T is sea surface temperature (SST), and the coefficients a_0–a_7 are 1.2956, 0.2749, 0.0617, –0.0205, 2.462×10^{-3}, -1.348×10^{-4}, 3.4132×10^{-6}, and -3.27×10^{-8}, respectively.

3.3.2.4. VGPM$_{KI05}$ (Kameda and Ishizaka, 2005)

Same as Eq. (3.8.1), but with

$$P^{\mathrm{B}}_{\mathrm{opt}} = \frac{\sum_{i=1}^{3} a_i T^i}{\mathrm{Chl}a} + \sum_{i=0}^{4} b_i T^i, \tag{3.9}$$

where the coefficients a_1–a_3 are 0.072, -3.2×10^{-3}, and 3.0×10^{-5}, respectively, and the coefficients b_0–b_4 are 1.0, 0.17, -2.5×10^{-3}, and 8.0×10^{-5}, respectively. This new parameterization of $P^{\mathrm{B}}_{\mathrm{opt}}$ is valid for T ranging between 0–30°C and for Chl*a* > 0.05 mg m^{-3}.

3.3.2.5. VGPM$_{H02}$ (Harding et al., 2002)

$$\begin{aligned} y &= a_0 + a_1 \times log_{10}(\mathrm{Chl}a) + a_2 \times log_{10}\left[d\mathrm{PAR}/(d\mathrm{PAR}+4.1)\right] + \\ &\quad a_3 \times log_{10}(Z_{\mathrm{eu}}) + a_4 \times log_{10}(\mathrm{DL}) + a_5 \times log_{10}(T) \\ \mathrm{IPP} &= 10^{y} \end{aligned} \tag{3.10}$$

where the regression coefficients a_0–a_5 are 0.1619, 0.7721, 2.0344, 0.8115, 0.0342, and 1.2817, respectively.

3.3.3. Accuracy assessment

Three statistical measures were used to evaluate the agreement between estimated (y_i') and measured (y_i) data in subsequent sections, including the root mean square difference (*RMSD*), the unbiased absolute percentage difference (*UAPD*), and the logarithmic bias (*Bias*). Here, y_i' refers to the *i*-th estimated photosynthetic parameters, e.g. $P^{\mathrm{B}}_{\mathrm{max}}$ and E_{k}, by the ERFR and the existing P-E algorithms, or the *i*-th estimated IPP by the TPM$_{ERFR}$ and the existing IPP models. These metrics are defined as follows:

$$RMSD = \sqrt{\frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}}\left[log_{10}(y_i') - log_{10}(y_i)\right]^2}, \tag{3.11}$$

$$UAPD = \frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}}\left|\frac{2(y_i' - y_i)}{y_i' + y_i}\right| \times 100\% , \tag{3.12}$$

$$Bias = (10^{Z} - 1) \times 100\%, \text{ where } Z = \frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}}\left[log_{10}(y_i') - log_{10}(y_i)\right], \tag{3.13}$$

where N is the total number of data involved in the calculation. The *RMSD* is a measure of overall algorithm skill and indicates a higher skill if the algorithm has a lower *RMSD*. The *UAPD* is similar to the *RMSD*, but it considers the uncertainty

in measured data and can be more robust for small values. The *Bias* assesses the extent to which y_i' differs from y_i. Model-II regression analysis was used to produce the fitted line and the coefficient of determination (R^2) to gauge the correlation between y_i' and y_i in log_{10}-space.

3.4. Generation of long-term global IPP product

3.4.1. Description of the implementation of TPM_{ERFR}

Global monthly products of ocean color (OC-CCI), SST (SST-CCI/MODIS), and *d*PAR (SeaWiFS/MODIS) together with the auxiliary products of Z_m (HYCOM) and three-dimensional water temperature (GODAS) (Table 2.4 and Table 2.6) over the period 1998–2021 were used to generate IPP product for global waters based on the TPM_{ERFR}, which were then used to analyze the spatiotemporal variability of IPP at different scales. HYCOM and GODAS products were resampled to a spatial resolution of 4 km, in alignment with other input products. In addition, to save computer memory and improve computation efficiency, the climatological monthly product of GODAS (1998–2021) was generated in advance and used for global IPP computation for each month.

The TPM_{ERFR} was run on a pixel-by-pixel basis according to the remote sensing scheme for oceanic waters (i.e., identifying stratified waters, estimating vertical profiles, and integrating within the productive zone), even for some inland waters (such as the Great Lakes) where the satellite data were available because of their relatively large areas. For the pixels that had valid ocean color, SST, and

*d*PAR data but lacked Z_m and/or assimilated *T* profile data (typically for those pixels in coastal and inland water areas), the TPM_{ERFR} was run on the assumption that these pixels had vertically homogeneous profiles. As such, the TPM_{ERFR} was practically applied to global waters to generate long-term satellite IPP product. Because the global monthly products of environmental parameters from multiple satellite sensors are very likely to have a certain degree of missing data, especially for ocean color (Chl*a* and Z_{eu}) products, a gap-filling algorithm was required to fill in the missing data in the generated IPP product.

3.4.2. Gap-filling procedure for IPP product

A widely-used gap-filling algorithm based on the empirical orthogonal functions (EOFs), known as the data interpolating empirical orthogonal function (DINEOF) (Alvera-Azcárate et al., 2005; Beckers and Rixen, 2003), was employed to fill in the data gap of the generated IPP product because of the contamination of cloud or cloud shadow and the failure of atmospheric correction as well as other issues. The performance of DINEOF has been well validated, such as the applications to generate cloud-free products of Chl*a* and Secchi disk depth in the Bohai Sea (Guo et al., 2022) or SST and Chl*a* in the Northern South China Sea (Ma et al., 2021). Recently, given its proven effectiveness and convenience, the DINEOF has been operationally adopted to generate global daily gap-free Chl*a* product for the Visible Infrared Imaging Radiometer Suite (VIIRS) missions (Liu and Wang, 2019, 2018). As such, the DINEOF was thought to be a suitable choice here. In general, the DINEOF reconstructs the missing data by iteratively

decomposing the initial data matrix (**X**) of spatial and temporal dimensions $m \times n$ ($m > n$) through singular value decomposition (SVD) until the difference between the validation data, which are valid data randomly chosen from **X** (about 3%) and set aside before the procedure, and the reconstructed data reaches its minimum. The decomposition can be formulated as

$$\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathrm{T}}, \tag{3.14}$$

where **S** is the singular values matrix with dimension $r \times r$, and r is the rank of the matrix ($r \leq \min(m, n)$). The matrixes **U** and **V** contain the spatial and temporal modes with dimensions $m \times r$ and $n \times r$, respectively. During the iteration, the first p EOFs ($p = 1, 2, 3, \ldots$) are successively included for the reconstruction, and the difference between estimated and known values will be calculated (as *RMSD*). Once the optimal number of EOFs, k, is determined, which corresponds to the minimum *RMSD*, the iteration will stop, and the missing data with indices i and j in **X**, $\mathbf{X}_{i,j}$, will be finally reconstructed as

$$\mathbf{X}_{i,j} = \sum_{p=1}^{k} s_p (\mathbf{u}_p)_i (\mathbf{v}_p^{\mathrm{T}})_j, \tag{3.15}$$

where $\mathbf{u}_p$ and $\mathbf{v}_p$ are the p-th column of **U** and **V**, respectively, and s_p is the corresponding singular value. Then, the flagged missing data in **X** will be replaced with the estimated ones, while other data that are originally known will remain intact. In practice, the missing data in **X** were set as zero, and **X** was normalized by subtracting the temporal mean after log-transformation. Since a powerful computation capability is necessary for performing the DINEOF directly on a

global product with a 4 km spatial resolution, the gap-filling procedure was run repeatedly for seven zonal sections divided equally with a latitude interval of ~20°. This process will not cause a distortion of global gap-free maps because information of geographical location is not required in the DINEOF.

Because of different lifetimes of ocean color sensors used in the OC-CCI version 5.0 project (Figure 3.4a), the global data coverage ranged between ~65% and 88% before Jul. 2002 when only the data from SeaWiFS were available, and it increased to the range between ~80% and 98% when the data from subsequent ocean color sensors were successively released in public, including MODIS-Aqua, MERIS, VIIRS onboard the Suomi National Polar-orbiting Partnership (SNPP), and OLCI-S3A (Figure 3.4b). In contrast, after the DINEOF-based gap-filling, the overall global data coverage was improved to the range between ~90% to 100% thorough the entire period of interest. To demonstrate the performance of DINEOF more explicitly, a case study was performed in the area of 115–145°E and 20–45°N, covering the marginal seas of China and Japan that have complex weather condition. The spatial distributions of monthly IPP in four selected months of 1998 (Dec., Jan., Feb., and Mar.) before and after gap-filling were shown in Figure 3.5. There are typically more missing data observed in these four months (winter) than other months for this area. It was found that the missing data in the Yellow Sea (YS), Eastern China Sea (ECS), and Japan Sea (JS) was successfully retrieved without abnormal artificial patches, implying a reasonable performance of the DINEOF preliminarily.

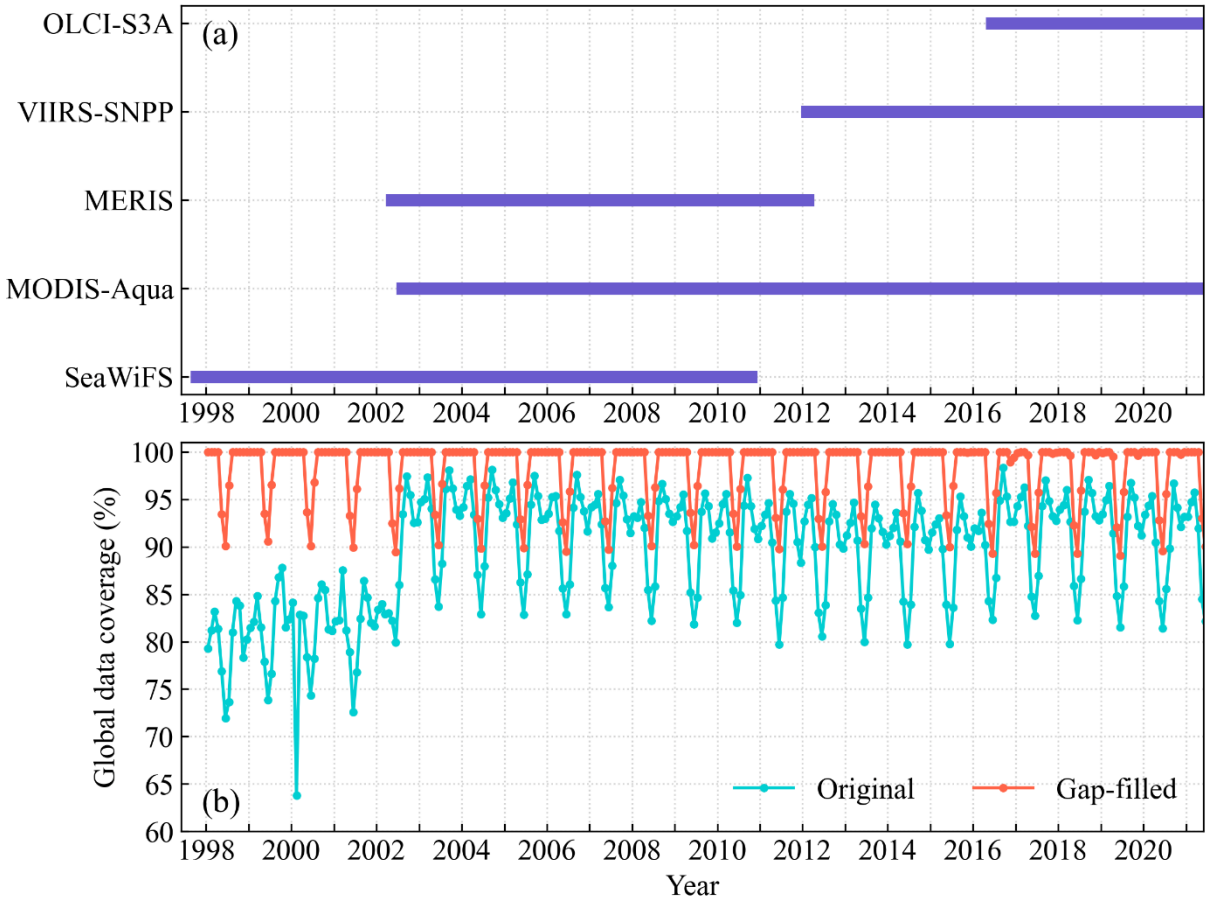

Figure 3.4. (a) The historical (SeaWiFS and MERIS) and current (MODIS-Aqua, VIIRS-SNPP, and OLCI-S3A) ocean color sensors that are included in the OC-CCI v5.0 products and their respective available time periods. (b) The global data coverage (%) (equals the ratio of total valid pixels to total water pixels; monthly sea ice mask of OSISAF and land mask were used to determine total water pixels) of the IPP product before (original) and after DINEOF-based gap-filling (gap-filled) for oceanic and inland water pixels.

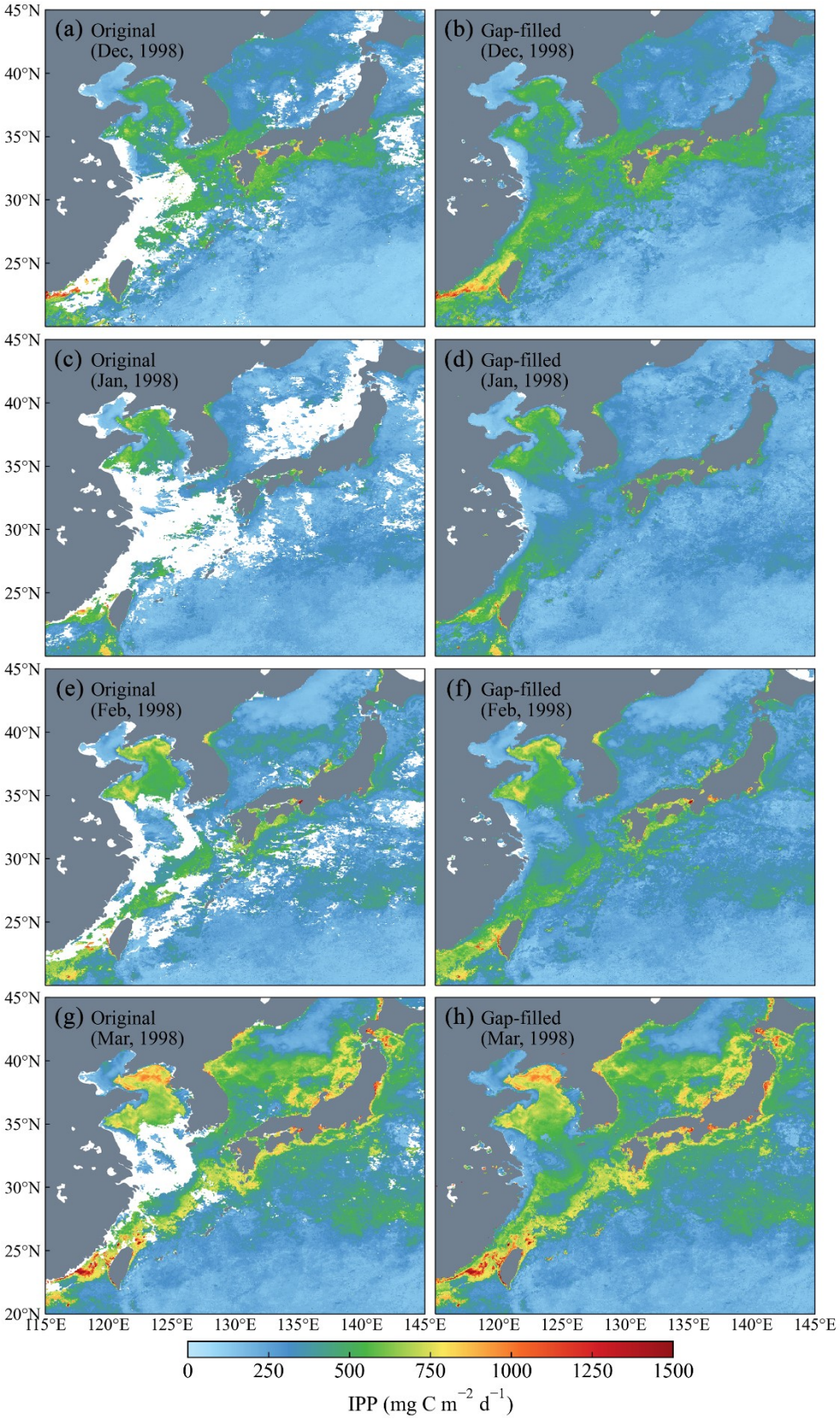

Figure 3.5. Demonstration the original (right panel) and gap-filled (left panel) IPP distributions in the area of 115–145°E and 20–45°N in Dec. (a and b), Jan. (c and d), Feb. (e and f), and Mar. (g and h) 1998, to visually inspect the performance of DINEOF.

Furthermore, the assessment result of cross-validation for this case study was shown in Figure 3.6, and the summarized accuracy statistics were tabulated in Table 3.2. For each month, the scatter points well distributed along 1:1 line without apparent bias, which corroborates the satisfactory performance of DINEOF. The values of R^2 ranged from 0.78 to 0.96, showing a high correlation between known and estimated IPP values. Besides, the values of *RMSD* generally stabilized at ~0.08, and the values of *UAPD* were lower than 15.4% with small values of *Bias*. Given its successful application in the marginal seas of China and Japan, the DINEOF were implemented on our global IPP product. Figure 3.7 shows the original and gap-filled global distributions of IPP in Jan., Apr., Jul., and Oct. 1998. As expected, only SeaWiFS is not enough to support a global gap-free observation. There were typically missing data observed in the eastern tropical Atlantic Ocean and the tropical Pacific Ocean. The data in high latitude oceans were also eroded by cloud contamination. After the gap-filling, the spatial pattern of IPP was more complete, with reasonable characteristics restored by the DINEOF. Note that the seasonal variation of upper and lower latitude bounds of all valid pixels was confirmed beforehand to rule out the sections with no data at all in each month for, on the one hand, saving computation time, and on the other hand, avoiding undesired estimation in polar oceans. In conclusion, the DINEOF is effective to produce gap-free IPP data at different spatial scale with reasonable performance, and thus the estimated total IPP and temporal trends at basin and global scales can be more realistic.

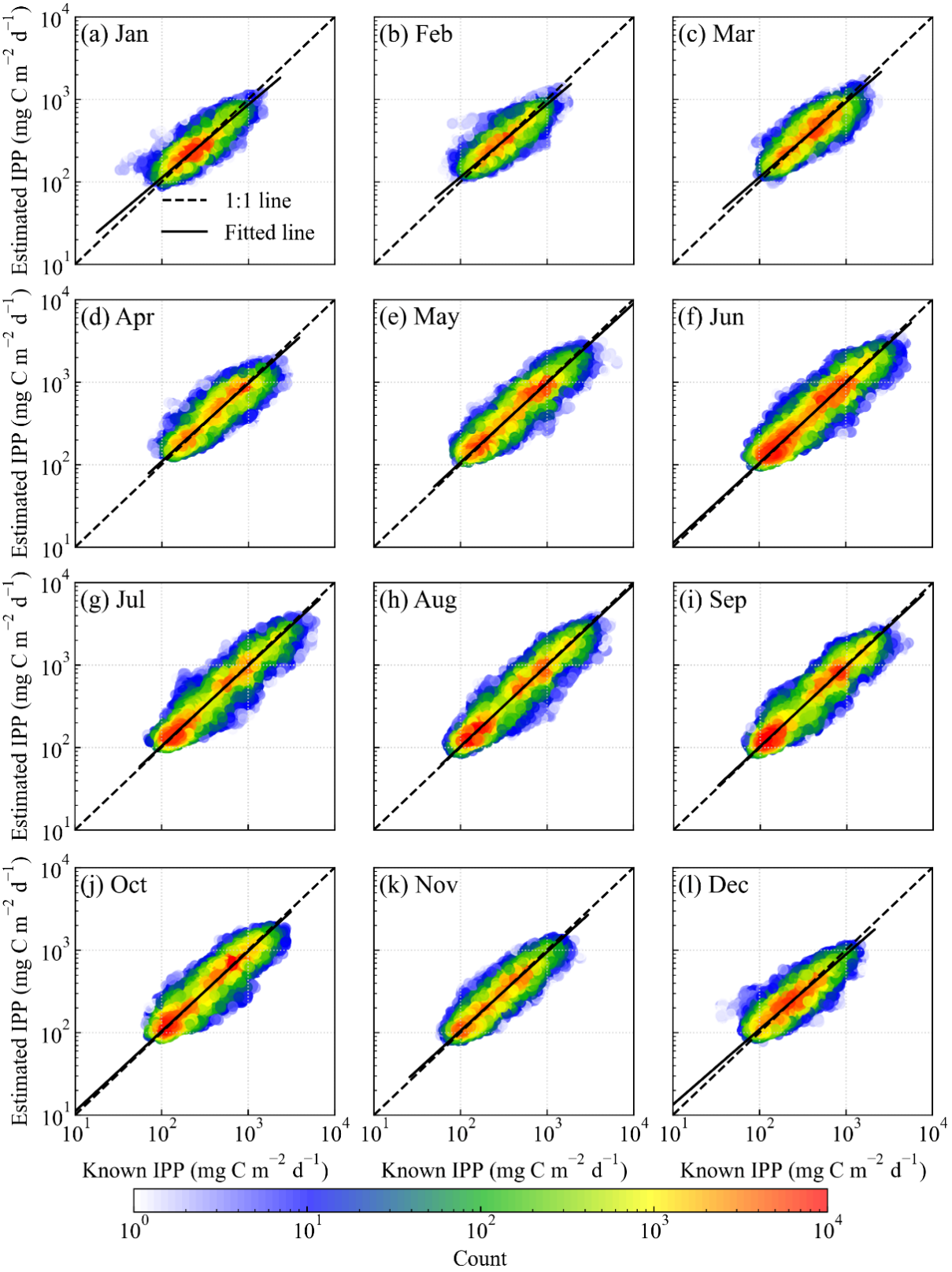

Figure 3.6. Scatter plots of the known IPP values set aside for cross-validation *versus* the estimated values based on the DINEOF for each month from 1998 to 2021.

Table 3.2. Accuracy statistics of the estimated IPP values used for cross-validation in each month over the period from 1998 to 2021.

Month	R^2	*RMSD*	*UAPD* (%)	*Bias* (%)	N
1	0.79	0.08	13.66	0.22	185568
2	0.78	0.09	14.61	0.23	192120
3	0.83	0.09	15.07	0.16	206088
4	0.89	0.09	15.43	0.02	210768
5	0.93	0.09	14.74	−0.01	211656
6	0.94	0.08	14.37	0.01	209400
7	0.95	0.09	14.55	−0.03	212544
8	0.96	0.08	13.06	0.05	214800
9	0.96	0.08	13.52	−0.01	214584
10	0.94	0.08	14.22	−0.02	213696
11	0.91	0.08	14.03	0.06	207864
12	0.83	0.09	14.81	0.22	192864

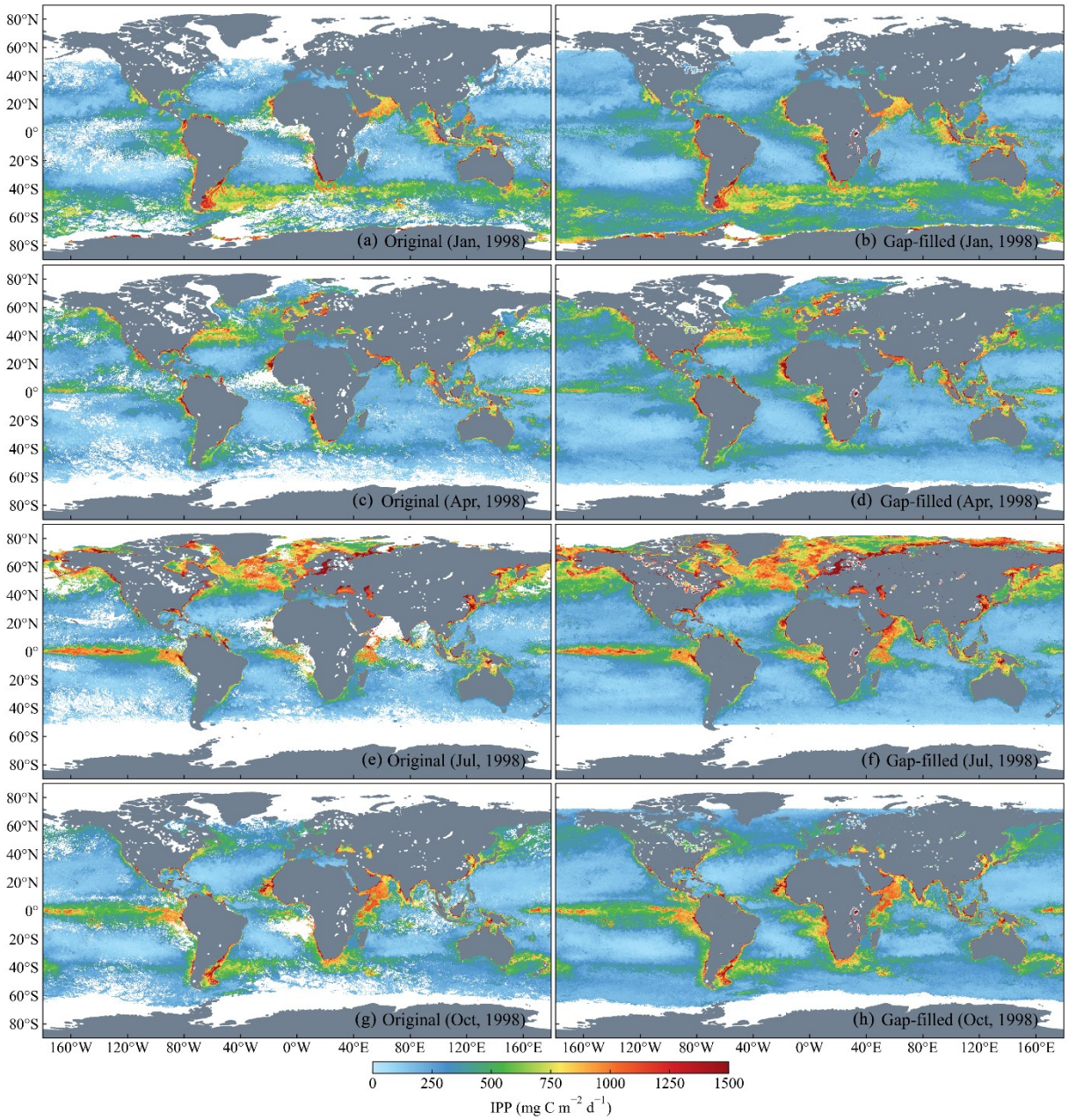

Figure 3.7. Spatial distributions of the original IPP (left panel) generated by the TPM$_{ERFR}$ and the gap-filled IPP (right panel) using the DINEOF in Jan. (a and b), Apr. (c and d), Jul. (e and f), and Oct. (g and h) 1998.

3.5. Analysis methods

3.5.1. Sensitivity analysis

The sensitivity of the TPM_{ERFR} to the uncertainties in the required input parameters obtained from satellite observations were evaluated by perturbation experiments using the ADDs. The input parameters we considered were T, Chl$a$, and K_d(PAR). A maximum uncertainty of 50% was assgined to all input parameters based on the matchup assessments for both water types (Section 2.2.3 and 2.2.4). Specifically, relative errors that varied between ±50% (with 11 steps) were used to perturb T, Chl$a$, and K_d(PAR). To analyze the sensitivity to each input parameter, only one of T, Chl$a$, or K_d(PAR) was randomly perturbed at a time (n.b., r_{PAR} would be perturbed simultaneously with K_d(PAR)) as follows:

$$\mathrm{M}^{err} = \mathrm{M} \cdot \boldsymbol{G}(\delta), \tag{3.16}$$

where M^{err} and M stand for the error-introduced and error-free measurements for one of three parameters, respectively. δ refers to the assigned relative error, and $\boldsymbol{G}$ is a Gaussian distribution function with a mean of 1 + δ and a standard deviation of 0.01. For each input parameter, the perturbation experiment was repeated 100 times for each of the eleven relative errors. The resultant uncertainties (ε) were investigated using the median of the unbiased percentage difference between the estimated (y_i^{err}) and measured values (y_i) as follows:

$$\varepsilon = median\left\{\left|\frac{2(y_i^{err} - y_i)}{y_i^{err} + y_i}\right|\right\} \times 100\%,\ i = 1,\ 2,\ \mathrm{L}\ ,\ \mathrm{N}. \tag{3.17}$$

3.5.2. Time series decomposition and trend detection

The technique of time series decomposition (TSD) is frequently used to split a given time series, such as the one that consists of monthly observations of CO_2, into the components of trend, seasonality, and noise. By means of TSD, we can easily remove the seasonality of a time series, and hence we can focus on its temporal trend. Here, the additive manner of TSD was employed, which decompose a time series based on the following equation:

$$y_{obs} = y_{trend} + y_{season} + y_{noise} \tag{3.18}$$

where y_{obs}, y_{trend}, y_{season}, and y_{noise} are the time series of observations and the components of trend, seasonality, and noise, respectively. To demonstrate the application of TSD, the long-term monthly observations of atmospheric CO_2 measured at Mauna Loa, Hawaii, was used (https://gml.noaa.gov/ccgg/). The result shows that the atmospheric CO_2 increased monotonically from 1998 to 2021 with an obvious seasonal fluctuation (Figure 3.8). TSD was implemented on the long-term monthly IPP, ocean color, and other environmental products to analyze the interannual trends of IPP with other parameters and to explain the underlying mechanisms at different scales.

Moreover, Mann-Kendall trend test (MKT) (Mann, 1945) was adopted to detect whether a given time series has a monotonic trend (consistently increasing or decreasing) and yield a robust estimate of trend slope. MKT was performed on the trend components of time series after TSD.

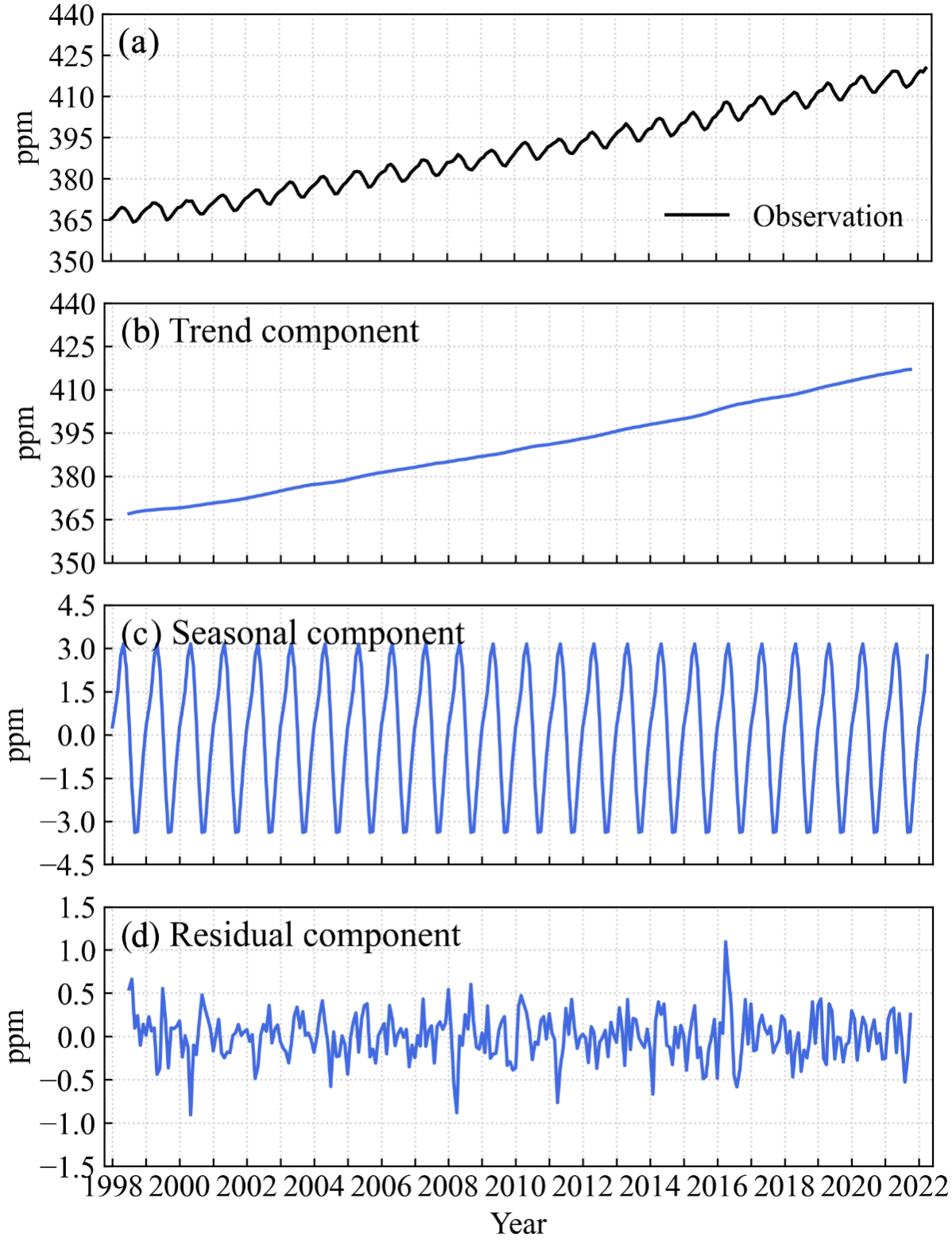

Figure 3.8. (a) Time series of monthly mean atmospheric CO_2 data (1998–2021) measured at Mauna Loa Observatory, Hawaii. The components of trend (b), seasonal (c), and residual (d) of CO_2 time series were obtained by TSD.

Chapter 4. Model evaluation

4.1. Performance evaluation of the ERFR

4.1.1. Algorithm development and independent validation

Through 10-fold cross-validations using the ADDs, the performances of individual ERFRs trained by four scenarios (S1–S4) were examined and compared against those of a traditional RFR and four selected existing P-E algorithms to determine the optimal ERFR. Algorithm accuracies were evaluated based on the testing data in each round of cross-validations (i.e., one-tenth of the data from each ADD) together with independent validation data from the AVDs.

Figure 4.1 shows the evaluation results for each algorithm and water type. The accuracies of both P^{B}_{max} and E_k retrievals gradually increased when T, Chl$a$, K_d(PAR), and r_{PAR} were successively included in the developments of ERFR according to S1–S4, as indicated by the decreasing trend of *RMSD*. The decrease of *RMSD* was significant for both oceanic and inland waters. The optimal ERFR was thus confirmed to be the one trained by S4 and hereafter will be simply referred to as the ERFR. The RFR was also trained by S4 for comparison, and it exhibited a pattern that was similar to that of the ERFR for oceanic waters (Figure 4.1a and Figure 4.1b) but failed to accurately estimate P^{B}_{max} and E_k for inland waters (in particular, for Lake Kasumigaura), especially for E_k (Figure 4.1c). The robustness of the ERFR for different water types evidenced by these results explain its improved extrapolation capability compared with a traditional RFR.

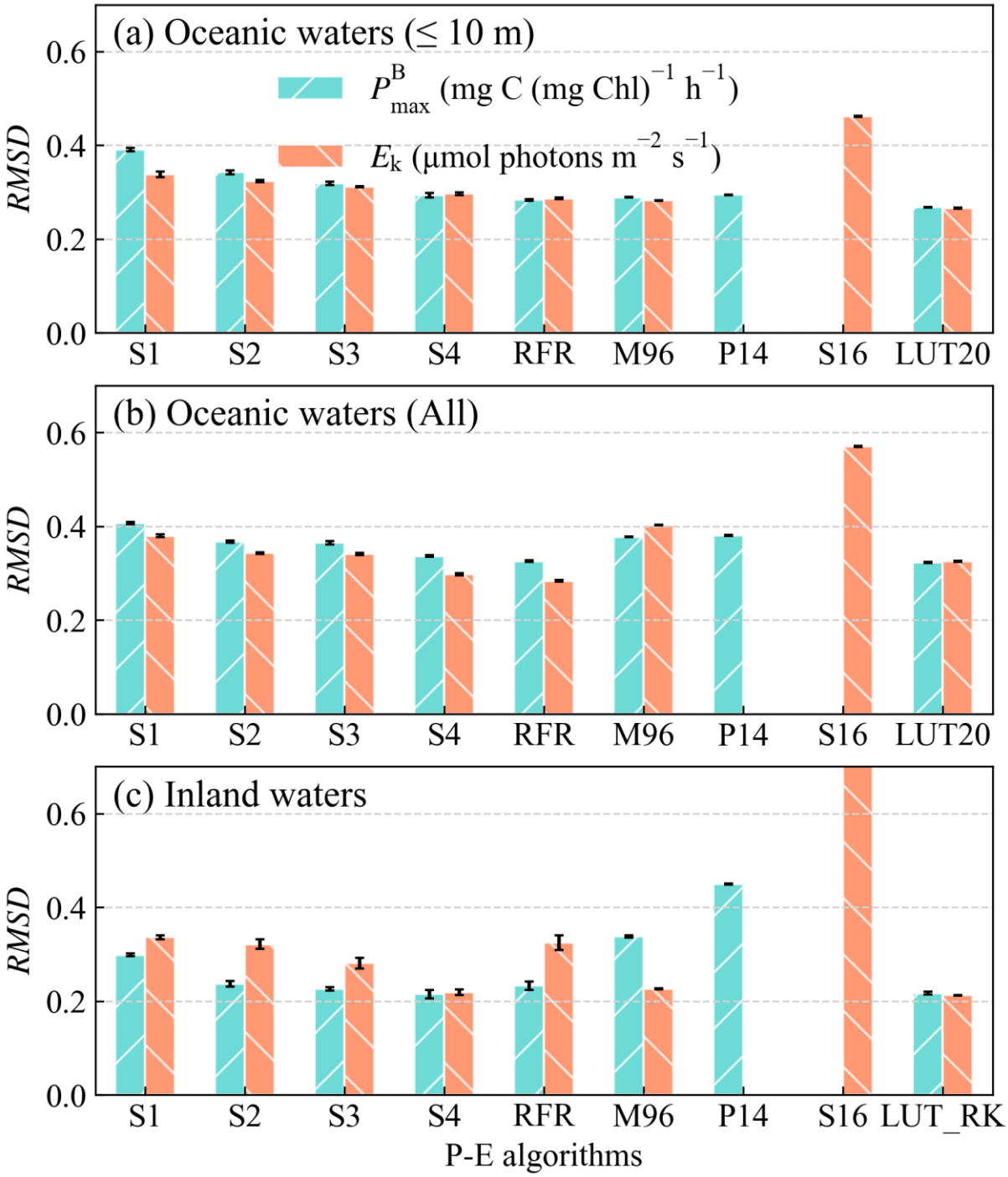

Figure 4.1. Performance comparison of different P-E algorithms based on the testing data in each round of cross validations together with independent data from the AVDs. Algorithm participants included four ERFRs (denoted by the corresponding training scenarios S1–S4), a traditional RFR (trained by S4), M96, S16, LUT20 (for oceanic waters), and LUT_RK (for inland waters). The bars represent the mean *RMSD* values, and the black error bars stand for standard deviations.

In addition, the M96, P14, and LUT20 generally worked well for surface waters (Figure 4.1a), where their performance was comparable to that of the ERFR, but it was evident that the M96 and P14 performed worse than the ERFR and LUT20 when all clear water data were used (Figure 4.1b). Similar results were also found for inland waters, where the M96 and P14 could not accurately estimate P^{B}_{max}, and the overall accuracy of the LUT_RK was very close to that of the ERFR. The E_k estimates of the M96 were surprisingly almost as accurate as those of the ERFR. In contrast, the S16 performed worst for both oceanic and inland waters. These results corroborate that the ERFR has a broader applicability compared with other P-E algorithms.

4.1.2. Satellite-based validation of photosynthetic parameters

A total of 649 valid matchups of P^{B}_{max} and E_k were obtained at different depths from oceanic waters based on most of the P-E algorithms. The exceptions were the P14 and S16, which yielded only 607 valid estimates because some Z_m and *d*PAR data were unavailable. As shown in Figure 4.2, it was evident that the ERFR performed best and provided a greater fidelity of P^{B}_{max} and E_k retrievals (*RMSD* $\leq$ 0.28, *UAPD* $\leq$ 48.3%, and *Bias* $\leq$ 7.4%). The estimated values of existing algorithms exhibited obvious discrepencies with measured values, especially for those at water depths > 60 m. The resultant *RMSD*s fell in the range 0.29–0.56 and *UAPD*s exceeded 48.3%. Note that several constant E_k estimates of 10 μmol photons m^{-2} s^{-1} were yielded by the S16 because it assumed that E_k would reach a lower limit under low-light conditions (Silsbe et al., 2016).

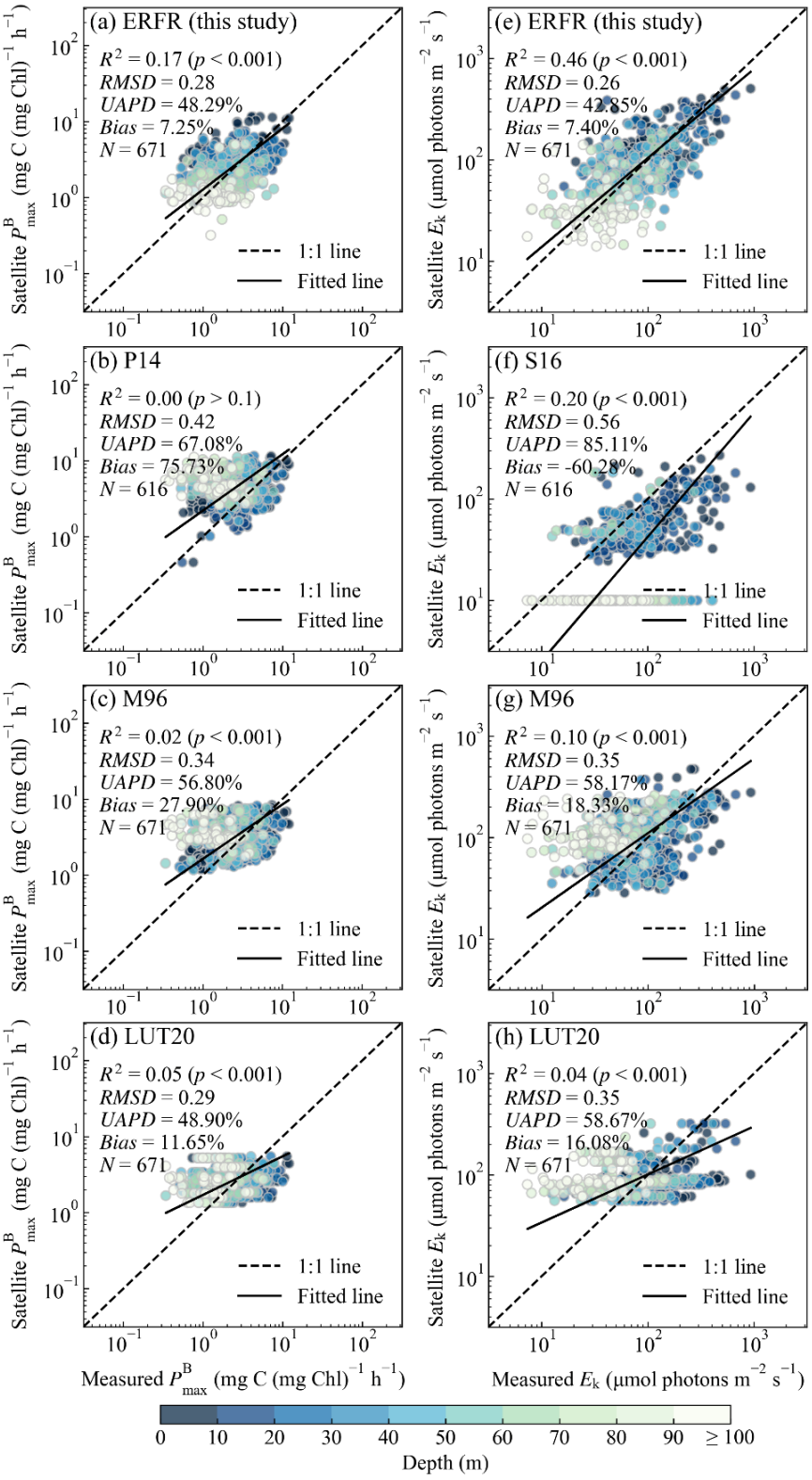

Figure 4.2. Comparison of satellite-derived P^B_{max} (top row) and E_k (bottom row) at different depths for oceanic waters using different P-E algorithms: ERFR (a and e), P14 (b), M96 (c and g), LUT20 (d and h), and S16 (f). Black dashed and solid lines represent the 1:1 line and fitted line, respectively.

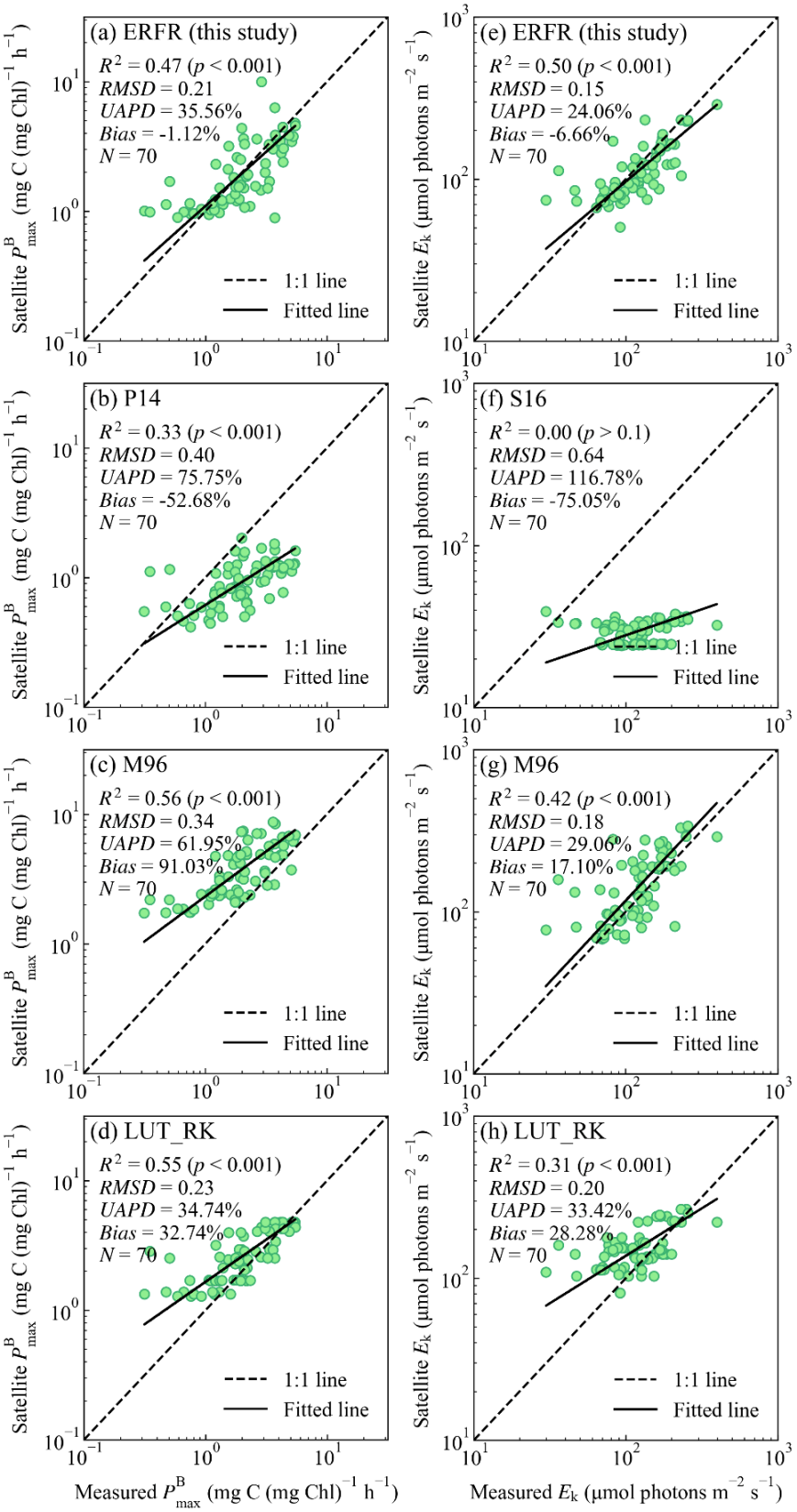

Figure 4.3. Comparison of satellite-derived P^B_{max} (top row) and E_k (bottom row) for inland waters using different P-E algorithms: ERFR (a and e), P14 (b), M96 (c and g), LUT20 (d and h), and S16 (f). Black dashed and solid lines represent the 1:1 line and fitted line, respectively.

Figure 4.3 shows the comparison of these P-E algorithms based on 70 valid matchups from inland waters. The ERFR exhibited the best performance for both P^{B}_{max} and E_k estimations as indicated by the lowest *RMSD* (0.21 and 0.15, respectively) and *Bias* (–1.1% and –6.7%, respectively). In contrast, the P14 and S16 were prone to underestimate P^{B}_{max} (*Bias* = –52.7%) and E_k (*Bias* = –75.1%), respectively. The M96 systematically overestimated P^{B}_{max} with the largest *Bias* of 91.0% among all P-E algorithms, but its performance in estimating E_k was clearly better than that of the S16 and LUT_RK as indicated by a higher R^2 and lower errors. Though the M96 performed slightly worse than the ERFR, it could be an alternative in addition to the ERFR for the estimation of E_k in inland waters because of its convenience. There was also an apparent overestimation observed for the LUT_RK as shown in Figure 4.3d and Figure 4.3h, especially when P^{B}_{max} was lower than 1 mg C (mg Chl)$^{-1}$ h^{-1} and E_k was lower than 100 μmol photons m^{-2} s^{-1}.

It is noteworthy that the LUT_RK (or LUT20, see Figure 4.2d and Figure 4.2h) performed worse on matchup data than on *in situ* data (Figure 4.1), which suggests that LUT methods are not suited for applications at smaller spatial and temporal scales. Because the numbers of matchups extracted from oceanic and inland waters are less than the data from the AVDs. It implies that LUT methods can capture the general variability of P^{B}_{max} and E_k in time and space, but they have limited performance in a certain region or in a short time period. Whereas the ERFR was proved to be more robust in spite of the change of validation data.

4.2. Performance evaluation of the TPM_{ERFR}

4.2.1. Independent validation of IPP

A total of five existing IPP models (Table 3.1) were used for the comparison against the TPM_{ERFR} to assess respective effect on the modeling of IPP in different water types. Table 4.1 provides a tabulation of the individual accuracy statistics based on the AVDs for each IPP model in oceanic and inland waters. For oceanic waters, a total of 500 valid IPP data were produced by all IPP models, apart from the $VGPM_{KI05}$ and $VGPM_{H02}$, which only yielded 499 and 481 valid results (IPP estimates less than 0 mg C m^{-2} d^{-1} are treated as invalid data), respectively. The accuracy of the TPM_{ERFR} was comparable to that of the TPM_{RFR}, as indicated by an identical *RMSD* of 0.26, because they performed similarly in retrieving P^{B}_{max} and E_k in oceanic waters (Figure 4.1). The TPM_{LUT20} and $AbPM_{AM96}$ also performed similarly since they had very close values of *RMSD*s and *UAPD*s that were around 0.30 and 54.0%, respectively, and they were prone to underestimate IPP with their values of *Bias* around –23.0%. The $VGPM_{BF97}$ and $VGPM_{KI05}$ were unable to produce IPP estimates as accurate as the TPM_{ERFR} and TPM_{RFR}, as indicated by the larger *RMSD*s of 0.33 and 0.29, respectively. Only the $VGPM_{BF97}$ showed a trend of overestimation (*Bias* = 8.3%) while other models showed a trend of underestimation. The $VGPM_{H02}$ was clearly not suited for application to oceanic waters, resulting in the largest *RMSD* (*UAPD*) of 0.74 (88%) among all IPP models, because it was originally designed for the turbid waters of the Chesapeake Bay (Harding et al., 2002).

Table 4.1. Accuracy statistics of the different IPP models based on independent validation using the AVDs. The results of the TPM_{ERFR} are highlighted in bold. N is the number of valid IPP data.

Water type	Method	IPP (mg C m^{-2} d^{-1})				
		R^2	*RMSD*	*UAPD* (%)	*Bias* (%)	N
Oceanic waters	**TPM_{ERFR}**	**0.59**	**0.26**	**47.34**	**−10.91**	**500**
	TPM_{RFR}	0.60	0.26	46.42	−13.17	500
	TPM_{LUT20}	0.54	0.31	53.99	−23.94	500
	$VGPM_{BF97}$	0.45	0.33	53.49	8.29	500
	$VGPM_{KI05}$	0.41	0.29	45.25	−13.82	499
	$VGPM_{H02}$	0.11	0.74	88.00	−59.13	481
	$AbPM_{AM96}$	0.52	0.30	54.89	−21.92	500
Inland waters	**TPM_{ERFR}**	**0.72**	**0.19**	**32.14**	**5.09**	**1463**
	TPM_{RFR}	0.69	0.22	36.43	20.21	1463
	TPM_{LUT_RK}	0.73	0.19	31.73	9.34	1463
	$VGPM_{BF97}$	0.62	0.62	111.22	273.91	1463
	$VGPM_{KI05}$	0.60	0.35	64.62	88.48	1463
	$VGPM_{H02}$	0.56	0.33	57.50	44.92	1463
	$AbPM_{AM96}$	0.74	0.36	66.80	101.63	1463

For the independent validation based on the turbid waters of Lake Kasumigaura, the TPM_{ERFR} yielded the lowest *RMSD* of 0.19 and *Bias* of 5.1%, while the TPM_{RFR} was worse than the TPM_{ERFR} and tended to overestimate IPP with a *Bias* of 20.2% (Table 4.1) because of the poor retrievals of P^{B}_{max} and E_k from the RFR (Figure 4.1c). The statistics for the TPM_{ERFR} and TPM_{LUT_RK} were, as expected, very close because the performances of the ERFR and LUT_RK were similar (Figure 4.1c). All three VGPM models and the $AbPM_{AM96}$ systematically overestimated IPP data, especially for the $VGPM_{BF97}$, which had the largest *Bias* of 273.9%. Although the $VGPM_{KI05}$ worked better than the $VGPM_{BF97}$ because of its improved performance in relatively productive coastal waters (Kameda and Ishizaka, 2005), it was not very effective in the case of Lake Kasumigaura, where the concentrations of both phytoplankton and suspended sediments are always high. The lower accuracy of the $VGPM_{H02}$ suggested that a localized empirical model is typically difficult to be used directly for other regions and thus has limited applicability.

4.2.2. Satellite-based validation of IPP

Figure 4.4 shows the comparison between satellite-derived and measured IPP based on 56 valid matchups extracted from oceanic waters for different IPP models. The agreement between estimated and measured IPP was, not surprisingly, degraded to a certain degree for all models, compared with their performances on *in situ* data, especially for DR models (i.e., the TPM_{ERFR}, TPM_{LUT20}, and $AbPM_{AM96}$). This degradation of performance might be partially attributed to the

uncertainty in the Chl*a* profiles estimated from satellite observations (see Section 4.5.2). The statistics in Figure 4.4 suggest that the TPM_{ERFR} outperformed other IPP models, yielding the highest R^2 of 0.52 and the lowest *RMSD* (*UAPD*) of 0.25 (41.5%). The TPM_{LUT20} performed slightly worse than the TPM_{ERFR} but better than other exisitng models, exhibiting the second highest accuracy. Whereas the $AbPM_{AM96}$ tended to overestimate IPP at low values (< ~800 mg C m^{-2} d^{-1}) and underestimate IPP at high values (> ~800 mg C m^{-2} d^{-1}). One of the reasons for this pattern could be the poor parameterization of phytoplankton absorption and energy transfer efficiency in the vertical direction. As expected, the IPP estimates yielded by the $VGPM_{H02}$ were highly disordered, showing some unrealistic values and the largest *RMSD* and *UAPD* of 0.51 and 75.6% among all models, respectively. Although part of the $VGPM_{H02}$-based IPP estimates distributed along the 1:1 line, the overall correlation with the measured values was insignificant (R^2 = 0.01). It implies that the utility of regression-based models can be efficient in the areas of interest with the support of local *in situ* datasets, however, the applicability of this kind of IPP models is possibly limited for other areas with different water properties. Moreover, the $VGPM_{BF97}$ significantly overestimated IPP, indicative of the largest *Bias* of 37.2%. With the inclusion of Chl*a* in the parameterization of P^{B}_{opt} in addition to T (Eq. (3.9)), the $VGPM_{KI05}$ worked better than the $VGPM_{BF97}$. But the performance of $VGPM_{KI05}$ was still not as good as the proposed TPM_{ERFR}, implying that both the estimation of photosynthetic parameters and model structure (DI or DR) are important for IPP modeling.

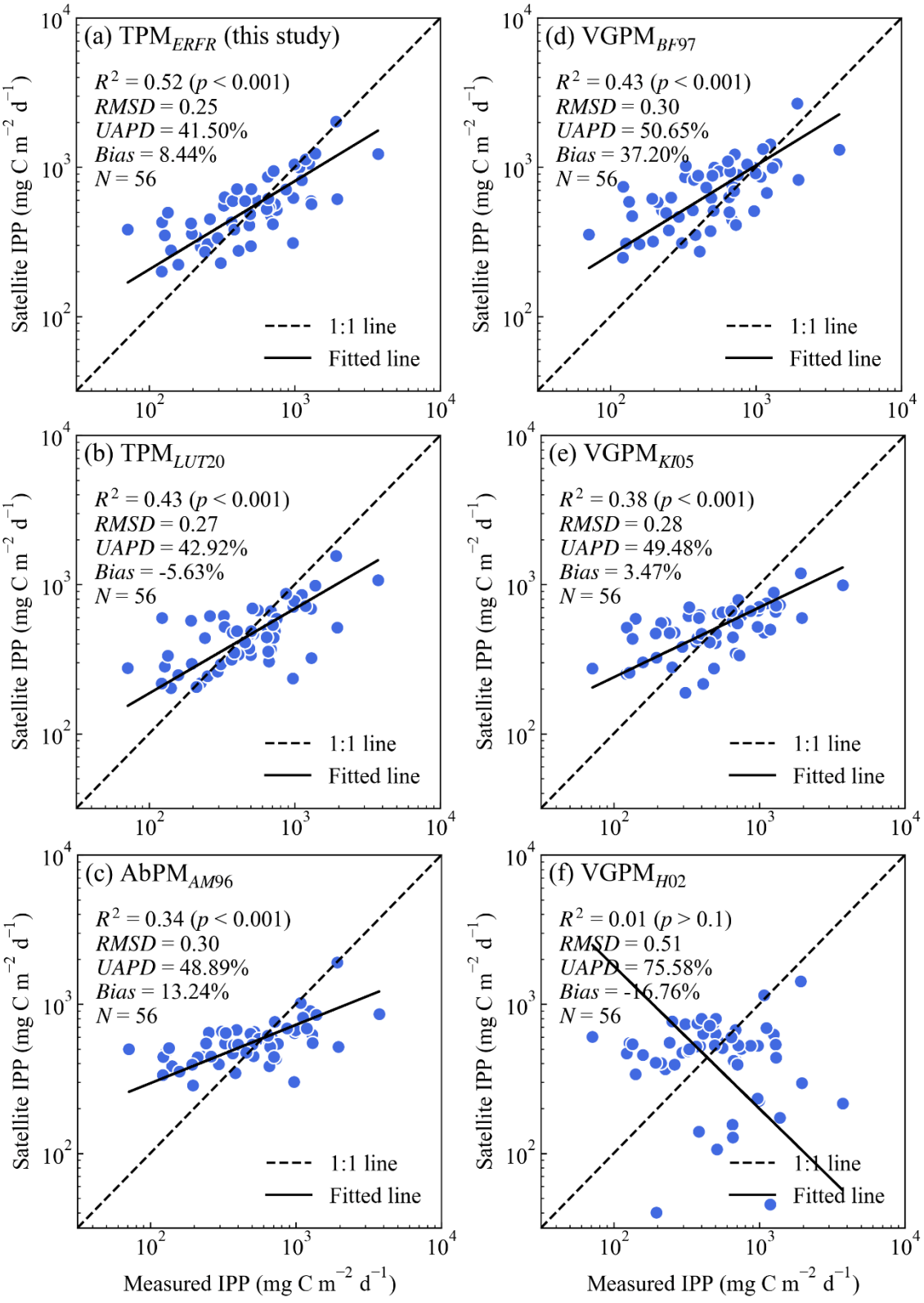

Figure 4.4. Comparison of satellite-derived IPP for oceanic waters based on different IPP models: TPM$_{ERFR}$ (a), TPM$_{LUT20}$ (b), AbPM$_{AM96}$ (c), VGPM$_{BF97}$ (d), VGPM$_{KI05}$ (e), and VGPM$_{H02}$ (f). The 1:1 line and fitted line are shown as dashed and solid lines, respectively.

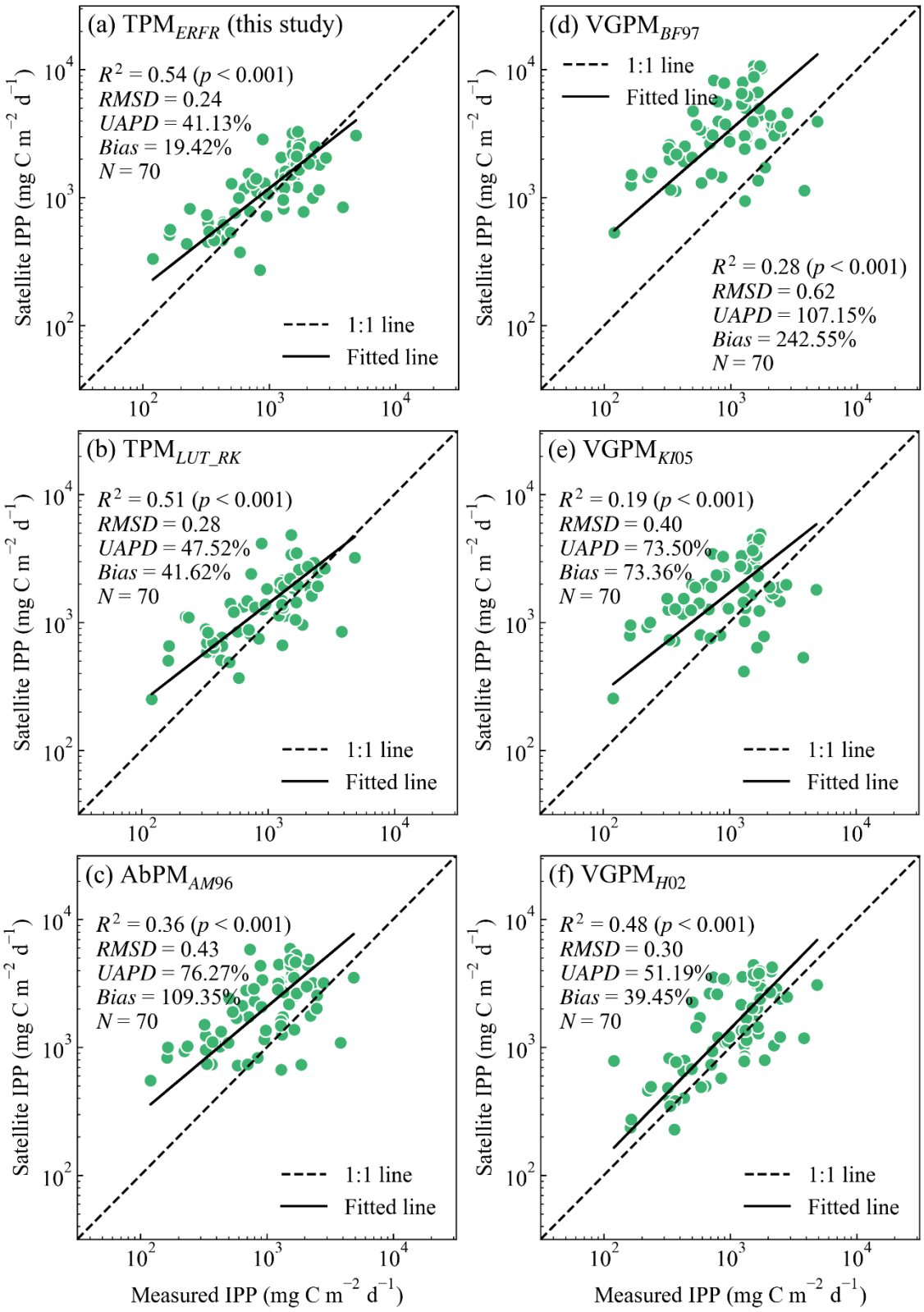

Figure 4.5. Comparison of satellite-derived IPP for inland waters based on different IPP models: TPM$_{ERFR}$ (a), TPM$_{LUT_RK}$ (b), AbPM$_{AM96}$ (c), VGPM$_{BF97}$ (d), VGPM$_{KI05}$ (e), and VGPM$_{H02}$ (f). The 1:1 line and fitted line are shown as dashed and solid lines, respectively.

Figure 4.5 compares the performances of different IPP models based on 70 MERIS/MSI matchups extracted from the Rhode River and Lake Kasumigaura. The TPM_{ERFR} provided the most accurate IPP estimates with the highest R^2 of 0.54 and the lowest error statistics. The accuracy of TPM_{LUT_RK} was clearly reduced by the fact that the LUT_RK generally tends to represent the climatological information on P^{B}_{max} and E_k. Both the $VGPM_{BF97}$ and $VGPM_{KI05}$ significantly overestimated IPP, as indicated by the large values of *Bias* (242.6% and 73.4%, respectively), even though the $VGPM_{KI05}$ performed relatively better than the $VGPM_{BF97}$. These results suggest that the overestimation was subject to both the parameterization of P^{B}_{opt} and the empirial volume function of the VGPM (Behrenfeld and Falkowski, 1997b), which may not be applicable for turbid and shallow waters. Because the $VGPM_{H02}$ retuned the coefficients of the original VGPM, it performed better than the $VGPM_{BF97}$ and $VGPM_{KI05}$ and returned a lower *RMSD* of 0.3. Nevertheless, the $VGPM_{H02}$ was not able to perform as well as the TPM_{ERFR} in these two inland waters. The poor performance of the $AbPM_{AM96}$ was probably subject to the configuration of the phytoplankton light utilization efficiency, which was originally parameterized for global oceans (Antoine and Morel, 1996).

4.2.3. Applications on the upper Chesapeake Bay and Lake Kasumigaura

The MSI-based IPP distributions were generated for the upper Chesapeake Bay on four selected dates (Jan. 21, Apr. 15, Jul. 29, and Oct. 7) in 2020. The IPP distributions on these dates could be used to demonstrate typical spatiotemporal

characteristics of IPP in winter, spring, summer, and autumn, respectively in the upper Chesapeake Bay (Figure 4.6a–d). In general, IPP values gradually increased from winter to summer, especially at inshore sites near river mouths, and reached a maximum in summer. Notably, a zone of low IPP was apparent in the northern watershed (39.3°N to 39.6°N) due to the local freshwater discharge, carrying abundant suspended sediments from the Susquehanna River (Balasubramanian et al., 2020; Prasad et al., 2010) and thus hindering phytoplankton from absorbing sunlight. High inputs of nutrients from northwestern tributaries would induce algae blooms and concomitant high IPP values in small estuaries (Wolny et al., 2020).

Figure 4.6e–h show the MSI-based IPP distributions generated on four dates (Mar. 29, Jun. 29, and Nov. 14, 2018 and Jan. 18, 2019) for the demonstation of local spatiotemporal change of IPP in Lake Kasumigaura. The spatial patterns of IPP in Lake Kasumigaura were not as complex as those in the upper Chesapeake Bay, showing a relatively more homogeneous distribution on each date. It has been reported that the northern coast of Lake Kasumigaura is usually dominated by high IPP values throughout the year (Takamura et al., 1987), and overall, the average IPP is generally high in summer (with some exceptions due to changes of phytoplankton populations) and low in winter (Takamura and Aizaki, 1991). In conclusion, the TPM_{ERFR} successfully captured the seasonal characteristics and subtle spatial patterns of IPP in these two turbid inland waters, such as those associated with freshwater streams (Figure 4.6f) and algae slicks (Figure 4.6g).

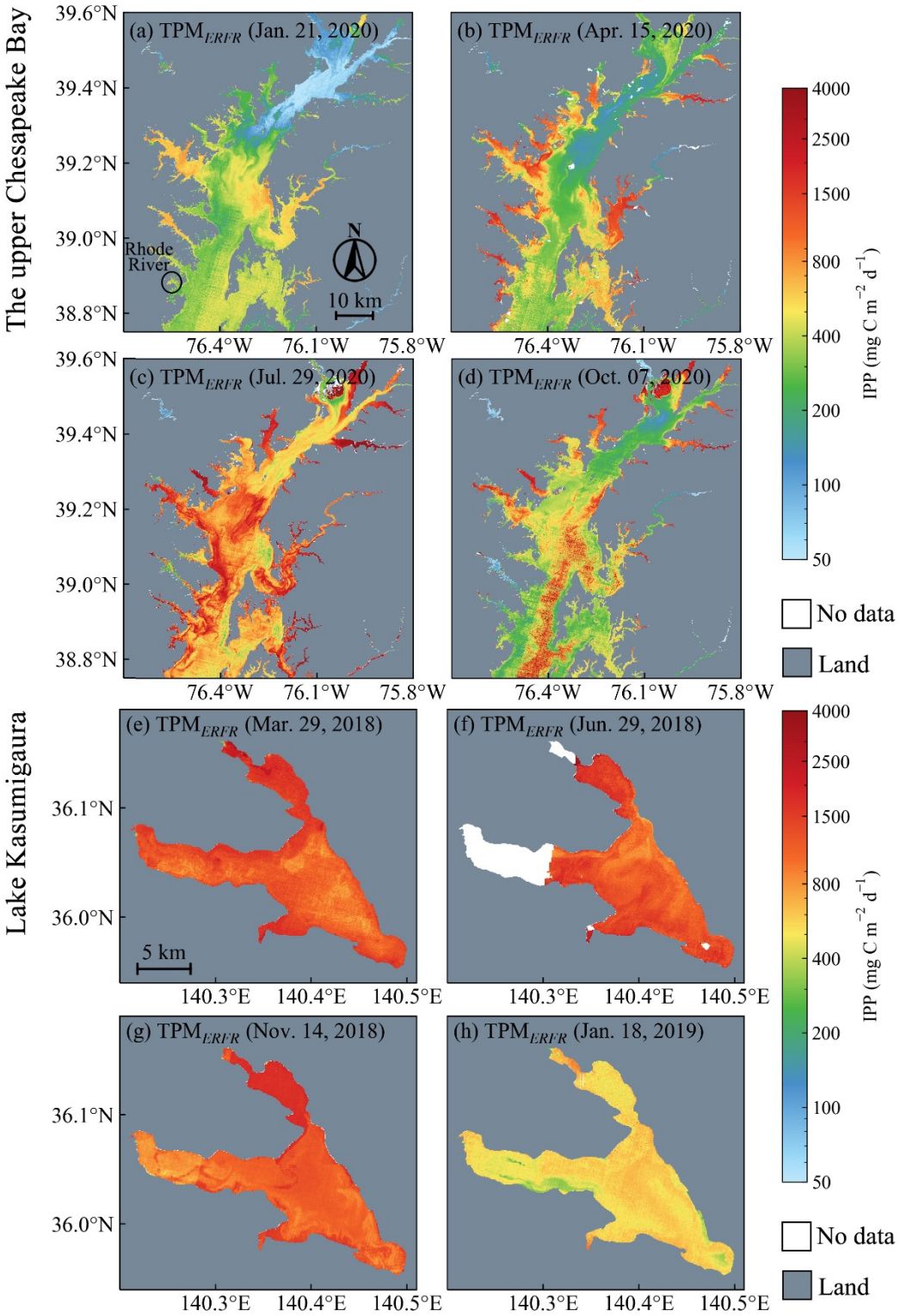

Figure 4.6. Spatiotemporal distributions of IPP generated by the TPM$_{ERFR}$ over the upper Chesapeake Bay on Jan. 21 (a), Apr. 15 (b), Jul. 29 (c), and Oct. 7 (d), 2020, and in Lake Kasumigaura on Mar. 29 (e), Jun. 29 (f), and Nov. 14 (g), 2018 and Jan. 18, 2019 (h). The location of the Rhode River is indicated by a black circle in (a).

4.3. Sensitivity of the TPM$_{ERFR}$

Figure 4.7 shows the responses of the P^{B}_{max}, E_k, and IPP estimates to the uncertainties in the three input parameters (T, Chl$a$, and K_d(PAR)) propagated through the ERFR and the subsequent TPM, where the corresponding signs of *Bias* are also displayed to facilitate interpretation. Recall that the uncertainty in the estimated IPP based on the TPM$_{ERFR}$ is subject to not only the uncertainties of the ERFR-derived P^{B}_{max} and E_k but also the uncertainties of Chl$a$ and K_d(PAR) introduced in the TPM. That is, both the uncertainties of the ERFR and the TPM will be eventually propagated into the TPM$_{ERFR}$-based IPP estimates. In particular, to quantify the uncertainty of the ERFR on IPP estimates only, another experiment scenario was designed, which was that the estimated P^{B}_{max} and E_k were substituted into the TPM while the inputs of Chl$a$ and K_d(PAR) in the TPM were kept as the measured ones for IPP calculation. The resultant IPP from this scenario was denoted as IPP*, which, in this case, would not be directly affected by the perturbed inputs of Chl$a$ and K_d(PAR), and thus IPP* was equivalent to IPP when only the input of T was perturbed (Figure 4.7a and Figure 4.7d)

As indicated by the minimum ε corresponded to error-free (δ = 0%) input parameters, the inherent uncertainties in P^{B}_{max} and E_k derived from the ERFR were very similar for all water bodies, varying between ~15–18% and ~10–12%, respectively. The inherent uncertainty in the TPM$_{ERFR}$-based IPP (same as IPP* since the inputs of Chl$a$ and K_d(PAR) were error-free) was consequently ~10% (*Bias* ≥ 0, overall overestimation) and ~18% (*Bias* < 0, overall underestimation)

for oceanic and inland waters, respectively. The larger changes of ε for $P_{\max}^{B}$ and E_k induced by error-introduced T and K_d(PAR) than of Chl*a* indicated that the ERFR was generally more sensitive to the uncertainties of T and K_d(PAR) than of Chl*a*. For example, in oceanic waters, the maximum ε of $P_{\max}^{B}$ was about 40% caused by the perturbed T and K_d(PAR) (Figure 4.7a and Figure 4.7c), but was only about 25% caused by the perturbed Chl*a* (Figure 4.7b).

More importantly, the analysis showed that the uncertainty was usually smaller in the resultant IPP* than in $P_{\max}^{B}$ (and even smaller than the uncertainty in E_k for oceanic waters) when the uncertainties of T, Chl*a*, and K_d(PAR) were propagated only into the ERFR but not into the TPM. It implies that there could be a certain degree of compensation between the uncertainties in $P_{\max}^{B}$ and E_k, which might suppress the uncertainty of IPP. For instance, in Figure 4.7a, when the ERFR-based retrievals of $P_{\max}^{B}$ and E_k were all biased high when T was overestimated, it is harder for phytoplankton to attain the enhanced $P_{\max}^{B}$ because of the accompanying increase of E_k, and vice versa. It was likely that the bias of IPP was mainly controlled by the bias of $P_{\max}^{B}$. For instance, in Figure 4.7c, when $P_{\max}^{B}$ was underestimated (or overestimated), so was IPP*, whereas E_k was overestimated. In contrast, the IPP estimates was clearly subject to the uncertainties in Chl*a* and K_d(PAR) when they were propagated into the TPM. As expected, the IPP estimates of all water bodies were generally overestimated (*Bias* ≥ 0) or underestimated (*Bias* < 0) because of overestimates of Chl*a* or K_d(PAR), respectively.

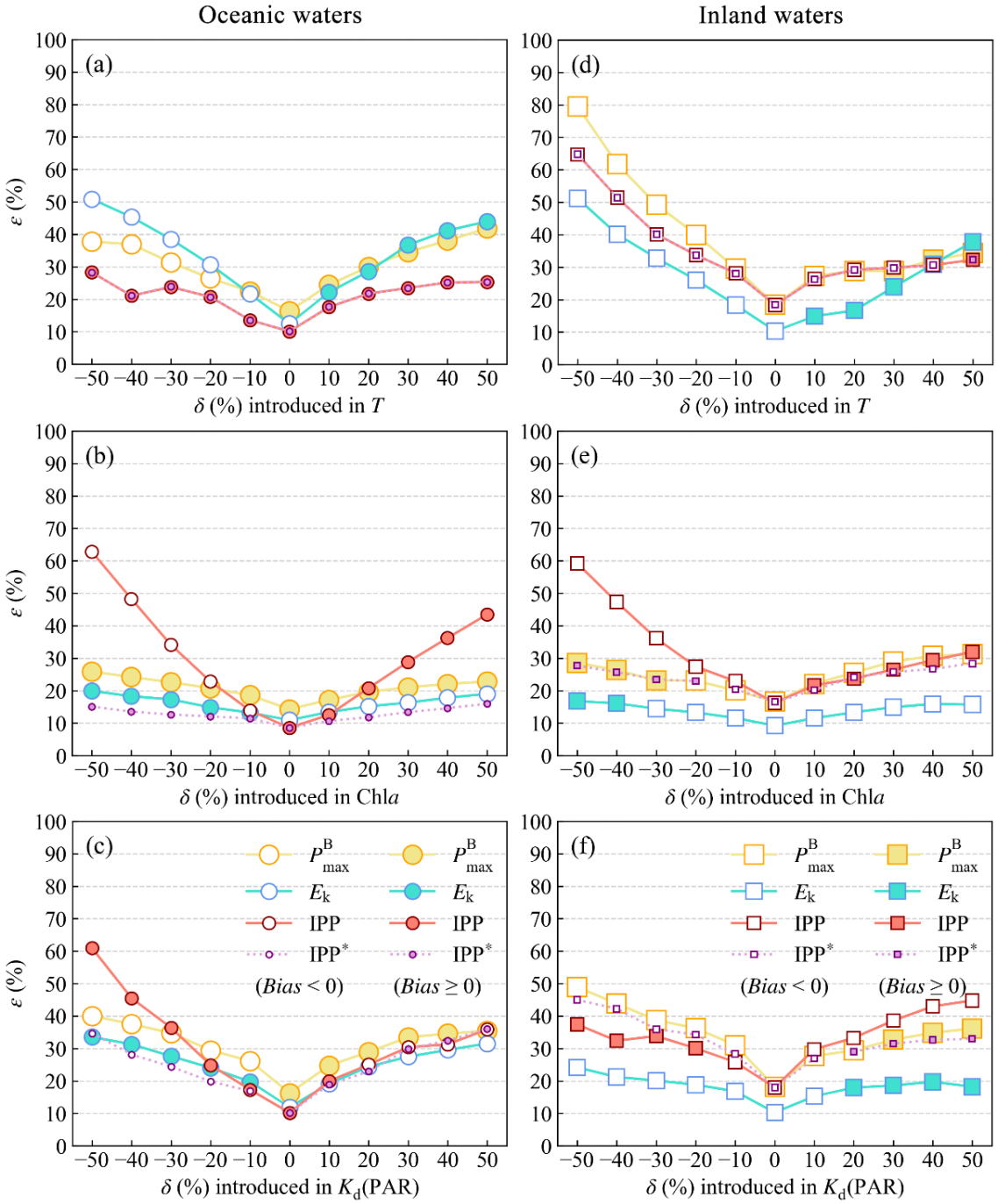

Figure 4.7. Uncertainties (ε) of different investigated estimates induced by introducing relative errors (δ) to T (a and d), Chl*a* (b and e), and K_d(PAR) (c and f) for oceanic (right panel) and inland waters (left panel). The investigated estimates were P^B_{max} and E_k from the ERFR, IPP from the TPM$_{ERFR}$. In addition, IPP calculated by the ERFR-estimated P^B_{max} and E_k with other parameters measured *in situ* was also included, denoted as IPP*, to investigate the uncertainty propagated from the ERFR to the TPM.

4.4. Rationality and applicability of the ERFR

The applicability of the ERFR, as a data driven algorithm, is heavily dependent on the selection of input parameters. In this study, it was confirmed that the combination of T, Chl$a$, K_d(PAR), and r_{PAR} was the optimal selection (Figure 4.1). This conclusion is consistent with the knowledge of potential drivers that influence phytoplankton photosynthesis (Behrenfeld and Falkowski, 1997b; Bouman et al., 2005; Eppley, 1972; Kameda and Ishizaka, 2005). Since K_d(PAR) reflected the overall optical property of water columns, which can be distinct in different water bodies, the inclusion of K_d(PAR) effectively improved the estimation of P^B_{max} and E_k in the upper layers of oceanic waters (Figure 4.1a) and inland waters (Figure 4.1c), as indicated by the comparison between S2 and S3. It implies that the bio-optical charateristics of water bodies should be taken into account for the development of a better generalized algorithm. Meanwhile, since the depth-dependence of photosynthetic parameters has already been revealed (Brewin et al., 2017; Uitz et al., 2008), including K_d(PAR) may still not be enough to interpret the variability of P^B_{max} and E_k along vertical direction. For this reason, the performance of S2 and S3 was nearly identical as shown in Figure 4.1b. The inclusion of r_{PAR} could largely address this issue and thus further increased the accuracy of estimation for all water bodies by introducing depth information into the ERFR, as indicated by the comparison between S3 and S4 (Figure 4.1). However, the parameterization of the vertical change of P^B_{max} and E_k has rarely been considered in previous studies (e.g., Kulk et al., 2020; Morel, 1991; Picart

et al., 2014), which can explain the overall superiority of the ERFR.

Development dataset is also essential to the success of the ERFR. Though we had only one dataset from turbid waters used for the development of the ERFR (i.e., RRCB), and this dataset is not fully representative of other turbid and shallow waters worldwide, its inclusion with the datasets from clear waters effectively expanded the dynamic range of measurements, especially for Chl*a* (0.017−214.80 mg m^{-3} with a CV of 131.0%) and K_d(PAR) (0.029−7.66 m^{-1} with a CV of 92.3%). As a result, the diversity of photosynthetic and environmental parameters in the ADDs contributed to the broad applicability of the ERFR in clear to turbid waters. The comparisons of the ERFR with existing P-E algorithms (Figures 4.1−4.3), which have been calibrated using the data with limited representative and quantity (Eppley, 1972; Morel, 1991; Picart et al., 2014; Sathyendranath et al., 2009), corroborate the important role of our sizable datasets in the development of ERFR. In addition, the enhanced extrapolation capability of the ERFR that has been preliminarily demonstrated by the comparison in Lake Kasumigaura against a traditional RFR (Figure 4.1 and Figure 4.3) is another reason for the promising performance. It is noteworthy that the M96 performed unexpectedly well for retrieving E_k in the Rhode River and Lake Kasumigaura (Figure 4.1c and Figure 4.3g). However, in order to convert the original output of the M96 to E_k, reliable retrieval of $a_{ph}(\lambda)$ is required (Morel et al., 1996), which could be difficult to achieve in global turbid waters. Besides, the fact that the S16 was originally calibrated using satellite data (Behrenfeld et al., 2016) and its

configuration for modeling the change of E_k with depth might not be directly adopted for other models (Silsbe et al., 2016) can explain the poor performance of the S16. Thus, we can conclude that the ERFR can be an effective alternative for the remote assignment of P^B_{max} and E_k because of its proven rationality and applicability.

4.5. Sources of uncertainty in the TPM$_{ERFR}$

4.5.1. Algorithm development dataset

Although the diversity of the ADDs contributed to the potentially wide applicability of the TPM$_{ERFR}$, the inconsistencies in Chl*a* data measured by different techniques and photosynthetic parameters obtained by P-E experiments with different incubation times and artificial light sources raise concerns about how these differences affect the development and future application of our model. On the one hand, it has been reported that fluorometric Chl*a* may underestimate HPLC Chl*a* by 2−19% (Bouman et al., 2018), and spectrophotometric Chl*a* may overestimate HPLC Chl*a* by ~16% for estuarine waters (Pinckney et al., 1994). Our sensitivity analysis revealed that P^B_{max} and E_k retrievals were relatively insensitive to the uncertainty of Chl*a*, mainly because Chl*a* was log-transformed before being used in the ERFR. For oceanic (inland) waters, as shown in Figure 4.7b (Figure 4.7e), a typical δ of ≤ 20% in Chl*a* data generally induces the uncertainties of ≤ 20% (25%) for P^B_{max} and ≤ 15% (12%) for E_k, and eventually the overall uncertainty propagated to IPP retrievals is ≤ 21% (28%). It suggests

that the potential bias in Chl*a* measurements will not have a significant impact on the ERFR-derived P^{B}_{max} and E_k, and the resultant uncertainty propagated to the TPM$_{ERFR}$-derived IPP over depth and time will not rise dramatically. Further correction and unification of Chl*a* measurements are still desired to reduce the bias, which may affect the temporal trend of IPP estimated by IPP models (Saba et al., 2010).

On the other hand, it has been known that the nature of PP measurements varies with incubation time (Marra, 2009; Milligan et al., 2015; Pei and Laws, 2013), and the effect of photoacclimation will become dominant if incubation time is relatively long (e.g., over 4 hours) (IOCCG, 2021). In that case, the obtained photosynthetic parameters may not be able to fully represent the photosynthetic response of phytoplankton in natural water condition. Since ~83% of the photosynthetic parameters used to develop the ERFR were obtained from incubations ≤ 2 h, the inconsistency of incubation time was thus thought to be acceptable. Note that the influence of incubation time on remote sensing application of IPP models is not evident (Campbell et al., 2002). The uncertainties in IPP models and satellite input are usually larger than the systematic errors caused by variable incubation times. Besides, the influence of different light sources on α^B and E_k has already been investigated in earlier studies, and the resultant errors can be minor to IPP retrievals because the response of α^B and E_k to light sources is usually more sensitive in dark layers rather than in well-illuminated upper layers (Bouman et al., 2018).

4.5.2. Satellite input and computation scheme

As shown in Figure 4.4a, the IPP retrievals from the TPM$_{ERFR}$ were generally overestimated at low values and underestimated at high values. Similar patterns were observed for other IPP models (except for the VGPM$_{H02}$) as shown in Figures 4.4b–4.4e but were not evident for the P^{B}_{max} and E_k retrievals (Figure 4.2a and Figure 4.2e). It was inferred that these trends of IPP were very likely caused by the uncertainties in satellite-retrieved Chl*a* and K_d(PAR) (especially the former) rather than in the ERFR-retrieved P^{B}_{max} and E_k. Because the uncertainties of Chl*a* and K_d(PAR) have a direct impact on IPP via the TPM (Eq. (1.1)), and a similar deviation trend was found in the assessment of extracted Chl*a* matchups (Figure 2.7).

It has been known that satellite-based SST and LSWT observations typically have a decent accuracy (< 1 K) (MacCallum and Merchant, 2012; Merchant et al., 2019), the major sources of error for retrieving IPP based on the TPM$_{ERFR}$ are thus the satellite observations of Chl*a* and K_d(PAR), especially for Case-II waters originated from coastal and inland areas. It is difficult to accurately estimate IPP in these areas because of the current challenges associated with atmospheric correction and inversion of ocean color parameters. Nevertheless, the encouraging results of this study (Figure 4.5) suggest that the improvement of IPP estimation in Case-II waters can be achieved by incorporation of preferred atmospheric correction processors and inversion algorithms of Chl*a* and K_d(PAR).

Despite that ocean color sensors only capture the characteristics of surface

waters from space, approaches to retrieve vertical distributions of required water properties for modeling IPP, such as Chl*a* and *T* in this study, are desired for a better delineation of three-dimensional distribution of IPP, especially for open oceans, where the stratification of water column is very common. In our remote sensing computation scheme, the vertical distribution of *T* was reconstructed by using satellite-measured SST and the normalized *T* profile from GODAS. The fact that the change of *T* is usually small within well-illuminated layers (for instance, in this study, the average difference between *T* at the surface and the depth where r_{PAR} is equal to 20% is only about 3%) suggests that the uncertainty in *T* indeed have a little effect on IPP computation. In contrast, the necessity of considering non-uniform Chl*a* profile for remotely estimating IPP based on DR models has already been emphasized in earlier studies (Bouman et al., 2020; Hill et al., 2013). Because if the Chl*a* profile is assumed to be uniform, there is very likely to be an underestimation of phytoplankton biomass for a water column, where water is stratified enough to form a DCM (note that satellite-measured Chl*a* is generally smalled than DCM). Nevertheless, the uncertainty in remotely retrieved Chl*a* profile could also be a source of error, and future efforts including accuracy evaluation and algorithm refinement are desired to provide satisfactory estimation of Chl*a* profile from satellites, which will greatly support the modeling of IPP. In conclusion, the approach to retrieve *T* profile proposed in this study is acceptable, and the accuracy of IPP modeling based on the TPM_{ERFR} is greatly dependent upon whether Chl*a* profile can be accurately estimated by satellite data.

4.6. Future perspectives of modeling IPP from space

Future efforts are necessary for the development of *in situ* datasets based on consistent measurement protocols in order to improve the quality and quantity of PP measurements and ensure that these measurements are inter-comparable (IOCCG, 2021). An incubation-free method, fast repetition rate fluorometry (FRRF), could be an efficient approach to extend the spatiotemporal coverage of photosynthetic parameters based on chlorophyll fluorescence (Schuback et al., 2021; Smyth et al., 2004), and the increasing data collections will ultimately provide new insights for algorithm development. Parallel IPP measurements obtained by long-term and short-term (calculated by photosynthetic parameters) incubations are required to investigate their consistency and understand the variable uncertainty of *in situ* IPP data measured by different approaches. Moreover, additional environmental measurements that are either directly or indirectly related to phytoplankton photosynthesis (especially those that can be estimated from space) should also be considered to increase data diversity, such as the particulate backscattering coefficient, phytoplankton carbon, and phytoplankton size classes (Roy et al., 2017; Thomalla et al., 2017; Xi et al., 2020). Such information may facilitate the interpretation of the generated IPP distributions.

Future work should consider inclusion of a component that accounts for the spectral and vertical distributions of incident light into the TPM_{ERFR} to delineate underwater light field more accurately for the modeling of IPP as emphasized in

previous studies (Lee et al., 2015; Sathyendranath et al., 2020). Forthcoming hyperspectral ocean-color sensors, such as the Plankton, Aerosol, Cloud, ocean Ecosystem (PACE) (Werdell et al., 2019), will provide new opportunity to the use of wavelength-resolved IPP models. An alternative method to model vertical distributions of Chl*a* based on biogeographic provinces and seasons may prove useful, especially for generating climatological IPP products (Kulk et al., 2020; Sathyendranath et al., 1995). In addition, the vertical measurements of downwelling PAR and Chl*a* obtained by advanced instruments, such as space-borne or airborne light detection and ranging (Lidar) (Behrenfeld et al., 2013; Jamet et al., 2019) and biogeochemical (BGC)-Argo floats (Xing and Boss, 2021), could be combined with the TPM_{ERFR} and other satellite observations to provide refined IPP distributions.

Chapter 5. Spatiotemporal variability analysis

5.1. Characteristics of climatological annual IPP

5.1.1. Spatial patterns of annual IPP

Global distributions of climatological annual IPP over the period 1998–2021 generated by different IPP models were shown in Figure 5.1. Most of these IPP distributions shared a common spatial pattern, and several major oceanographic features were apparent. Generally, IPP was relatively low in ocean gyres, such as the north and south gyres in the Pacific and Atlantic oceans, whereas equatorial oceans and coastal water areas were generally dominated by high IPP, especially for those upwelling areas where the waters are usually nutrient-rich and so more productive. The belt of high IPP along the southern subtropical front (around 40°S) was also a recognizable characteristic in most of IPP distributions. By comparison, the $CAFE_{OP}$ provided a relatively higher IPP distribution for almost global oceans, and the IPP features in ocean gyres were less obvious as displayed in Figure 5.1a. On the contrary, the $AbPM_{GC}$ provided a lower IPP distribution as shown in Figure 5.1d, where the IPP values along the equatorial and southern subtropical zones were significantly lower. As expected, the IPP distributions from the $VGPM_{OP}$ and $VGPM_{BF97}$ were in well agreement (Figure 5.1b and Figure 5.1e), though there was small difference in their IPP magnitudes because different satellite products were used as input. The $VGPM_{BF97}$ might provide overestimated IPP in coastal and inland water areas as indicated by our validations (Section 4.2).

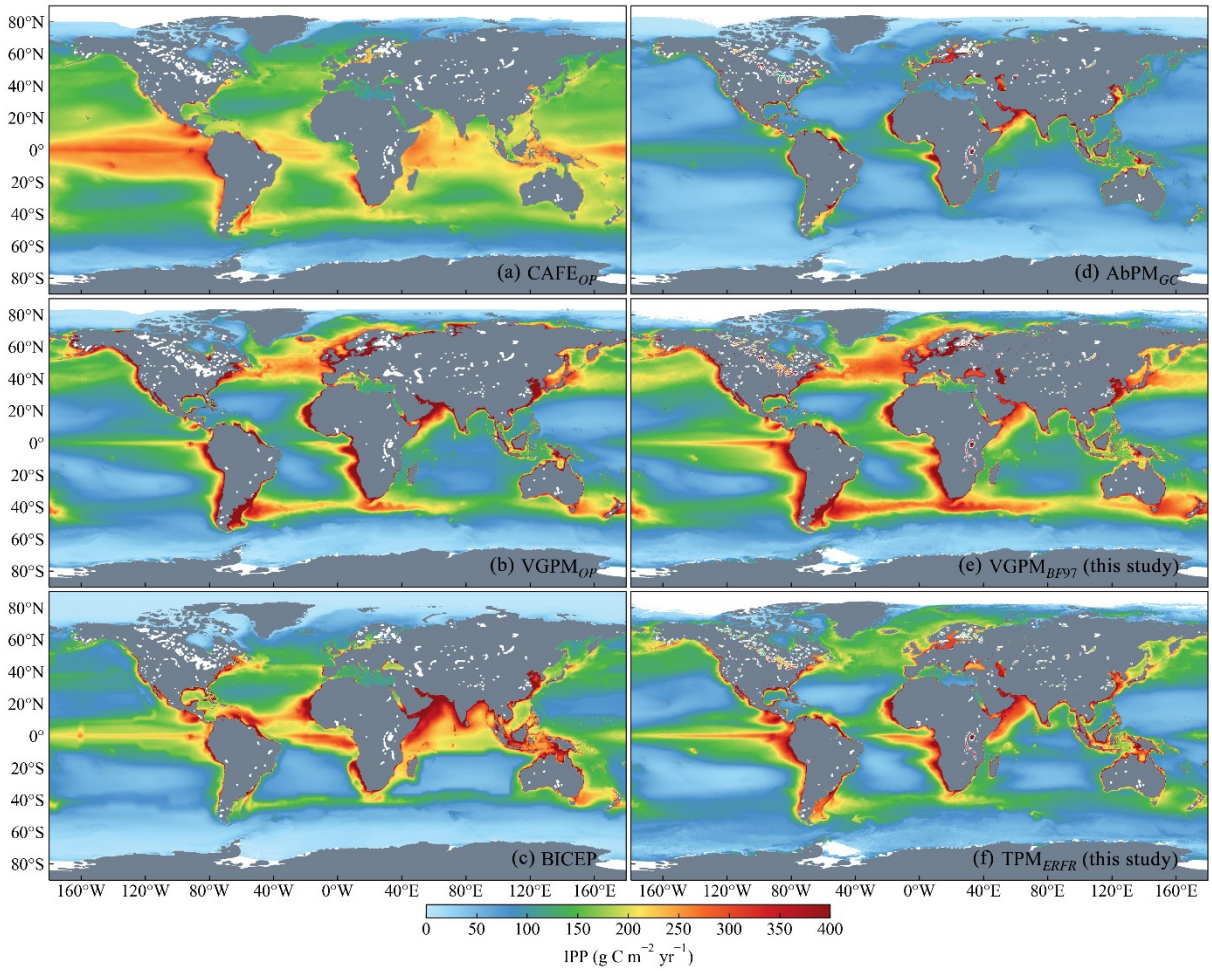

Figure 5.1. Global distributions of climatological (1998–2021) annual IPP (g C m^{-2} yr^{-1}) generated by different models/products: CAFE$_{OP}$ (a), VGPM$_{OP}$ (b), BICEP (c), AbPM$_{GC}$ (d), VGPM$_{BF97}$ (e), and TPM$_{ERFR}$ (f).

Figure 5.2 shows differences (ΔIPP) between IPP distributions generated by individual existing models and the one generated by the TPM_{ERFR}, where ΔIPP higher than 0 g C m^{-2} yr^{-1} indicates that the estimated IPP is higher than that of the TPM_{ERFR}. The $CAFE_{OP}$ typically yielded higher IPP values than the TPM_{ERFR} in the area between 40°S and 40°N (ΔIPP varied from ~30 g C m^{-2} yr^{-1} to 110 g C m^{-2} yr^{-1}), except for some coastal areas, such as the nearshore area of the Arabian Sea and the marginal seas of China (Figure 5.2a). Whereas the $AbPM_{GC}$ generally provided lower IPP values in high latitude water areas poleward of 50°S or 50°N and relatively more consistent estimates in ocean gyres, compared with those of the TPM_{ERFR} (Figure 5.2d). As shown in Figure 5.2b and Figure 5.2e, the estimated IPP values from the $VGPM_{OP}$ and $VGPM_{BF97}$ were higher than those from the TPM_{ERFR} in medium or high latitude ocean regions (20°N to 60°N, 20°S to 50°S), especially along the southern subtropical front and in the North Atlantic Ocean (ΔIPP larger than ~50 g C m^{-2} yr^{-1}). However, these two models had lower IPP values in the polar waters of the Antarctic Ocean (poleward of 50°S), with ΔIPP of –30 g C m^{-2} yr^{-1} approximately. Moreover, recall that the BICEP utilized LUT methods to estimate IPP on basis of biogeographic provinces (Kulk et al., 2020), and thus there were inevitably distinct boundaries found in the resultant IPP distribution (Figure 5.1c) and the corresponding ΔIPP map (Figure 5.2c). The BICEP had higher IPP values mainly in the ocean gyres of the Pacific and North Atlantic oceans, the Arabian Sea, and the marginal seas of China, but it gave lower IPP values mainly in polar oceans and several upwelling zones.

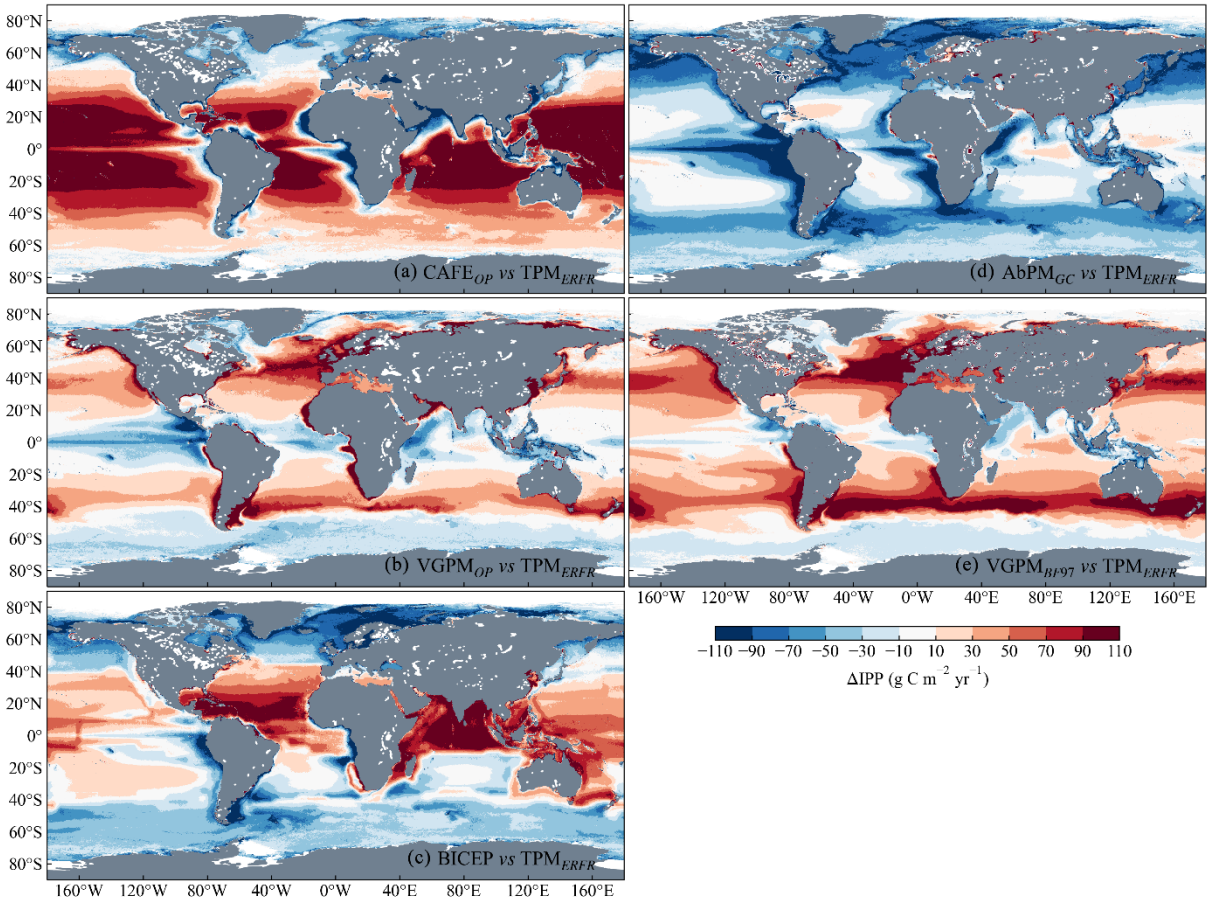

Figure 5.2. Differences (ΔIPP) between the global distributions of climatological (1998–2021) annual IPP (g C m^{-2} yr^{-1}) generated by the existing IPP models/products (CAFE$_{OP}$ (a), VGPM$_{OP}$ (b), BICEP (c), AbPM$_{GC}$ (d), and VGPM$_{BF97}$ (e)) and the one generated by the TPM$_{ERFR}$. The value of ΔIPP higher than 0 g C m^{-2} yr^{-1} means that the estimated IPP of a given existing IPP model is higher than that of the TPM$_{ERFR}$.

Figure 5.3 presents the density histograms of climatological annual IPP and corresponding ΔIPP compared with the TPM_{ERFR}. Typically, the IPP estimates from the $CAFE_{OP}$ and $AbPM_{GC}$ varied between ~0 and 300 g C m^{-2} yr^{-1}, while those from the $VGPM_{OP}$, $VGPM_{BF97}$, BICEP, and TPM_{ERFR} could reach about 400 g C m^{-2} yr^{-1}, as the spatial patterns revealed by these four models were similar (Figure 5.1). On the other hand, the calculated ΔIPP generally had a common range between –100 and 100 g C m^{-2} yr^{-1}, except for the ΔIPP values from the $AbPM_{GC}$ (almost all of them were lower than 0 g C m^{-2} yr^{-1}). Overall, the $CAFE_{OP}$ and $VGPM_{BF97}$ were particularly prone to yield higher IPP estimates, whereas the $AbPM_{GC}$, the BICEP, and the $VGPM_{OP}$ were more likely to provide lower IPP estimates.

Given the fact that the global distributions of climatological annual IPP from different products showed some common spatial characteristics but also exhibited different magnitudes in different regions, the uncertainties in current satellite-based IPP products were confirmed in this study, which has been stated in the recent IPCC report (Gulev et al., 2021). The remote estimation of photosynthetic parameters can be one of the main factors that cause the discrepancy. Nevertheless, the inter-comparison among these IPP products can greatly contribute to identifying the areas with high confident IPP estimates, such as the areas where all IPP models gave a consistent estimation. Furthermore, the inter-comparison can also highlight the areas with low confidence of IPP estimation, and thus will point out a direction for future assessments and improvements.

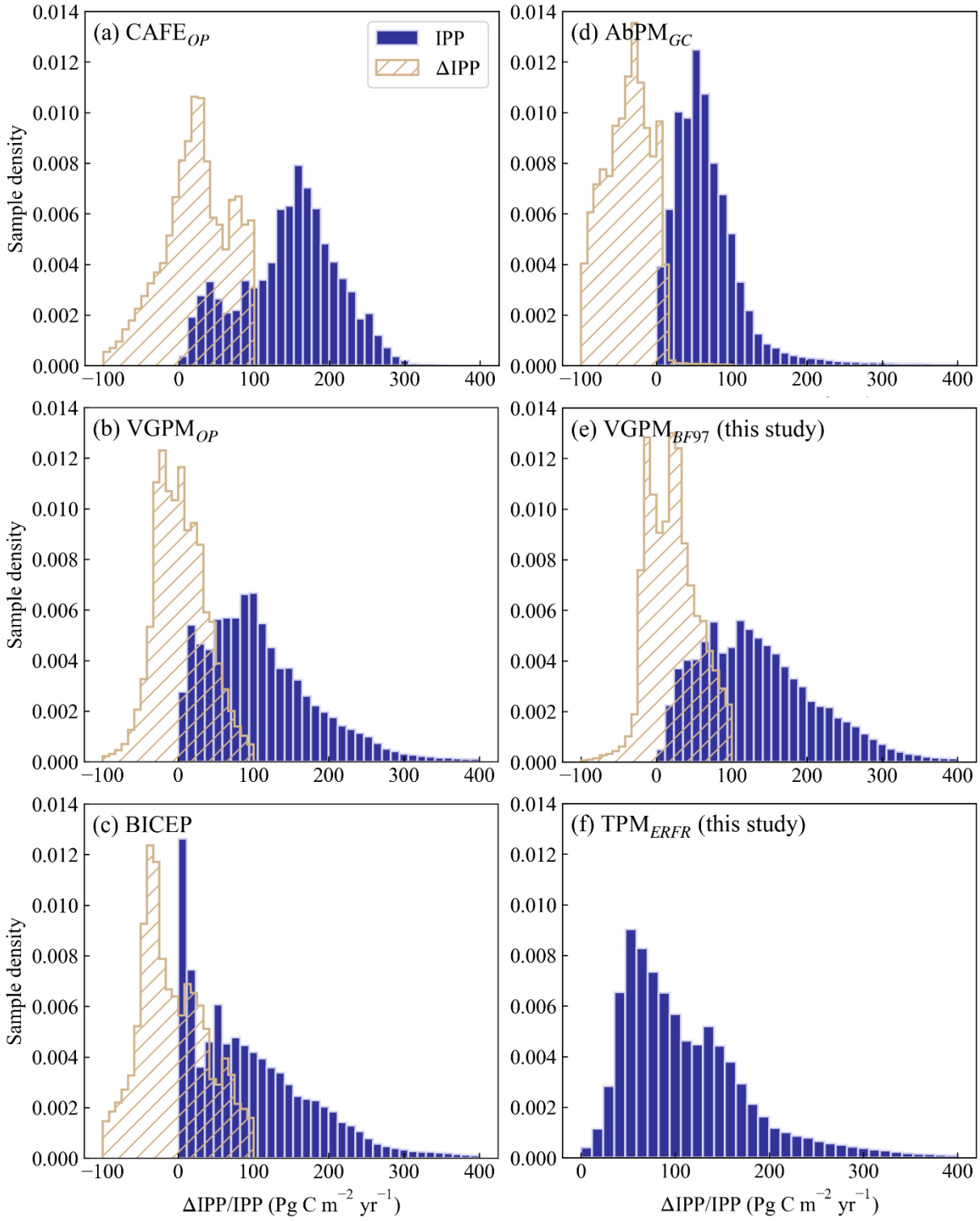

Figure 5.3. Density histograms of global climatological (1998–2021) annual IPP (g C m^{-2} yr^{-1}) generated by different IPP models/products and corresponding differences (ΔIPP) compared with the TPM$_{ERFR}$.

5.1.2. Contribution analysis of ocean basins and biomes

The percent contributions of individual ocean basins (the Atlantic, Pacific, Indian, Antarctic, and Arctic basins) and biomes (the Polar, Westerlies, Trades, and Coastal biomes) (see Figure 3.3) to global climatological annual IPP were summarized in Figure 5.4 and Figure 5.5, respectively. In addition to the four biomes defined earlier, the Inland biome was defined here as those inland areas that had valid IPP estimates in satellite products. Only the IPP products from the $AbPM_{GC}$, $VGPM_{BF97}$, and TPM_{ERFR} had significant estimates in the Inland biome because most of inland water pixels were excluded in other products (Figure 5.1). It is noteworthy that the missing data in polar regions were not the same in the IPP products due to different procedures for masking sea ice pixels. However, since the number of missing data is small, it has negligible effect on the inter-comparison among percent contributions of the Polar biome from different products.

As indicated by the results shown in Figure 5.4, although the total amounts of IPP in individual ocean basins were different for these IPP products, their basin contributions could be generally similar. The Pacific Ocean had the largest contribution, with the propotions of global IPP ranging between 41.3 and 48.4%. As the ocean basin that had the second largest contribution, the Atlantic Ocean typically contributed 20.3 to 27.7% of annual IPP. The percent contributions of the Indian Ocean were larger than those of the Antarctic Ocean for the IPP products estimated by the $CAFE_{OP}$, $VGPM_{OP}$, BICEP, and $AbPM_{GC}$. Especially

for the BICEP, the annual IPP estimated in the Indian Ocean (8.3 g C m^{-2} yr^{-1}, 18.0%) was twofold larger than that in the Antarctic Ocean (4.1 g C m^{-2} yr^{-1}, 9.0%). In contrast, the TPM_{ERFR} and $VGPM_{BF97}$ gave larger contributions of the Antarctic Ocean than those of the Indian Ocean. Generally, the percent contributions of the Indian and Antarctic oceans lie in the ranges of 13.5–18.0% and 9.0–15.4%, respectively. The smallest contribution came from the Arctic Ocean for all IPP products, varying between 1.5 and 3.8%.

Biome contributions to global annual IPP calculated from different IPP products also offer useful information on the variable characteristics of carbon fixation by phytoplankton unveiled based on different IPP models. As shown in Figure 5.5, the Trades biome contributed the largest part of annual IPP, with the percent contributions varied from 36.4 to 47.1%, except for the products generated by the $VGPM_{OP}$ and $VGPM_{BF97}$ (Figure 5.5b and Figure 5.5e), which only contributed 32.9% and 34.8% of annual IPP, respectively. Likewise, the Westerlies biome was prone to have the second largest contribution, ranging between 26.6 and 35.7%, for the IPP products of the $CAFE_{OP}$, BICEP, $AbPM_{GC}$, and TPM_{ERFR}. Moreover, even though the coverage of coastal waters is relatively smaller than open oceans (Figure 3.3b), the Coastal biome typically had a considerable contribution, yielding 20.2 to 25.6% of global annual IPP, except for the $CAFE_{OP}$. It implies that the $CAFE_{OP}$ may not be applicable to turbid and shallow waters, which is mainly because its strategy of parameterization for photosynthetic parameters was built for open oceans (Silsbe et al., 2016).

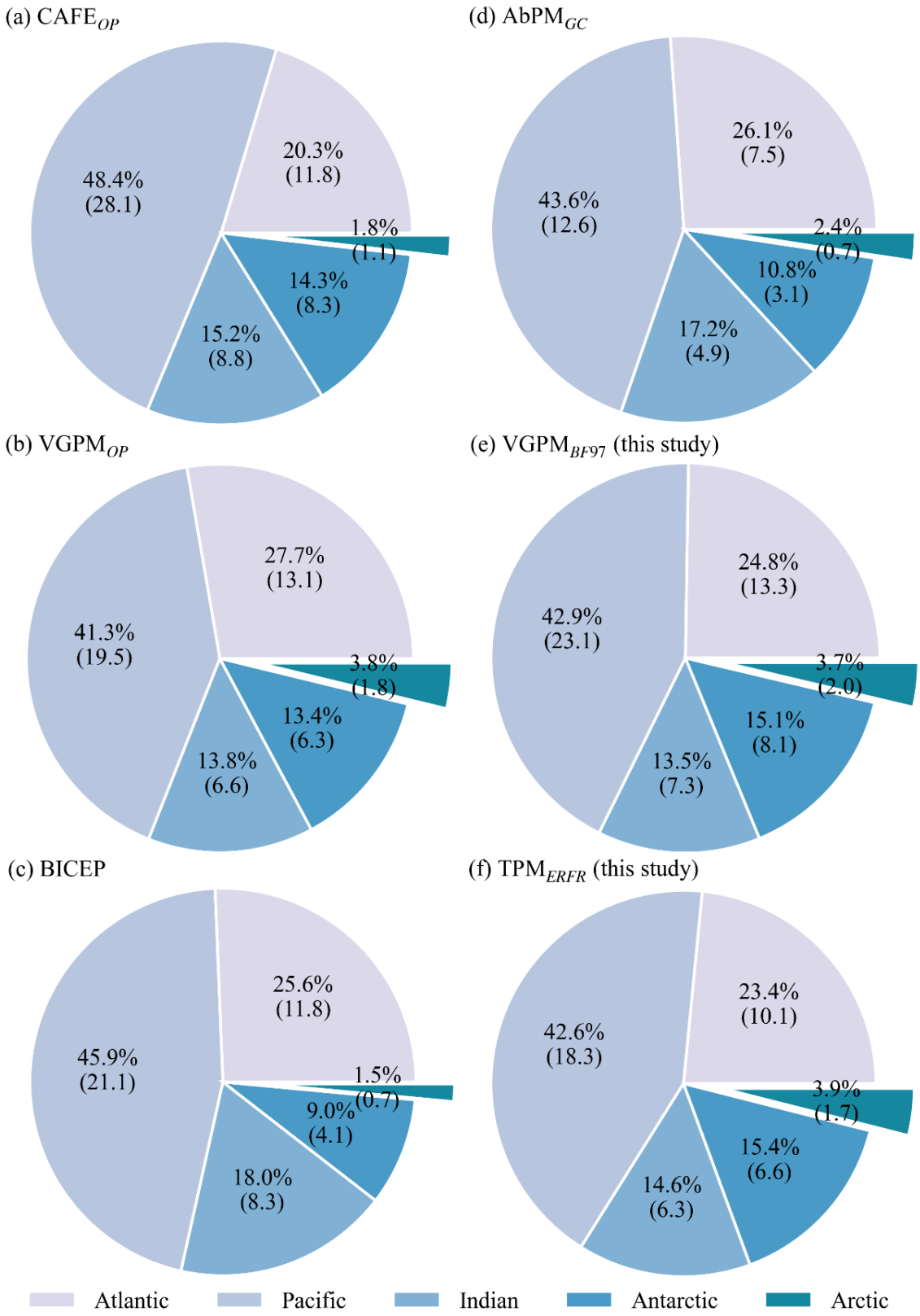

Figure 5.4. Percent contributions (%) of ocean basins to global climatological annual IPP (Pg C yr^{-1}) calculated from different IPP products for the period 1998–2021. The values in parentheses are the corresponding climatological annual IPP at basin scale.

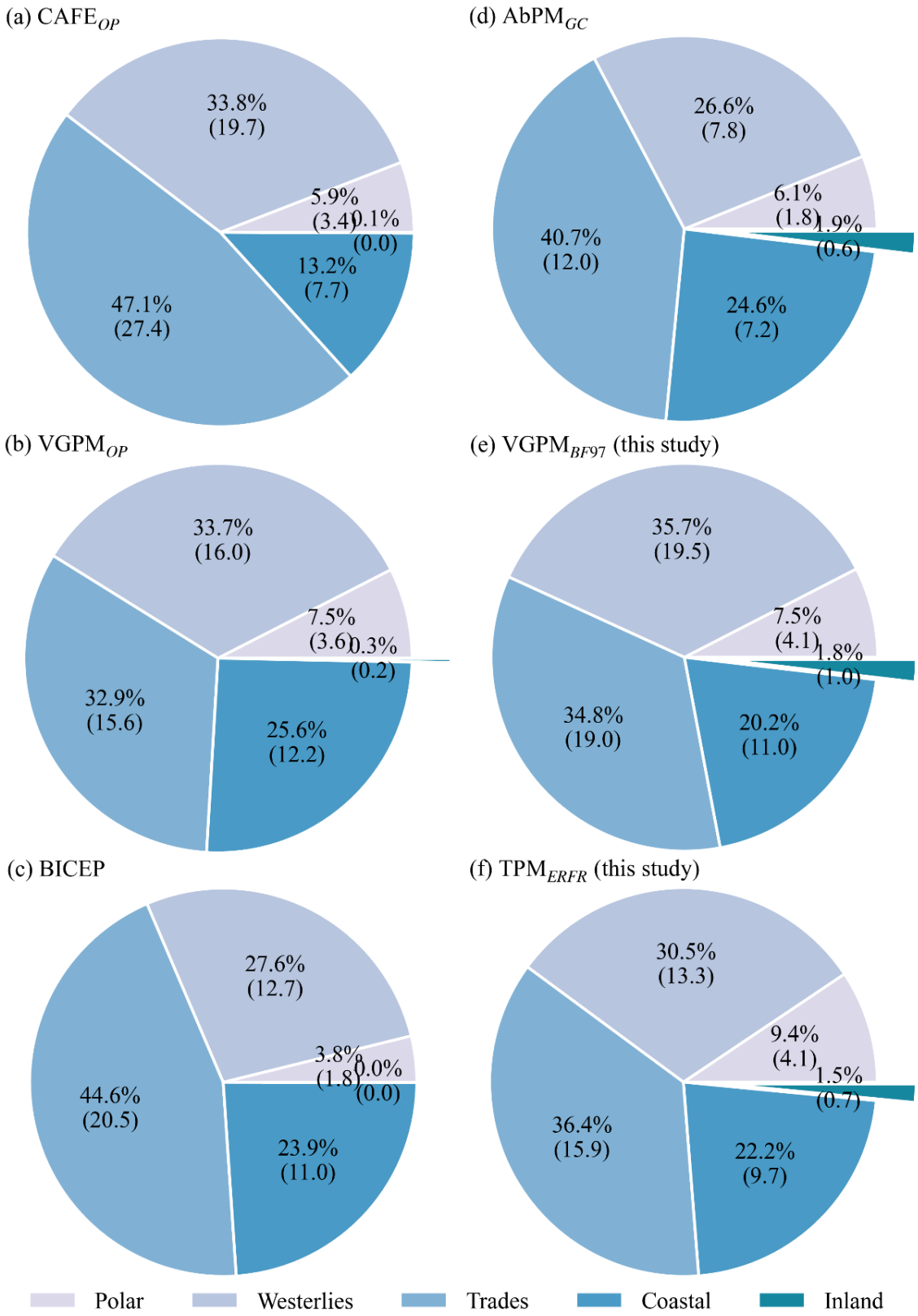

Figure 5.5. Percent contributions (%) of ocean biomes to global climatological annual IPP (Pg C yr^{-1}) calculated from different IPP products for the period 1998–2021. The values in parentheses are the corresponding climatological annual IPP at biome scales.

The BICEP-based IPP product gave the smallest contribution of the Polar biome, which was only 3.8%, whereas the percent contributions calculated by other products could reach 5.9–9.4%. Note that the Inland biome generally contributed 1.5 to 1.9% of global annual IPP for the products of the $AbPM_{GC}$, $VGPM_{BF97}$, and TPM_{ERFR}, and given this small proportion for inland waters, the percent contributions in other biomes from different IPP products can still be safely compared.

Furthermore, Figure 5.6 shows the annual IPP estimates in individual ocean basins and biomes from the six satellite-based products. As expected, the $CAFE_{OP}$ and $AbPM_{GC}$ provided the highest and lowest estimates of global annual IPP, as 58.2 Pg C yr^{-1} and 29.4 Pg C yr^{-1}, respectively. The $VGPM_{BF97}$ gave the second highest estimate of 54.8 Pg C yr^{-1}. By comparison with the $VGPM_{OP}$-based estimate of 47.5 Pg C yr^{-1}, the effect of input satellite products on IPP estimation for a given model is proved to be fundamental, in particular, for the studies about carbon sinks/sources of aquatic ecosystems. The TPM_{ERFR} yielded an estimate of global annual IPP as 43.7 Pg C yr^{-1}, which was relatively closer to that of the BICEP (46.0 Pg C yr^{-1}) as another product based on the OC-CCI data. Typically, as the results shown in Figure 5.1–5.5, relatively large differences of annual IPP estimates existed in the Trades biome and the Pacific Ocean, resulting in a bias of ~15 Pg C yr^{-1}. The Westerlies biome also had obvious discrepancy, especially for the Pacific and Antarctic Ocean sectors, causing the biases of ~6.6 Pg C yr^{-1} and 4.5 Pg C yr^{-1}, respectively.

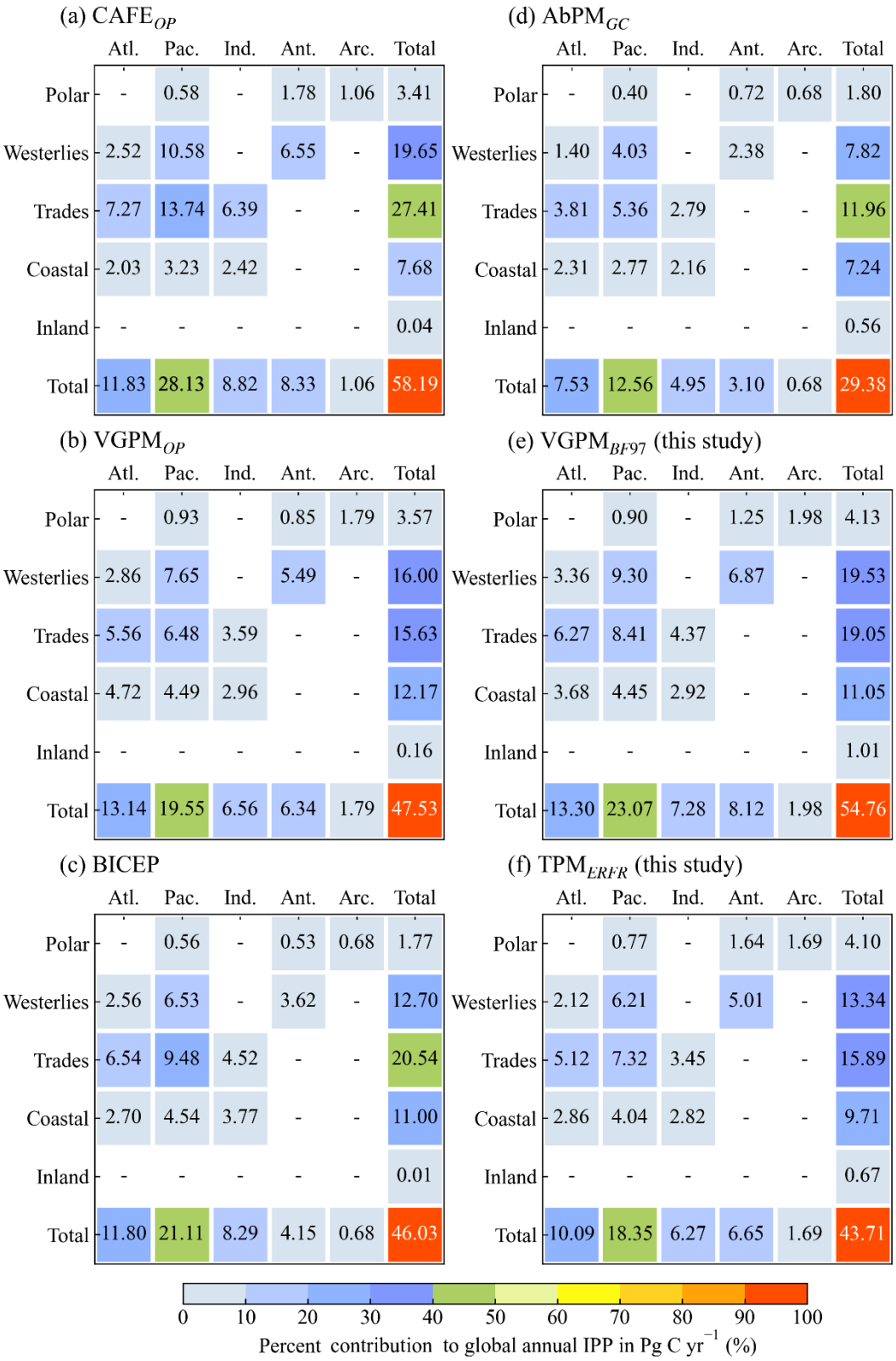

Figure 5.6. Climatological (1998–2021) annual IPP (Pg C yr^{-1}) in each ocean basin (Atl. – Atlantic, Pac. – Pacific, Ind. – Indian, Ant. – Antarctic, and Arc. – Arctic) and biome (Polar, Westerlies, Trades, Coastal, and Inland) estimated by different IPP models/products.

5.1.3. Reassessment of annual IPP: consistency and discrepancy

Earlier estimates of global annual IPP based on mainstream models generally ranged from 36.5 to over 60.0 Pg C yr^{-1} (Carr et al., 2006; Sathyendranath et al., 2019). The TPM_{ERFR} estimated that the global climatological annual IPP was 43.7 Pg C yr^{-1} for the last 24 years, which lies in the range of these reported inter-comparisons. Table 5.1 summarizes several estimates of annual IPP at both basin and global scales that have been published over the past two decades, including the results from the TPM_{ERFR} and other IPP models/products used in this study. By partitioning the global annual IPP among ocean basins, we could further investigate discrepancies between our estimates and previously reported values. Note that the periods of time associated with the IPP estimates and the ways used to distinguish ocean basins in the cited studies and this study are not the same, and the result from a regional study was thought to be more confident than that from a global study. It is nevertheless instructive to assess these results in the same context. On basis of the $AbPM_{AM96}$ with a pre-generated LUT, Antoine et al. (1996) had reported an estimate of 36.5 Pg C yr^{-1}, however, the official IPP product of the $AbPM_{GC}$ provided a lower estimate of 29.4 Pg C yr^{-1} for the last 24 years. Rousseaux and Gregg (2014) also gave a similar estimate of 39.0 Pg C yr^{-1}, by using a numerical simulation model, the NASA Ocean Biogeochemical Model (NOBM). But the Arctic Ocean was not included in their research, which means that the actual estimate of global annual IPP based on the NOBM could be about 40.0 Pg C yr^{-1}, approaching to the estimate of the TPM_{ERFR}.

Table 5.1. Summary of annual IPP (unit: Pg C yr^{-1}) estimates at basin and global scales. The results reported in previous studies and those provided in this study based on the TPM_{ERFR} (highlighted in bold) and other IPP models/products are presented. The IPP estimates in the Mediterranean (Med.) were extracted from the Atlantic provinces to be consistent with other results.

Reference/Source	Year	Global	Ocean basins					Med./Inland
			Pacific	Atlantic	Indian	Antarctic	Arctic	
Longhurst et al., 1995	1978–1986	50.3	19.7	14.8	6.5	8.2	1.1	–/–
Antoine et al., 1996	1978–1986	36.5	15.9	9.9	6.6	4.0	0.2	–/–
Smyth et al., 2005	2000–2003	52.2	23.1	11.1	8.1	9.0	0.5	0.4/–
Carr et al., 2006[1]	1998	50.7	21.0	12.8	9.9	2.6	0.3	0.5/–
Arrigo et al., 2008	1998–2006	–	–	–	–	1.9	–	–/–
Tilstone et al., 2009	1998–2005	–	–	10.2	–	–	–	–/–
Uitz et al., 2010	1998–2007	45.6	20.0	12.2	9.0	3.5	0.5	0.5/–
Hill et al., 2013	1998–2007	–	–	–	–	–	1.0	–/–
Rousseaux and Gregg, 2014[2]	1998–2011	39.0	17.2	8.7	8.6	4.5	–	–/–
Brewin et al., 2017	2007	–	–	7.9	–	–	–	–/–
Kulk et al., 2021b	1998–2018	50.7	22.3	14.3	8.2	5.9	–	–/–
Silsbe et al., 2016/ $CAFE_{OP}$	1998–2021	58.2	28.1	11.8	8.8	8.3	1.1	–/–
Behrenfeld and Falkowski, 1997/ $VGPM_{OP}$	1998–2021	47.5	19.6	13.1	6.6	6.3	1.8	–/–
Antoine et al., 1996/ $AbPM_{GC}$	1998–2021	29.4	12.6	7.5	5.0	3.1	0.7	–/0.60
Kulk et al., 2020/ BICEP	1998–2020	46.0	21.1	11.8	8.3	4.2	0.7	–/–
Behrenfeld and Falkowski, 1997/ This study	1998–2021	54.8	23.1	13.3	7.3	8.1	2.0	0.6/1.0
This study	**1998–2021**	**43.7**	**18.4**	**10.1**	**6.3**	**6.7**	**1.7**	**0.4/0.7**

[1] A total of 31 IPP models were tested, including 24 ocean-color IPP models and 7 general circulation models (GCMs). The mean results of all tested model are shown here.

[2] The NASA Ocean Biogeochemical Model (NOBM), which belongs to GCM, was used to estimate global phytoplankton primary production.

Obvious discrepancies are apparent in the Pacific and Indian oceans, where the estimated annual IPP values in previous studies have fallen in the range 12.6−28.1 Pg C yr^{-1} and 5.0−9.9 Pg C yr^{-1}, respectively. The corresponding annual IPP values estimated from the TPM_{ERFR} in these two areas, 18.4 Pg C yr^{-1} and 6.3 Pg C yr^{-1}, respectively, fell below the lower bounds of the reported ranges. Friedrichs et al. (2009) have confirmed that the modified model of Longhurst et al. (1995) performs best in the tropical Pacific Ocean, where its *RMSD* is only 0.23. For the Indian basin, by a similar assessment, Saba et al. (2011) have verified that the model of Smyth et al. (2005) is the best model in the Arabian Sea because of its improved performance in Case-II waters. But similar validations have not been made in the Westerlies biome of the Pacific basin and the Trade biome of the Indian basin. Without adequate *in situ* data for further assessment, it is difficult to reach a confident conclusion about the causes of the discrepancy of annual IPP estimates in these two undersampled ocean basins.

The investigations have been more exhaustive in the Atlantic basin than the Pacific and Indian basins. Tilstone et al. (2009) have shown that a wavelength-resolved model based on the P-E curve works best in the Atlantic basin and have estimated an annual IPP of 10.2 Pg C yr^{-1}. More recently, by using a TPM-like model with parameters that were well calibrated using cruise datasets from the Atlantic basin to model size-fractionated IPP, Brewin et al. (2017) obtained a more conservative estimate of 7.9 Pg C yr^{-1}. These results suggest that a reasonable range of annual IPP in the Atlantic basin is likely ~7.9−10.2 Pg C yr^{-1}

without significant interannual variability, which includes our estimate of 10.1 Pg C yr^{-1}. The annual IPP estimates for the Mediterranean region are quite similar and lie in the range 0.4−0.6 Pg C yr^{-1}. Because *in situ* data from the Mediterranean were included for the development of the ERFR (Table 2.1), the TPM_{ERFR}-based estimate of 0.4 Pg C yr^{-1} was expected to be accurate in the Mediterranean.

It has been widely acknowledged that accurate modeling of IPP based on satellite observations is challenging in polar areas, especially in the Southern (Antarctic) Ocean (Carr et al., 2006). Mainstream IPP models typically perform poorly in the Arctic Ocean ($RMSD > 0.5$) mainly because of the uncertainty in satellite-estimated Chl*a* and their failure to account for DCMs (Lee et al., 2015). Based on both *in situ* and satellite data, Hill et al. (2013) have reported that the estimated annual IPP in the Arctic Ocean is about 1.0 Pg C yr^{-1} for the period 1998−2007, if the effect of DCMs is included. Though this estimate is close to those estimated by the model of Longhurst et al. (1995) (1.1 Pg C yr^{-1}), $CAFE_{OP}$ (1.1 Pg C yr^{-1}), $AbPM_{GC}$ (0.7 Pg C yr^{-1}), and BICEP (0.7 Pg C yr^{-1}), it has been reported that the loss of sea ice in the Arctic Ocean has leaded to an increase of phytoplankton PP (Lewis et al., 2020). It means that the relatively higher estimates, such as the one from the TPM_{ERFR} (1.7 Pg C yr^{-1}), could be theoretically more reasonable for the period 1998−2021. For the Antarctic Ocean, the fact that the annual IPP estimates have fallen in the range 1.9−9.0 Pg C yr^{-1} implies a large uncertainty in the modeling of IPP in these polar waters. The wide range of the estimates can in part be attributed to the inconsistent definition of the Antarctic

Ocean. Longhurst (2007) and Rousseaux and Gregg (2014) have considered the area poleward of 40°S to be the Antarctic Ocean (Figure 3.3). In contrast, the annual IPP of ~1.9 Pg C yr^{-1} estimated by Arrigo et al. (2008) applies to the area poleward of 50°S. This area is relatively closer to the Polar biome of the Antarctic basin, and a more consistent estimate of 1.6 Pg C yr^{-1} was yielded by the TPM_{ERFR} for this biome (Figure 5.6f). Besides, Hirawake et al. (2011) have demonstrated that the VGPM was prone to significantly overestimate IPP along the southern subtropical front. It indicates that, in practice, the annual IPP of the Westerlies biome of the Antarctic Ocean is very probably less than 5.5 Pg C yr^{-1} according to the results from the $VGPM_{OP}$ and $VGPM_{BF97}$ (Figure 5.6b and Figure 5.6e). These results suggest that the TPM_{ERFR}-based estimate of IPP in the Antarctic Ocean may well be reasonable.

The tendency of the $VGPM_{BF97}$ to overestimate IPP has been found not only in the Rhode River and Lake Kasumigaura but also in other coastal and inland regions as suggested in several previous studies (Harding et al., 2002; Hirawake et al., 2012; Yoo et al., 2019). Saba et al. (2011) have concluded that the $VGPM_{KI05}$ was the best of 21 tested models for modeling IPP in shallow waters (< 250 m), but our validation shows that the $VGPM_{KI05}$ has limited applicability in turbid inland waters. Given the fact that the TPM_{ERFR} enabled more accurate estimates of IPP in the Rhode River and Lake Kasumigaura than the other benchmark models, we can reasonably expect that the TPM_{ERFR} has the potential to improve IPP estimates in the Coastal and Inland biomes. Compared with traditional

investigations based on field data collections (Cloern et al., 2014), future utility of the TPM_{ERFR} is promising to reveal the variability of IPP in global coastal and inland aquatic ecosystems by providing synoptic observations.

Based on the reassessment of annual IPP at basin and global scales, the consistency between the TPM_{ERFR}-based and previously reported estimates was found in most of ocean basins, and the improvement in Coastal and Inland biomes can be expected with reliable satellite products as input. The global climatological (1998–2021) annual IPP of 43.7 Pg C yr^{-1} was yielded by the TPM_{ERFR}, which lied in a reasonable range but was relatively lower than the average estimate of 50.0 Pg C yr^{-1} based on historical records of remote sensing studies. Given different time periods of IPP estimates, the fact that a decline trend in global IPP of –0.8 Pg C yr^{-1} $decade^{-1}$ from 1998 to 2015 has recently been reported (Gregg and Rousseaux, 2019) could be a plausible explanation for the discrepancy. In addition, the loss of phytoplankton biomass due to the uncertainty of remotely retrieved Chl*a* profile (especially in oligotrophic waters where the presence of DCMs is common) could be another reason, which is a common issue for depth-resolved IPP models. Still, by the inter-comparisons with the previously reported estimates at basin scale and the forementioned *in situ*/satellite-based validation results, we can conclude that the TPM_{ERFR} can provide confident estimation of IPP at global scale. Nevertheless, only with a sufficient number and coverage of *in situ* measurements could we more thoroughly evaluate the TPM_{ERFR} for every ocean basin and biome, which will require a great deal of efforts in the future.

5.2. Seasonal variation of IPP

5.2.1. Seasonal signals of global IPP

The monthly changes of global climatological IPP estimated by different models/products are presented in Figure 5.7. Similar variation of monthly IPP can be identified for the $AbPM_{GC}$, $VGPM_{OP}$, $VGPM_{BF97}$, and TPM_{ERFR}, though their IPP amplitudes were variable and the seasonal signal of the $AbPM_{GC}$-based IPP was not as significant as those of other three models. Among these four IPP products, the $VGPM_{OP}$ and $VGPM_{BF97}$ exhibit the strongest seasonal fluctuation, varying from ~3.4 to 4.8 Pg C $month^{-1}$ and from ~4.0 to 5.2 Pg C $month^{-1}$, respectively, with a seasonal amplitude of IPP over 1 Pg C $month^{-1}$. A plausible explanation is the linkage between P_{opt}^{B} and SST as the parameterization of $VGPM_{OP}$ and $VGPM_{BF97}$ (Behrenfeld and Falkowski, 1997). In contrast, the $AbPM_{GC}$- and TPM_{ERFR}-based IPP estimates have relatively less prominent seasonal variations, varying from ~2.2 to 3.0 Pg C $month^{-1}$ and from ~3.1 to 4.2 Pg C $month^{-1}$, respectively. This kind of mild seasonal signal of global IPP has been unveiled in several earlier studies (e.g. Antoine et al., 1996; Smyth et al., 2005), and the reason can be the opposite seasonal variations in the Northern and Southern hemispheres. Typically, global oceans are more productive in two time periods, one is between Jun. and Aug. and another one is between Dec. and Jan., as indicated by two peaks of IPP found in monthly time series. For the $VGPM_{OP}$, $VGPM_{BF97}$, and $AbPM_{GC}$, the IPP peak occurred bewteen Jun. and Aug. is larger than the one occurred between Dec. and Jan., especially for the first two models.

Whereas for the TPM_{ERFR}, the two peaks of IPP are comparable. A distinctive seasonal variation can be found in the monthly time series of the $CAFE_{OP}$-based IPP, which is that global oceans generally reach their highest IPP values between Oct. and Jan., followed by a sharp decrease from Jan. to Feb., and become the least productive in Jun., as displayed in Figure 5.7a. Likewise, the monthly change of BICEP-based IPP also exhibits a different pattern compared with others. According to the result from the BICEP, global oceans commonly experience high level of IPP through the period between Mar. and Oct., and become less productive between Nov. and Feb., as shown in Figure 5.7c. On basis of the quantitative evaluation for IPP estimates at different scales (Section 4.2 and Section 5.1) and the qualitative inter-comparison for seasonal signals of global IPP, we can conclude that the TPM_{ERFR} is able to provide realistic estimation of IPP amount and seasonal variation. Therefore, following analyses will focus on the result from the proposed TPM_{ERFR}.

Figure 5.8 presents the global distributions of climatological monthly IPP generated by the TPM_{ERFR} to demonstrate detailed seasonal changes in different regimes. The permanent feature of enhanced IPP can be observed in the equatorial upwelling zone, apart from the Indian Ocean sector, and in the coastal upwelling zones, such as those along the western coastline of Africa and the western Atlantic, because of the modulation of plentiful nutrient supplies in the upper water layers. In contrast, seasonal change of IPP in the Arabian Sea is mainly controlled by local seasonal upwelling system, which will be discussed in next section.

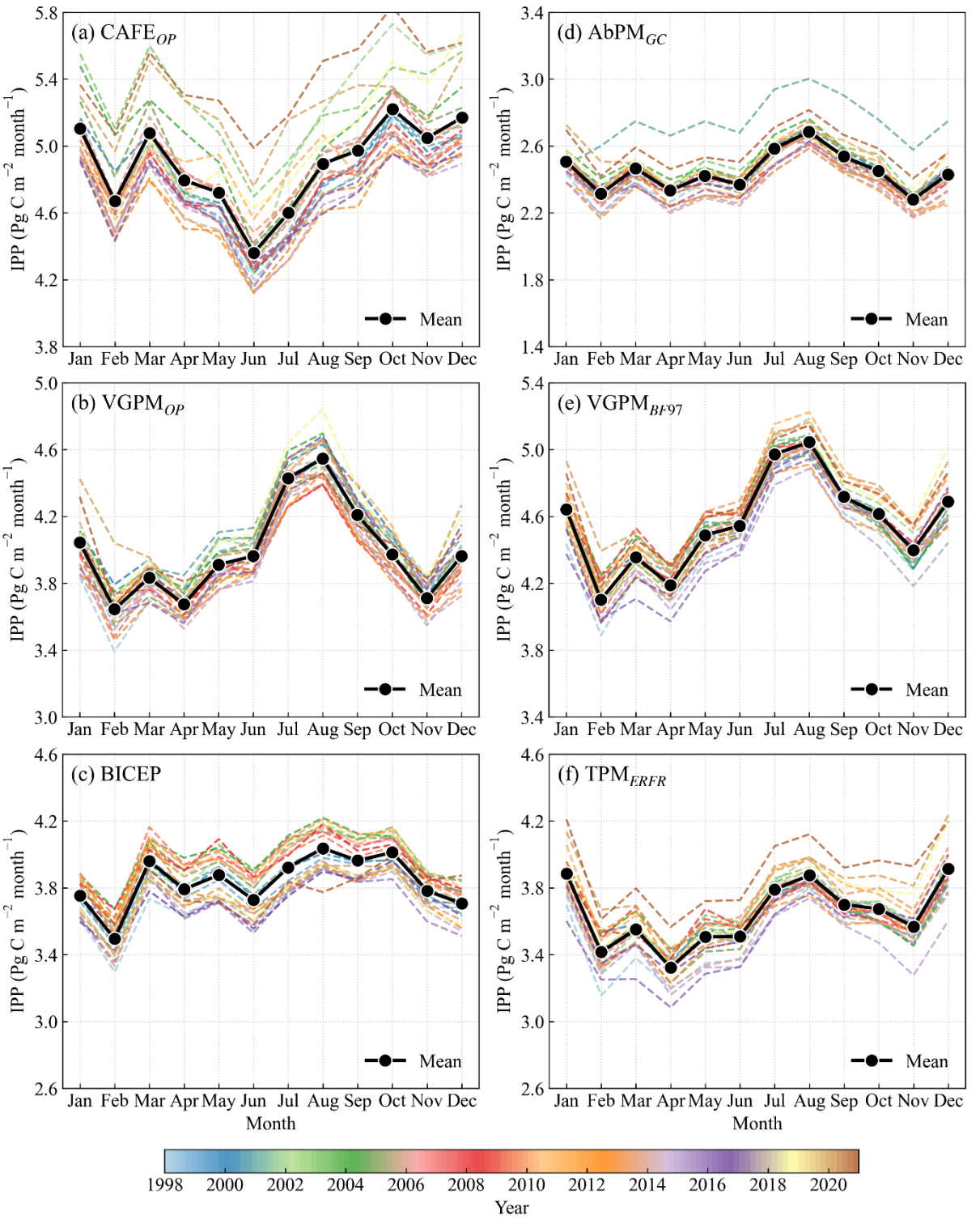

Figure 5.7. Monthly changes of global IPP estimated by different models/products: CAFE$_{OP}$ (a), VGPM$_{OP}$ (b), BICEP (c), AbPM$_{GC}$ (d), VGPM$_{BF97}$ (e), and TPM$_{ERFR}$ (f). The data in all available years are shown together with the climatological monthly mean IPP.

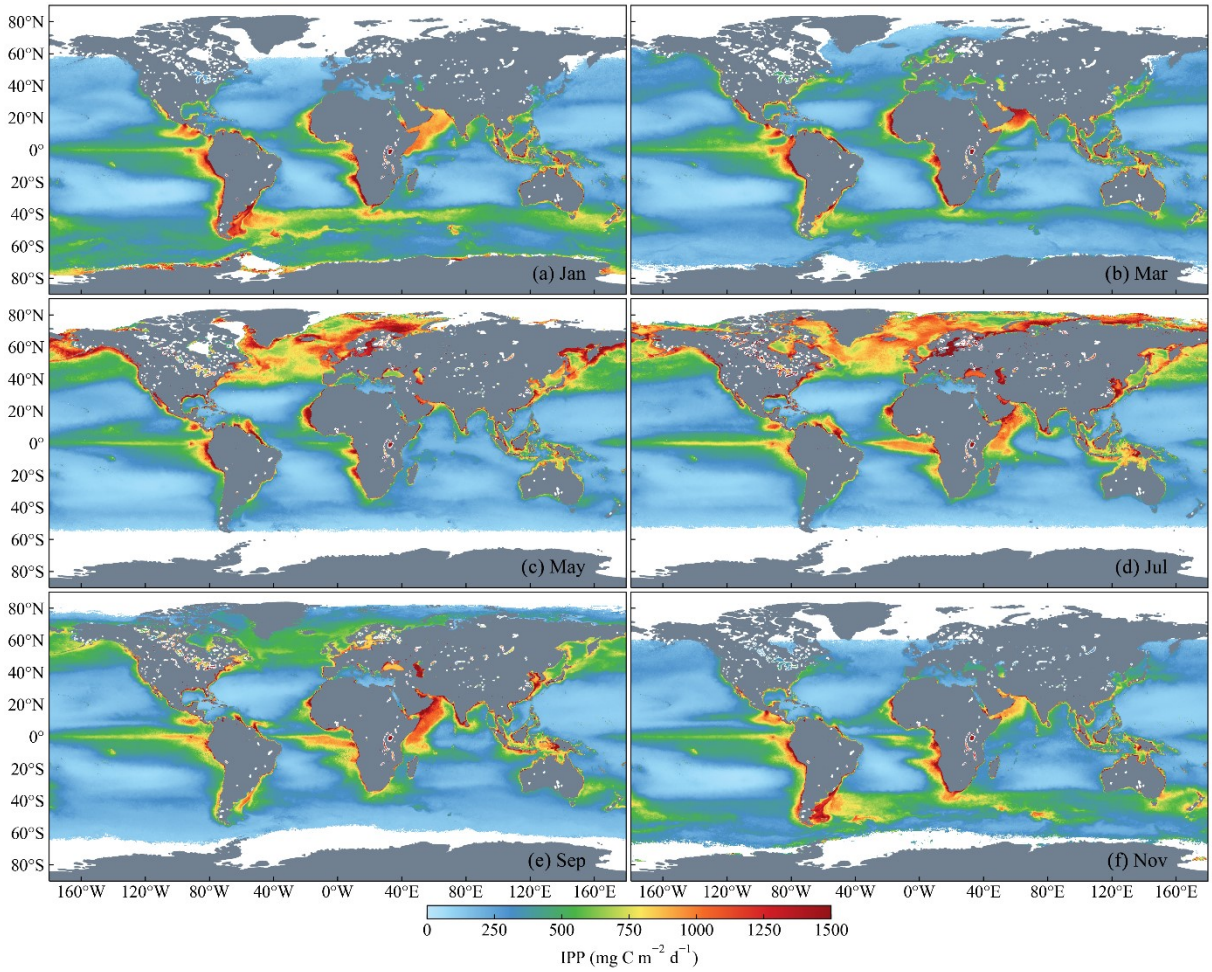

Figure 5.8. Global distributions of climatological (1998−2021) monthly IPP generated by the TPM$_{ERFR}$ in Jan. (a), Mar. (b), May (c), Jul. (d), Sep. (e), and Nov. (f) as a demonstration of typical seasonal change of IPP.

Morevoer, the monthly changes of IPP in the Arctic and Antarctic Ocean show an obvious seasonal pattern, that is, the waters gradually become more and more productive from the spring period for each hemishphere, and have the strongest bloom in the corresponding summer period. The ocean gyres are commonly characterized by stable features of low IPP due to the lack of nutrient. For the coastal and inland water areas in the Northern hemishpere, such as the Great Lakes and the marginal seas of China, the change of IPP follows the rule of low in winter and high in summer.

5.2.2. Interpretation of seasonal pattern in each ocean basin

The seasonal signals of IPP in the Pacific and Atlantic oceans were extracted from the global singal, as displayed in Figure 5.9, together with the monthly changes of related environmental parameters, including Chl*a*, SST, and *d*PAR. The seasonality in the Pacific and Atlantic oceans is not very prominent, with the IPP ranges of ~1.3–1.7 Pg C month^{-1} and ~0.7–1.0 Pg C month^{-1}, respectively. It is most likely because that the characteristics of both the equatorial and coastal upwelling zones and the gyre regimes that cover a large area of these two oceans are stable, as nutrient concentrations in these two types of water areas are permanently high and low, respectively. There are generally two peaks in Chl*a*, SST, and *d*PAR in the Pacific Ocean due to the integration of seasonalities from two hemispheres, which are not recognized in IPP change. While the seasonality of the Atlantic Ocean is more subject to the boreal summer conditions as indicated by the feature of one peak in IPP, SST, and *d*PAR.

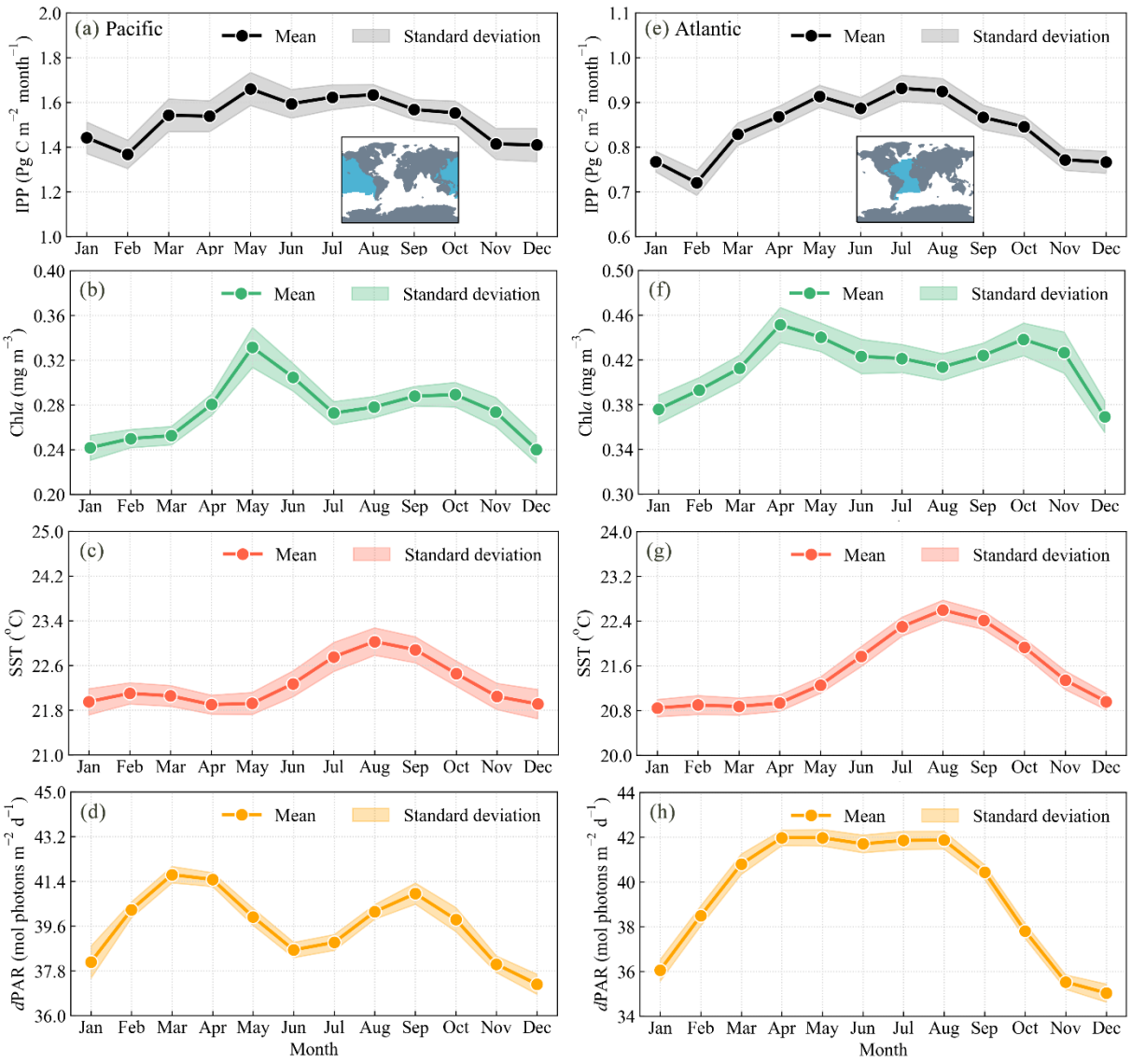

Figure 5.9. Monthly changes of the TPM$_{ERFR}$-based IPP (a and e), Chl*a* (b and f), SST (c and g), and *d*PAR (d and h) in the Pacific (right panel) and Atlantic oceans (left panel), respectively, in the period 1998−2021. Blue areas in the inset figures of (a) and (e) show the defined coverage of the Pacific and Atlantic oceans, respectively.

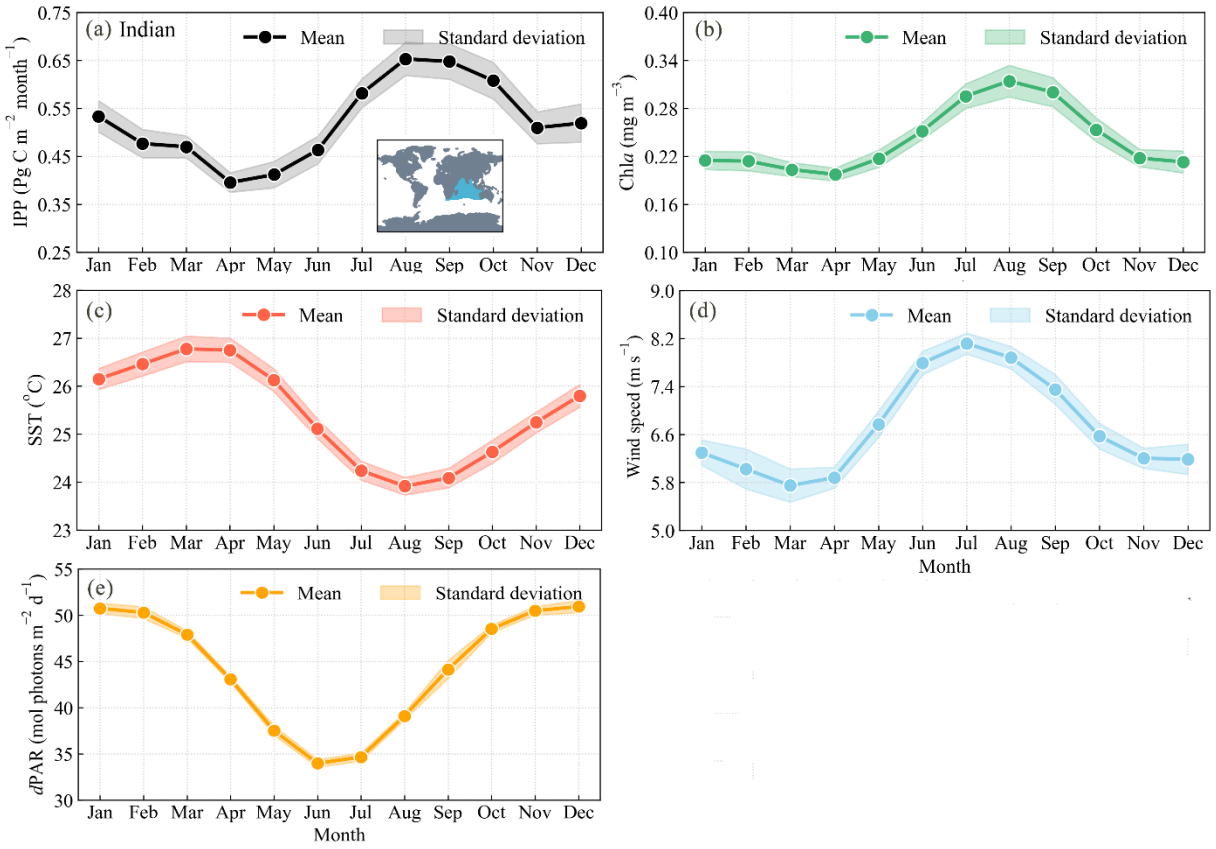

Figure 5.10. Monthly changes of the TPM$_{ERFR}$-based IPP (a), Chl*a* (b), SST (c), wind speed (d), and *d*PAR (e) in the Indian Ocean in the period 1998−2021. Blue area in the inset figures of (a) shows the defined coverage of the Indian Ocean.

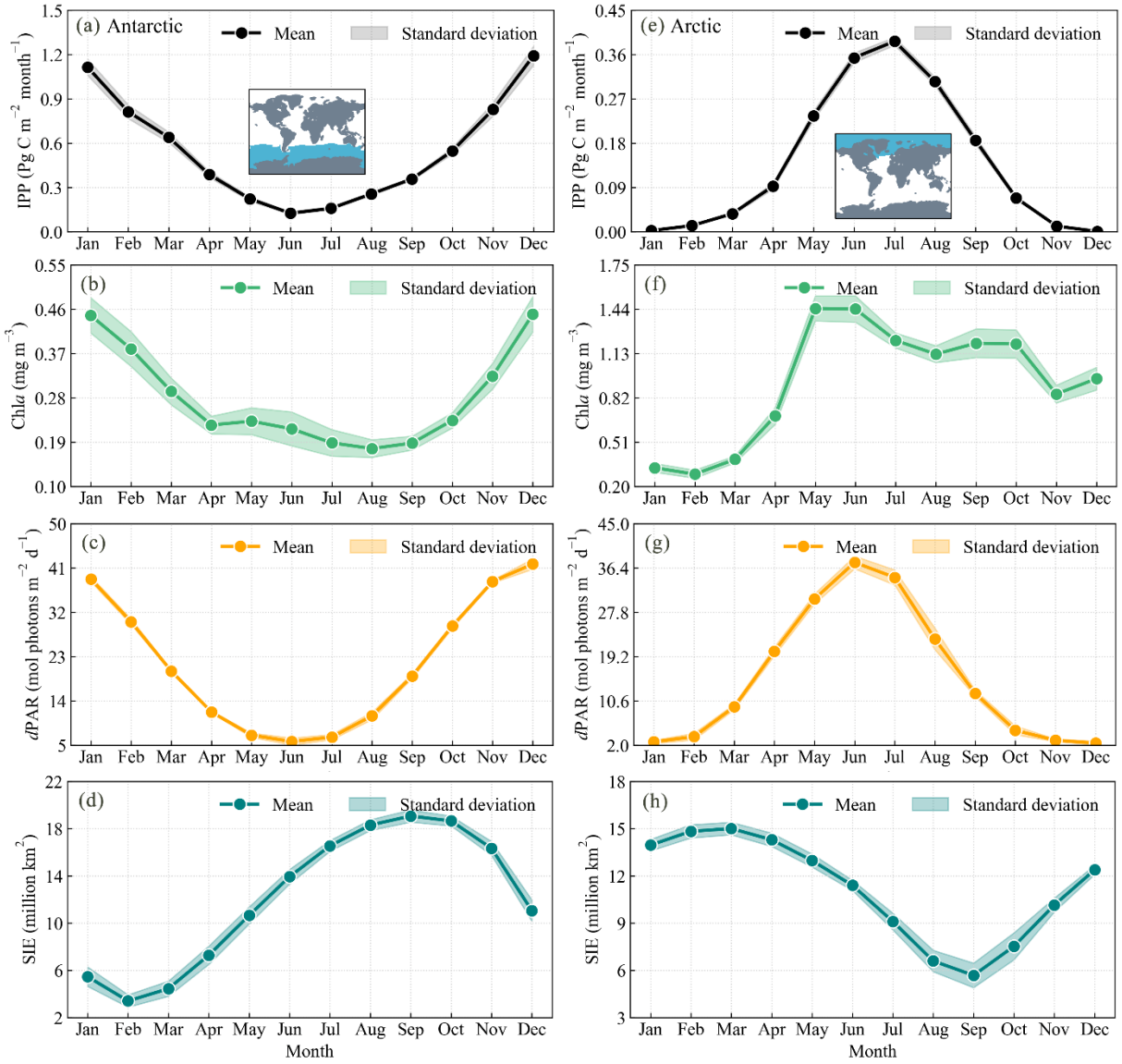

Figure 5.11. Monthly changes of the TPM$_{ERFR}$-based IPP (a and e), Chl*a* (b and f), *d*PAR (c and g), and SIE (d and h) in the Antarctic (right panel) and Arctic oceans (left panel), respectively, in the period 1998−2021. Blue areas in the inset figures of (a) and (e) show the defined coverage of the Antarctic and Arctic oceans, respectively.

Strong seasonality of IPP can be observed in the Indian Ocean as shown in Figure 5.10, with two peaks occurred in boreal winter and summer, which has been found in several previous studies (Antoine et al., 1996; Smyth et al., 2005; Kulk et al., 2020). This distinctive pattern is known to be triggered by seasonal upwelling system in the Arabian Sea, the coastal area in the northern Indian Ocean (see Figure 5.8), due to the local monsoonal wind forcing (Dalpadado et al., 2021). During the summer period (Jun.–Sep.), the western Indian Ocean is dominant by the southwest monsoon (wind speed > 7.0 m s^{-1}, Figure 5.10d), which will induce strong vertical mixing in the Arabian Sea and thereby cause the cooling of sea surface (SST < 25°C, Figure 5.10c) and local bloom (Chl*a* > 0.24 mg m^{-3}, Figure 5.10b). As a result, the Indian Ocean has its largest peak of IPP in boreal summer. During the winter period (Dec.–Feb.), the northeast monsoon becomes dominant in the northern Arabian Sea, bringing cool dry air from the Asian continent, which will also intensify the convection process of local waters due to the cooling effect (Madhupratap et al., 1996). Though the winter monsoon is not as strong as the summer monsoon (wind speed < 6.6 m s^{-1}), it is still enough to sustain fairly high IPP in the northern Arabian Sea, leading to the second round of local bloom and thus the second peak of IPP (but smaller one) in winter. In this case, the seasonal change of *d*PAR is very unlikely a key driver of IPP for the Indian Ocean (Figure 5.10e).

Compared with the other three ocean basins, the Antarctic and Arctic oceans have the strongest seasonal variation as shown in Figure 5.11. Although the

monthly change of SIE exhibit marked seasonal cycle in these two oceans, the seasonal variation is clearly induced by the cycle of polar night and polar day. The polar night period is typically between Nov. and Jan. for the Northern hemishpere and between May and Jul. for the Southern hemishpere. As a result, both IPP and Chl*a* show a similar pattern that is in agreement with *d*PAR in the Antarctic and Arctic oceans, corresponding to the negligible value of IPP during the polar night period for each hemishpere.

5.3. Interannual variability of IPP

5.3.1. Trend analysis of IPP over the last 24 years

The 24-year time series of the anomalies in global annual IPP from 1998 to 2021 based on different models/products were shown in Figure 5.12, where abnormal spikes in the $CAFE_{OP}$- and $AbPM_{GC}$-based IPP occurred in periods 2001–2005 and 2000–2002, respectively. It is most likely because of the artificial issues during the implementation of IPP models, such as quality-control or bias-correction for multi-source input satellite data, especially for the transition period from SeaWiFS to MODIS. Based on the same model but different input products, the $VGPM_{OP}$ and $VGPM_{BF97}$ exhibited different interannual trends, suggesting the importance of climatological satellite records of the required environmental parameters on IPP modeling. In contrast, based on the OC-CCI products, the interannual trends of global annual IPP were consistent for the $VGPM_{BF97}$ and TPM_{ERFR}. Note that only trend signal is considered here but not absolute value.

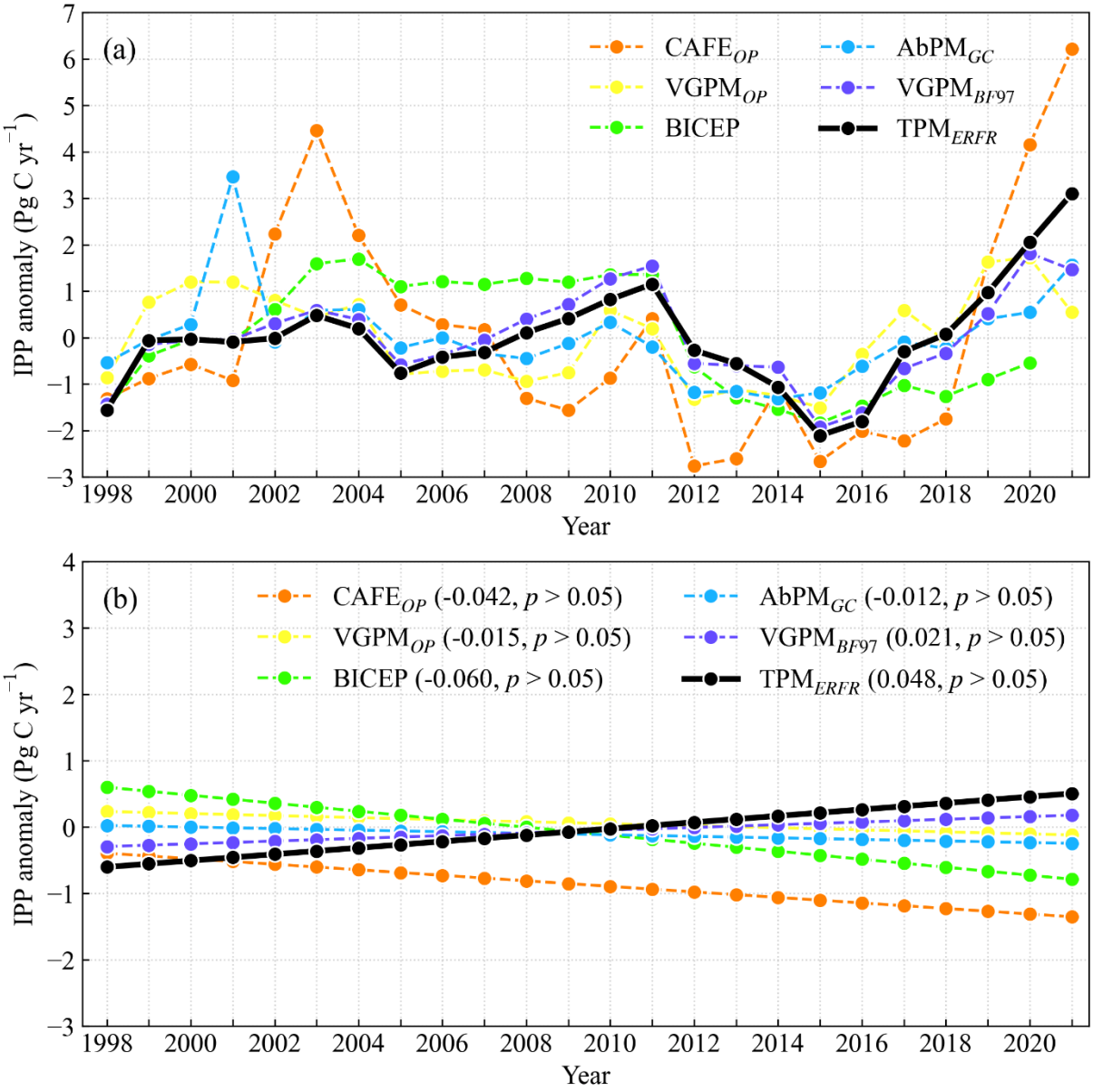

Figure 5.12. Time series of IPP anomalies (a) from different products and the corresponding monotonic linear trends detected by the MKT (b).

According to the TPM_{ERFR}-based time series, several distinct increasing and decreasing periods of global annual IPP can be found. For instance, there was a modest increase of ~2.0 Pg C in period 2005–2011, followed by a decrease of ~3.0 Pg C to 2015. Similar decreasing trends from 2011 to 2015 also appeared in other time series, except for the $CAFE_{OP}$. In recent years, from 2015 to 2021, a dramatic increase of ~5.0 Pg C was revealed by the TPM_{ERFR}, and other time series also showed varying degrees of increasing patterns since 2015. In despite of the certain increasing or decreasing periods, the $VGPM_{BF97}$ and TPM_{ERFR} revealed overall increasing trends of 0.02 Pg C yr^{-1} and 0.05 Pg C yr^{-1} in global annual IPP for the last 24 years, respectively (Figure 5.12b). Whereas other products showed overall decreasing trends. However, all these 24-year trends of global annual IPP are not significant ($p > 0.05$), which means that current satellite record of ocean IPP may not be long enough to sustain a significant climatological trend (Gulev et al., 2021; Pinkerton et al., 2021).

Furthermore, the 24-year time series of annual IPP in individual ocean basins have variable trend signals as summarized in Table 5.2. Compared with the annual IPP trends that are not significant at global scale, the Atlantic and Pacific oceans had experienced an increase of 0.02 Pg C yr^{-1} revealed by the TPM_{ERFR} and a decrease of 0.04 Pg C yr^{-1} revealed by the $VGPM_{OP}$, respectively. There was also a decrease trend of 0.01–0.03 Pg C yr^{-1} found in the Indian Ocean. In contrast, the annual IPP in the Antarctic and Arctic oceans showed overall increase trends of 0.02–0.04 Pg C yr^{-1} and ~0.01 Pg C yr^{-1}, respectively.

Table 5.2. Monotonic trends of time series of annual IPP (Pg C yr^{-1}) and IPP trend components (Tg C $month^{-1}$, converted from Pg C $month^{-1}$ because the trends can be very small) from different products for different oceanic regions over the last 24 years (1998–2021). The trend that is not significant at 95% confidence level ($p > 0.05$) is omitted (denoted as "–"). In particular, the trends detected between 2015 and 2021 for time series of annual IPP and between Dec. 2015 and Dec. 2021 for time series of IPP trend components are shown in parentheses to represent the recent changes of IPP.

Time series	Region	Monotonic trends for different products					
		$CAFE_{OP}$	$VGPM_{OP}$	BICEP	$AbPM_{GC}$	$VGPM_{BF97}$	TPM_{ERFR}
Annual IPP (Pg C yr^{-1})	Global	–	–	–	–	–	–
		(1.54)	(–)	(0.24)	(0.40)	(0.62)	(0.90)
	Atlantic	–	–	–	–	–	0.02
		(–)	(–)	(–)	(0.07)	(0.05)	(0.08)
	Pacific	–	−0.04	–	–	–	–
		(0.62)	(0.24)	(0.13)	(0.16)	(0.39)	(0.42)
	Indian	–	−0.02	−0.03	–	−0.01	–
		(0.23)	(–)	(–)	(–)	(0.13)	(0.19)
	Antarctic	0.04	0.03	–	–	0.04	0.02
		(0.31)	(–)	(–)	(–)	(0.19)	(0.17)
	Arctic	0.01	0.01	–	–	0.01	0.00
		(0.07)	(–)	(–)	(–)	(–)	(0.01)
IPP trend component (Tg C $month^{-1}$)	Global	−0.51	−0.23	−0.53	−0.21	–	0.14
		(11.25)	(2.37)	(1.36)	(2.31)	(5.32)	(6.40)
	Atlantic	−0.27	–	−0.09	–	–	0.10
		(2.54)	(−0.47)	(−0.31)	(0.30)	(0.36)	(0.63)
	Pacific	−0.28	−0.37	−0.33	−0.17	−0.27	−0.15
		(4.55)	(1.94)	(0.89)	(1.03)	(2.93)	(3.40)
	Indian	−0.20	−0.18	−0.20	−0.10	−0.10	–
		(1.64)	(–)	(–)	(0.28)	(0.70)	(0.85)
	Antarctic	0.23	0.24	–	–	0.31	0.16
		(2.14)	(0.72)	(0.32)	(0.15)	(0.89)	(1.09)
	Arctic	0.06	0.09	0.00	–	0.06	0.03
		(0.41)	(–)	(–)	(0.03)	(–)	(0.07)

In order to demonstrate the recent changes of annual IPP, Table 5.2 also presents the trends detected between 2015 and 2021. As expected, all time series of annual IPP in the recent period exhibited significant increasing trends at global scale, except for the $VGPM_{OP}$. Similarly, the 7-year time series of annual IPP in the Atlantic (0.05–0.08 Pg C yr^{-1}), Pacific (0.13–0.62 Pg C yr^{-1}), Indian (0.13–0.62 Pg C yr^{-1}), Antarctic (0.02–0.04 Pg C yr^{-1}), and Actic (0.01–0.07 Pg C yr^{-1}) oceans had varying degrees of increasing trends. However, although the revealed trends based on different IPP models/products had the same directions, their magnitudes could often be different, and thus much caution should be exercised when performing a quantitative analysis.

5.3.2. Mechanism analysis of interannual IPP changes

The trend components of IPP were extracted from the monthly data between 1998 and 2021 based on TSD for interpreting more subtle interannual changes of IPP and underlying mechanisms at global and basin scales. Time series of anomalies in IPP trend components of the global, Pacific, and Indian oceans generally generally had a similar interannual pattern (Figure 5.13). Behrenfeld et al. (2006) have stated that the change of global IPP is largely dependent upon the changes in expansive low-latitude oceans that are characterized by permanent stratification, in accordance with a total contribution of over 55% to global IPP for the Pacific and Indian oceans (Figure 5.4). As a result, the apparent trends of IPP in these two oceans will be projected into the time series of global IPP, such as the turning point around the end of 2015.

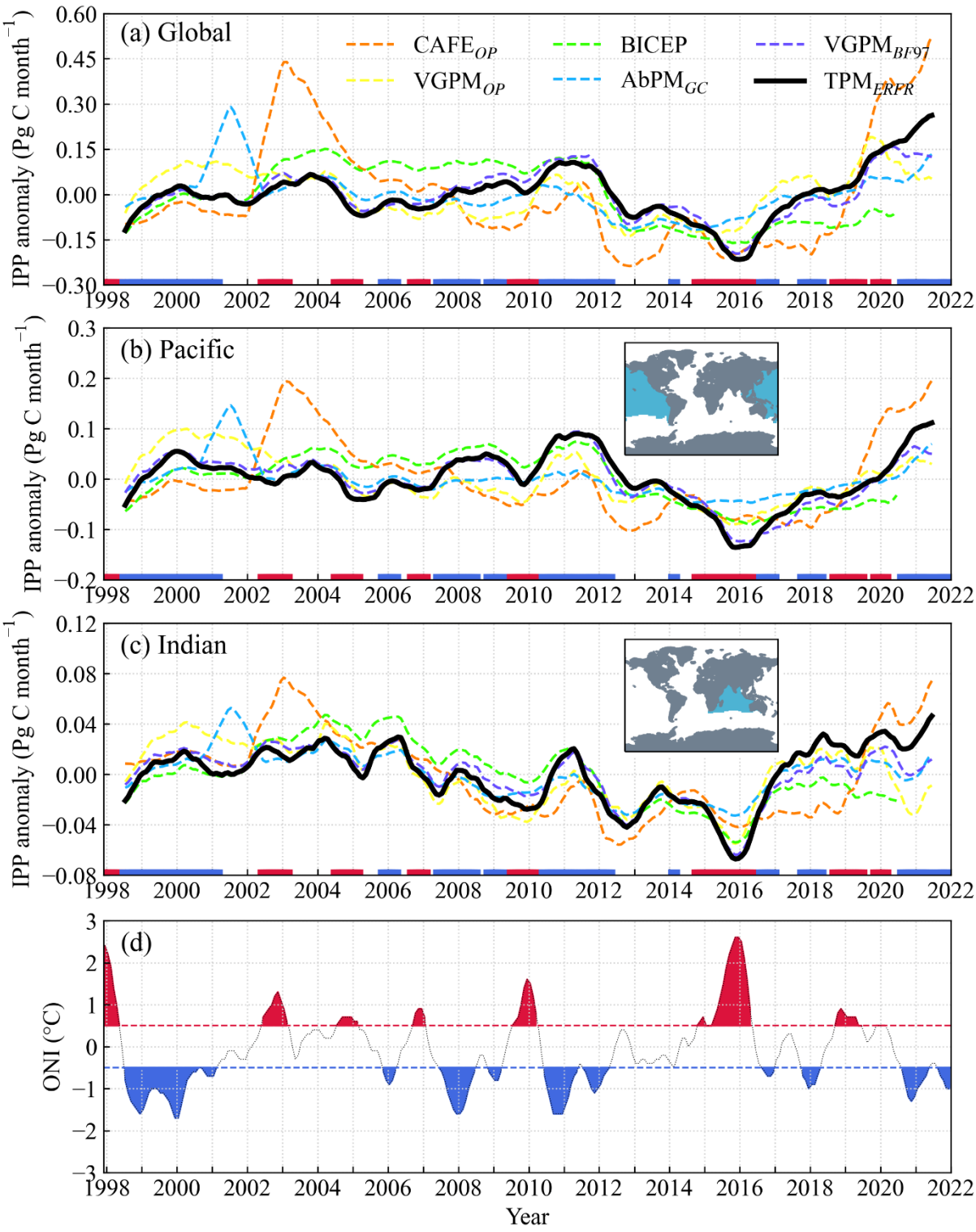

Figure 5.13. Times series of anomalies in IPP trend components from different products for the global (a), Pacific (b), and Indian (b) oceans. Blue areas in the inset figures of (b) and (c) show the defined coverage of the Pacific and Indian oceans, respectively. The time periods of El Niño and La Niña are shaded in red and blue along the x-axis, respectively, according to the time series of ONI that is shown in (d).

Unlike the trends in time series of annual IPP, there was a significant increase trend of IPP (0.14 Tg C month^{-1}, Table 5.2) in the global ocean revealed by the TPM$_{ERFR}$-based time series of IPP trend component for the last 24 years. Moreover, based on other products but except for the VGPM$_{BF97}$, there were significant decreasing trends (0.21–0.53 Tg C month^{-1}) observed in global IPP. While both the Pacific and Indian oceans suffered significant declines of 0.15–0.37 Tg C month^{-1} and 0.10–0.20 Tg C month^{-1}, repsectively, as derived by all products apart from the TPM$_{ERFR}$ for the Indian Ocean. The overall trend of IPP for over two decades is important to quantify the response of global marine ecosystems to climate change, it is nevertheless necessary to analyze the detailed changes of IPP at a smaller temporal scale that eventually affect the overall trend.

Several studies have already confirmed the connection between ocean biological conditions and climate-driven ocean-atmospheric interactions (Di Lorenzo et al., 2008; Martinez et al., 2009; Taboada et al., 2019), such as the effects of ENSO transition on interannual variabilities of global ocean Chl*a* (Vantrepotte and Mélin, 2011) and IPP (Chavez et al., 2011; Kulk et al., 2020). In addition, the mechanisms hidden behind the interannual changes of global IPP associated with climatic events have been discussed. By using the 24-year time series of IPP newly generated based on the TPM$_{ERFR}$, we intend to verify the existing hypotheses and to offer a new piece to the unsovled puzzle in interannual variability of IPP. For the Pacific and Indian oceans, the ENSO-driven interannual changes of IPP are investigated.

According to the time series of ONI (Figure 5.13d), a total of seven El Niño events happened bewteen 1998 and 2021, and two of them were unusually strong, which happened in the periods 1997–1998 and 2014–2016, with the ONI larger than 2°C. On the other hand, a total of six La Niña events can be found in the last 24 years. Combining the time series of IPP trend component derived by the TPM_{ERFR} and the ONI (Figure 5.13b–d), it is noteworthy that El Niño period typically corresponds to a dcrease or local minimum of IPP, such as the low IPP at the beginning of the time series in 1998 and the even lower IPP at the end of 2015 (corresponding to the largest ONI in that El Niño period). On the contrary, La Niña period corresponds to an increase or local maximum of IPP, such as the IPP peaks at the end of 1999 and 2011. As expected, the correlation coefficients (r) between the TPM_{ERFR}-based IPP trend component and the ONI were –0.60, –0.32, and –0.38 ($p < 0.001$) for the Pacific, Indian, and global oceans, respectively (Table 5.3). Other IPP products generally yielded consistent signs of r values, except for the $CAFE_{OP}$. The underlying mechanism is that abnormally warmer ocean surface during El Niño period enhances water stratification and thus suppress the nutrient supplies from deep layer to the surface, leading to a decline of IPP. Conversely, abnormally cooler ocean surface during La Niña period intensifies vertical mixing and thereby draws nutrient-rich waters to the surface layer, casuing a rise of IPP. As such, the transition from a strong El Niño to La Niña induced a sharp rebound of IPP since Dec. 2015, followed by overall elevated IPP in recent 7 years (Table 5.2) due to the dominant features of La Niña.

Table 5.3. Summary of correlation coefficients between IPP trend components from different products and the climate-related (ONI and AMO) and environmental (SIE) variables for different oceanic regions. The IPP trend components in the Antarctic and Arctic oceans are only correlated with the SIE trend components of the Northern and Southern hemispheres, respectively, and in addition, the results for the Polar biome of the Antarctic (poleward of 50°S) are shown. The correlation coefficient that is not significant at 95% confidence level ($p > 0.05$) is omitted (denoted as "–"), and the one that is significant at 99.9% confidence level ($p < 0.001$) is highlighted in bold.

Variable	Region	Correlation coefficients with IPP trend components					
		CAFE$_{OP}$	VGPM$_{OP}$	BICEP	AbPM$_{GC}$	VGPM$_{BF97}$	TPM$_{ERFR}$
ONI	Global	–	**–0.24**	–0.19	–0.14	**–0.34**	**–0.38**
	Atlantic	–	–	–	–	–	–
	Pacific	–	**–0.49**	**–0.31**	**–0.26**	**–0.56**	**–0.60**
	Indian	–	**–0.23**	**–0.21**	–0.13	**–0.36**	**–0.32**
AMO	Global	–	–	–	–	–	–
	Atlantic	–	–	–	–	–0.16	–
	Pacific	–	**–0.20**	–0.13	–0.16	**–0.21**	–0.18
	Indian	–	–	–	–	–	–
SIE trend component	Antarctic	**–0.29**	**–0.38**	–0.19	**–0.56**	–0.18	**–0.24**
	Polar biome of the Antarctic	**–0.37**	**–0.50**	**–0.54**	**–0.58**	**–0.34**	**–0.36**
	Arctic	**–0.64**	**–0.75**	**–0.36**	**–0.23**	**–0.86**	**–0.86**

Although the time series of IPP trend component of the Indian Ocean was generally consistent with that of the Pacific Ocean (Figure 5.13b and Figure 5.13c), in theory, the influence of El Niño and La Niña events on interannual variability of IPP is more straightforward in the Pacific Ocean than in the Indian Ocean, resulting in a lower *r* between IPP trend component and the ONI in the Indian Ocean. In practice, the parallel change of SST in the Indian Ocean, as a response to ENSO transition, is mostly regulated by atmospheric teleconnection (such as the changes of cloud coverage and wind stress induced by El Niño or La Niña events) (Kug et al., 2004; Yu et al., 2021). It can be an explanation for the subtle differences in the time series of these two ocean basins. In conclusion, the inverse correlation between climate-driven SST variations and IPP changes in the Pacific, Indian, and global oceans as a reflection of the influence of ENSO transition on ocean biology, which has been illustrated earlier (Behrenfeld et al., 2006; Martinez et al., 2009), is corroborated by the TPM_{ERFR}-based 24-year time series of IPP.

As tabulated in Table 5.2, an overall increase of 0.10 Tg C $month^{-1}$ in the Atlantic Ocean from 1998 to 2021 was revealed by the TPM_{ERFR}, which could be more confident than the trends detected in the $CAFE_{OP}$ and BICEP-based time series because of our assessments for IPP products (Section 5.1) and the irregular patterns in the $CAFE_{OP}$ data (Figure 5.14). In the period between Dec. 2015 and Dec. 2021, the IPP in the Atlantic Ocean also showed a significant increasing trend of 0.63 Tg C $month^{-1}$ as derived by the TPM_{ERFR}, which was similar to the

trends in the Pacific and Indian oceans. However, the difference is that, as shown in Figure 5.14a, this overall increase of IPP in the Atlantic Ocean could date back to end of 2013 rather than 2015. It suggests that ENSO transition is not an apparent factor of the interannual variability of IPP in the Atlantic Ocean, with no significant correlation with time series of the ONI (Table 5.3). Furthermore, the time series of AMO (Figure 5.14c) was introduced as a new clue to interpret the interannual change of IPP in the Atlantic Ocean. Unfortunately, there was no statistically significant correlation between the AMO and IPP trend components for all IPP products (Table 5.3), except for the $VGPM_{BF97}$, which makes the potential link with the AMO less confident for the last 24 years. Interestingly, the IPP in the Pacific Ocean was negatively correalated with the AMO, with significant r values ranging between –0.21 and –0.13 ($p < 0.05$), but this connection was not as strong as the one with the ONI. Martinez et al. (2009) have stated that the AMO is clearly related to the multidecadal changes of global Chl*a* by using the ocean color time series from the CZCS (1979–1983) and SeaWiFS (1998–2002). However, it is not the case in the period 1998–2021, because the AMO can often last for over two decades and it has been a warm phase since the mid-1990s. It means that only with more older satellite data (in the era of CZCS) can we analyze the potenial influence of phase shifting of the AMO on the interannual varibility of IPP in the Atlantic Ocean, which is out of the scope of this study and needs future efforts, especially for the process and bias-correction of the CZCS products.

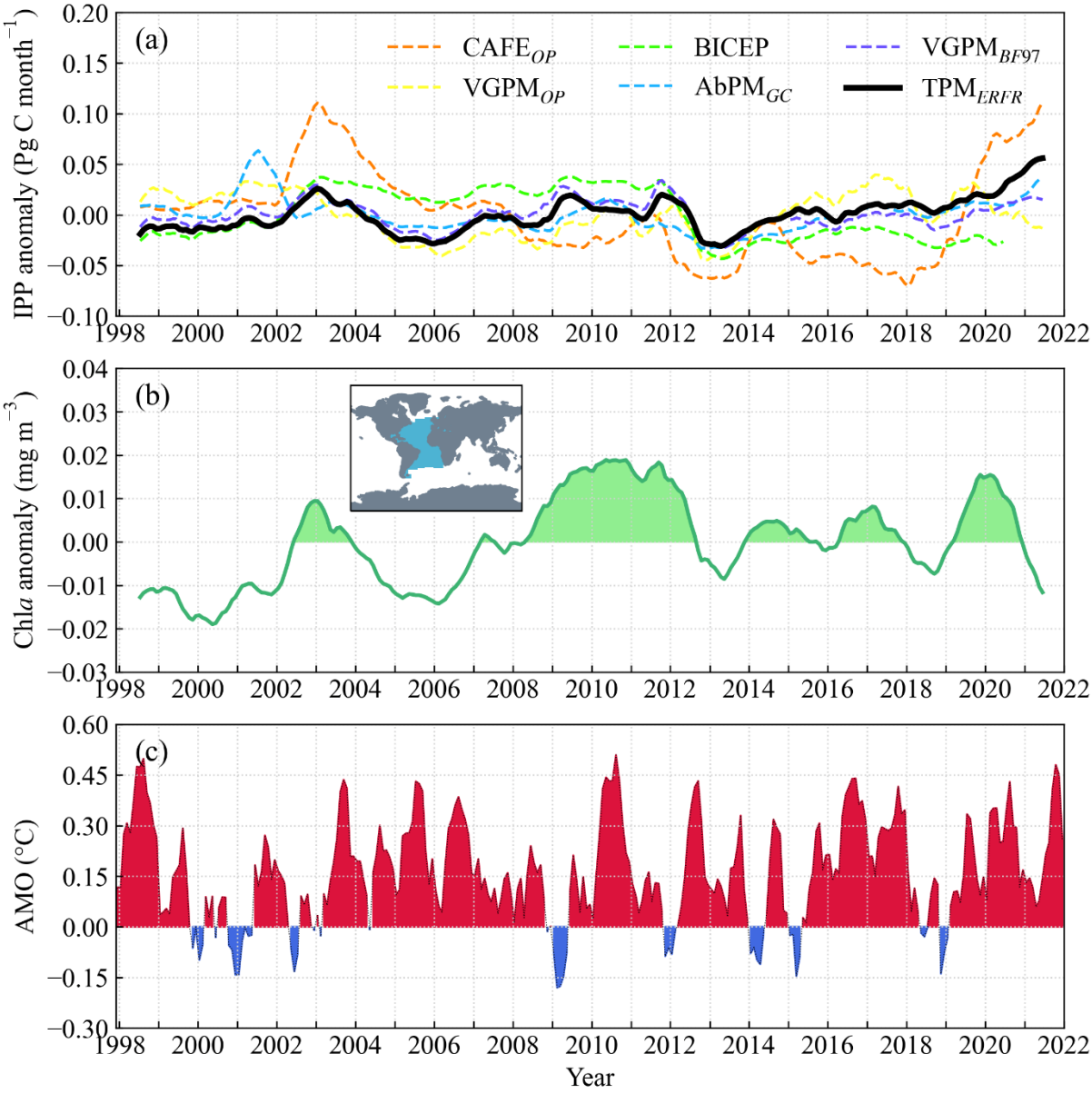

Figure 5.14. Times series of anomalies in IPP trend components (a) from different products along with anomalies in Chl*a* trend component (b) for the Atlantic Ocean. Blue area in the inset figure of (b) shows the defined coverage of the Atlantic Ocean. Time series of the AMO is also presented (c).

Despite the fact that the climate variabilities of AMO and ENSO were not apparently related to the interannual change of IPP in the Atlanci Ocean for the last 24 years, the temporal changes of Chl*a* may partially contributed to the variations in IPP. As shown in Figure 5.14b, the time series of IPP trend component showed several features that were in common with those of Chl*a* time series. For example, both IPP and Chl*a* tended to increase between 2002 and 2003, and then started to reduce until 2006, followed by an overall increasing trend but with modest fluctuations to the end of 2011. After a decline observed in 2012, the Chl*a* in the Atlantic Ocean had experienced obvious fluctuations with the anomalies biasd in two directions. It is clearly inconsistent with the overall increase of IPP since 2013. Hence, the mechanisms of the environment- and climate-related interannual variability of IPP in the Atlantic Ocean in the period 1998–2021 remain to be illustrated in the future work by considering more data and also by performing region-scale analyses instead of whole basin.

The 24-year time series of IPP trend components from the TPM_{ERFR} in the Antarctic Ocean exhibited a significant increasing trend of 0.16 Tg C month^{-1}, with a raising rate of 1.09 Tg C month^{-1} from Dec. 2015 to Dec. 2021 (Table 5.2). Other IPP products also indicated the elevation of IPP in the Antarctic Ocean with varying rates, except for the BICEP and $AbPM_{GC}$. As for the Arctic ocean, the increasing trends were also detected for the last 24 (0.03 Tg C month^{-1}) and 7 years (0.07 Tg C month^{-1}) based on the TPM_{ERFR}, which could be related to the changes in ambient conditions of phytoplankton community in polar oceans.

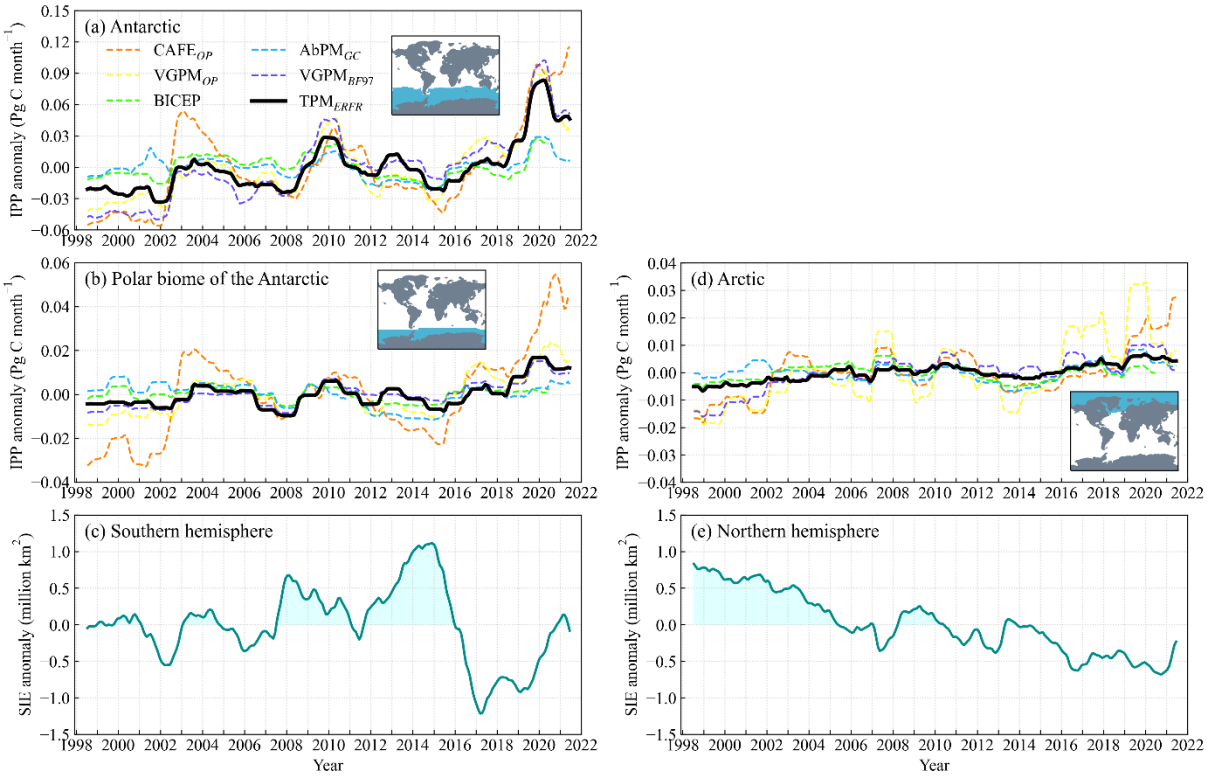

Figure 5.15. Times series of anomalies in IPP trend components from different products for the entire (a) and Polar biome of the Antarctic Ocean (b) as well as the Arctic Ocean (d). Blue areas in the inset figures of (a), (b), and (d) show the defined coverage of the entire and Polar biome of the Antarctic Ocean as well as the Arctic Ocean, respectively. Time series of anomalies in SIE trend components for the Southern (c) and Northern hemispheres (e) are shown in the bottom panel.

Previous studies have stated that the prolonged increase of IPP in polar oceans is greatly attributed to the decline in sea ice cover, which is particularly true in the Arctic Ocean (Arrigo et al., 2008; Lewis et al., 2020; Petrenko et al., 2013). Because there will be a greater living space for phytoplankton community in polar oceans as a result of sea ice reduction and raising area of open waters, which is favorable for phytoplankton photosynthesis and hence the enhancement of carbon fixation. Therefore, the time series of IPP trend components in the Antarctic and Arctic oceans were compared with those of SIE (Figure 5.15) to investigate the interannual responses of IPP to SIE variability. Since the defined area of the Antarctic Ocean covers not only polar zone but also part of southern temperate zone, the results for the Polar biome of the Antarctic Ocean (poleward of 50°S) were also shown (Figure 5.15b).

As anticipated, the increasing and decreasing periods of IPP in the Antarctic Ocean generally corresponds to the reduction and expansion of sea ice, respectively (Figure 5.15a–c). For example, the IPP tended to increase from 2008 to 2010 when the SIE started to decrease, and in particular, there was a local minimum of IPP occurred around 2015 corresponding to the maximum of SIE anomalies. For all products, as shown in Table 5.3, time series of IPP trend components showed a negative correlation with SIE ($r = –0.24$, $p < 0.001$, derived by the TPM_{ERFR}) in the Antarctic Ocean, while this negative correlation was closer for the IPP in the Polar biome ($r = –0.36$, $p < 0.001$, derived by the TPM_{ERFR}). These results suggest that the SIE variability only partially contributes to the

interannual changes of IPP in the Antarctic Ocean, because this scenario of mechanism could not explain every turning points of IPP and SIE, such as the parallel increase in IPP and SIE from 2019 to 2020.

Compared with the strong fluctuation of SIE in the Southern hemishpere, the anomalies in SIE for the Arctic Ocean suggests an obvious decline of sea ice cover in the Northern hemishpere for the last 24 years (Figure 5.15e). This reduction of the Arctic sea ice well corresponds to the prolonged increase in the Arctic IPP (Figure 5.15d, Table 5.2), with a significant r of –0.86 ($p < 0.001$) dervied by the TPM_{ERFR} (Table 5.3). Therefore, the hypothesis raised in previous studies is corroborated in both theoretical and statistical ways. Nevertheless, to further decipher the interannual variability of IPP in the Antarctic and Arctic oceans, future works should be dedicated to illustrate the physical processes in oceanic and atmospheric environments induced by climate change and the related biological responses of phytoplankton community.

Chapter 6. Conclusions

Phytoplankton serve as the base of aquatic food web, and thereby support the heterotroph community across the globe. Monitoring phytoplankton primary production (PP) via remote sensing is vital to understanding changes in biological conditions of global aquatic ecosystems, and subsequently responses of marine environments to climate change. In this study, we have proposed a novel and generalized model to improve satellite-based estimation of PP in various types of water bodies, and produced refined satellite records of global PP for the last 24 years from 1998 to 2021. We demonstrated the advantages of the proposed model in seamlessly monitoring global PP, and discussed the characteristics and causes of spatiotemporal variabilities in basin-scale and global PP. The principal results and conclusions are summarized as follows.

The improvement of remote estimation of phytoplankton daily, depth-integrated primary production (IPP) in clear to turbid waters has been achieved in this study by means of the integration of machine learning algorithm (ERFR) and photophysiological model (TPM). Based on a large *in situ* dataset acquired from several representative aquatic regimes that spanned a broad range of trophic conditions, the ERFR trained with four input parameters (T, Chl$a$, K_d(PAR), and r_{PAR}) was confirmed to be optimal. This optimal training scenario corresponds to the fact that phytoplankton growth through photosynthesis is subject to both biotic and abiotic factors, and the inclusion of depth-dependent optical characteristics (i.e., r_{PAR}) of water enables better estimation of vertical distribution of P^{B}_{max} and

E_k. Consequently, the ERFR significantly improved the remote assignment of P^B_{max} and E_k in oceanic and inland waters compared with existing P-E algorithms. The enhanced estimations of P^B_{max} and E_k in optically complex turbid (Case-II) waters was demonstrated in the Rhode River and Lake Kasumigaura, especially in the latter case as a completely independent validation. As a combination of conventional MLR and RFR, the ERFR preliminarily exhibited an improved extrapolation capability, implying its potential applications in other independent inland waters. Given the advantages of the ERFR and TPM, the TPM_{ERFR} was proved to have the best performance ($RMSD \leq 0.25$) compared with other benchmark IPP models (*RMSD*s ranged between 0.27 and 0.62) in the two turbid water areas and the clear water areas tested in this study (mainly the Atlantic Ocean and the Australian sector of the Antarctic Ocean). The major uncertainty of the TPM_{ERFR} may come from the satellite-measured Chl*a* and K_d(PAR) as required model input, especially for optical complex waters, and the remote retrieval of vertical distribution of Chl*a*, especially for the stratified open oceans. Future validation and improvement of the TPM_{ERFR} can be facilitated through the collection and use of high-quality field data from different water bodies with additional environmental parameters of interest based on a unified protocol. Observations from advanced instruments, such as detailed spectral variations of light penetration obtained by hyperspectral sensors and vertical distributions of Chl*a* measured by the BGC-Argo floats or ocean Lidar, may enable better three-dimensional retrieval of IPP based on the TPM_{ERFR}.

The second achievement of this study is the generation of gap-free global IPP product using the 24-year (1998–2021) multi-source satellite data based on the TPM_{ERFR}. The utility of merged remote sensing observations from multiple sensors promotes the reconstruction of climate data records for environmental parameters. The 24-year ocean color products from the OC-CCI (v5.0) project provided high-quality merged global data of Chl*a* and K_d(PAR), which could be directly and safely used in the TPM_{ERFR}. Whereas potential systematic biases might exist in the multi-source products of *d*PAR (SeaWiFS/MODIS) and SST (SST-CCI/MODIS) for the transition periods of data sources. To reduce the uncertainty of input satellite products that would be propagated into the IPP estimation, the quantitative analysis was performed to investigate the potential biases in different sets of *d*PAR and SST data in the overlapping time periods. The analysis used linear regression to quantify the bias between SeaWiFS and MODIS monthly *d*PAR data for all available pixels in each month. The fitted linear slopes and intercepts indicated that the monthly *d*PAR of SeaWiFS is generally larger that of MODIS, and thus the monthly mean slopes and intercepts were used to correct the bias of SeaWiFS *d*PAR. This bias-correction approach is convenient and effective to generate refined and continuous time series of *d*PAR from SeaWiFS and MODIS. Similar analysis was repeated for the multi-source SST data from the SST-CCI and MODIS, and no apparent bias was identified in these two sets of SST data as indicated by small intercepts and slopes close to 1.0. The time series of monthly SST of SST-CCI was consistent with that of MODIS

in the overlapping time period. Hence, bias-correction for the multi-source SST data was deemed to be unnecessary. As such, the uncertainty from model input could be suppressed by the procedure of bias-correction. Furthermore, as the common issues like the failure of atmospheric correction and the contamination of cloud or cloud shadow, ocean color products often have a certain proportion of missing data. To address this problem, a frequently-used gap-filling approach, namely the DINEOF, was adopted to fill in the missing data in the generated IPP products. The performance has proved to be satisfactory based on a case study in the marginal seas of China and Japan, as indicated by the *RMSD* between known and reconstructed IPP values less than 0.09 and high R^2 of 0.79–0.96. By the implementation of the DINEOF on the generated 24-year IPP product, the global data coverage for all water pixels has been successfully elevated from ~80–98% to ~90–100%, without noticeable artificial patches in IPP distributions after gap-filling. Consequently, estimates of total IPP at different spatial and temporal scales based on the reconstructed gap-free global IPP product could be of higher fidelity.

This study has then proceeded to comprehensively analyze detailed features in spatiotemporal variability of IPP and underlying mechanisms associated with climate-related and environmental drivers. Additional IPP products generated via different models or satellite products were introduced for the inter-comparisons against our results. Global distributions of climatological annual IPP on basis of different IPP products showed similar spatial patterns and several apparent oceanographic features but with a variety of magnitudes. Typically, ocean gyre

regimes are permanently dominated by low levels of IPP, whereas equatorial and coastal upwelling zones together with coastal and inland shallow water areas are usually fertilized by local nutrient-rich waters and hence can be more productive than most of open oceans. Percent contributions of ocean basins and biomes to global IPP for individual models/products were systematically quantified. In spite of the discrepancy in total amounts of IPP derived by different products, most of these products yielded relatively consistent contributions in each ocean basin and biome. The rationality of the TPM_{ERFR}-based IPP distributions at basin and global scales was first qualitatively assessed through the inter-comparisons with other products and the results of model evaluation reported in previous studies in addition with those presented here. More realistic distributions of IPP were revealed by the TPM_{ERFR}, especially for the water environments of polar, coastal, and inland areas. Moreover, basin-scale and global estimates of annual IPP derived by differet models/products were assessed by exhausive literature reviews for over the past two decades. The TPM_{ERFR} gave an estimate of global climatological annual IPP of 43.7 Pg C yr^{-1} in the period 1998–2021, and gave basin-scale estimates of 18.4 (42.1%), 10.1 (23.1%), 6.3 (14.4%), 6.7 (15.3%), and 1.7 Pg C yr^{-1} (3.9%) for the Pacific, Atlantic, Indian, Antarctic, and Arctic oceans, respectively, along with an estimate of 0.7 Pg C yr^{-1} (1.6%) for inland water areas. The assessment suggested that the TPM_{ERFR} has the potential to provide reasonable estimates of annaul IPP, especially for the Atlantic, Antarctic, and Arctic oceans. Nevertheless, further works of validation and improvement are

desired under the support of expanded *in situ* data collections.

Seasonal variations of IPP in the Antarctic and Arctic oceans are the strongest compared with other ocean basins, because of the cycle of polar night and polar day that regulates the distribution of solar energy available for phytoplankton community in polar oceans. Seasonal changes of IPP in the Indian Ocean are characterized by local seasonal upwelling system in the Arabian Sea, where strong summer monsoon will intensify water stratification and thus enrich the nutrient concentration in the upper mixed layer, leading to phytoplankton bloom. During the period of local winter monsoon, the surface cooling effect, though it is not as strong as the wind-driven effect in summer period, will also induce vertical exchanege of waters that favors elevated IPP. In contrast, there are no significant seasonl signals in the Pacific and Atlantic oceans, which mainly because that most of these two ocean basins are characterized by stable oceanographic features, such as low and high levels of IPP in ocean gyres and upwelling systems, respectively. For the last 24 years, based on the TPM_{ERFR}, no significant trends of annual IPP were detected in the global, Pacific, and Indian oceans, while the annual IPP data in the Atlantic, Antarctic, and Arctic oceans exhibited significant increasing trends. In view of the recent 7 years between 2015 and 2021, both basin-scale and global annual IPP data have experienced an apparent increasing trend. Time series of IPP trend components were extracted from monthly products based on TSD to facilitate the interperation of subtle interannual changes of IPP. It was found that the interannual variabilities of IPP in the Pacific and Indian oceans is obviously

attributed to ENSO transition, the patterns of which will be projected into the changes of global IPP. El Niño (La Niña) events accompanied with abnormally warmer (cooler) ocean surface will weaken (enhance) water stratification, resulting in less (more) nutrient supplies in the surface layer, and consequently suppress (promote) phytoplankton grow. This mechanism can be a plausible explanation for the dramatic rise of IPP in global, Pacific, and Indian oceans since 2015, because there was a period of transition from a strong El Niño to La Niña starting at the end of 2015. The correlation coefficients (r) between IPP trend components and ONI were –0.38, –0.60, and –0.32 ($p < 0.001$) in the global, Pacific, and Indian oceans, which could statistically sustain this mechanism. On the other hand, the interannual changes of IPP in the Atlantic Ocean has proved to be uncorrelated with ENSO and AMO, with no significant values of r, but could be partially explained by the changes in Chl*a*. In agreement with the earlier findings, the interannual varibilities of IPP in polar oceans are subject to the changes of sea ice cover to a certain degree, as indicated by significant negative correlations between SIE and IPP trend components in the Antarctic Ocean ($r = –0.24$ and $r = –0.36$ for its Polar biome only, $p < 0.001$) and particularly in the Arctic Ocean ($r = –0.86$, $p < 0.001$). Ultimately, the diversity of interannual variability of IPP over the last 24 years has been partially deciphered by the TPM_{ERFR}-based product, nevertheless, future efforts are expected to unravel more detailed variations in IPP by using the TPM_{ERFR}, which relies on favourable climate records and the extension of satellite observations.

Acknowledgments

The acquisition of cruise datasets in the LEFE-CYBER were funded through Centre national de la recherche scientifique (CNRS)-Institut national des sciences de l'Univers (INSU) grants. The Long-term Environmental Monitoring data in Lake Kasumigaura was provided by the National Institute for Environmental Studies (NIES), Japan. Thanks to ESA for providing access to the CCI, MERIS and Sentinel-2 MSI data, and NASA for SeaWiFS and MODIS data. The financial support provided by the China Scholarship Council (No. 201908320540) was gratefully acknowledged.

I am extremely grateful to my supervisors, Dr. Wei Yang and Dr. Akihiko Kondoh, for their patience and assistance at every stage of my PhD study. I would also like to thank the scientists: Dr. Marcel Babin, Dr. Bangqin Huang, Dr. Yuyuan Xie, Dr. Julia Uitz, Dr. Gemma Kulk, and Dr. Shubha Sathyendranath, for their kind help and insightful comments that have made this research better. I would like to thank all the members in the Kondoh & Yang lab for their inspiring ideas. Finally, I would like to express my sincere gratitude to my mother and my girlfriend for their tremendous understanding and encouragement all through my PhD life. I am very happy that I will be finally granted the title of PhD. From the day I met my girlfriend, I was convinced that the greatest prize of my life was her, and nothing will change that. I would love to share my future with her and keep doing better, hoping we can walk through the rest of our life together.

Reference

Alvera-Azcárate, A., Barth, A., Rixen, M., Beckers, J.M., 2005. Reconstruction of incomplete oceanographic data sets using empirical orthogonal functions: Application to the Adriatic Sea surface temperature. Ocean Model. 9, 325–346. https://doi.org/10.1016/j.ocemod.2004.08.001

Antoine, D., André, J.M., Morel, A., 1996. Oceanic primary production: 2. Estimation at global scale from satellite (Coastal Zone Color Scanner) chlorophyll. Global Biogeochem. Cycles 10, 57–69. https://doi.org/10.1029/95GB02832

Antoine, D., Morel, A., 1996. Oceanic primary production: 1. Adaptation of a spectral light-photosynthesis model in view of application to satellite chlorophyll observations. Global Biogeochem. Cycles 10, 43–55. https://doi.org/10.1029/95GB02831

Arrigo, K.R., van Dijken, G.L., Bushinsky, S., 2008. Primary production in the Southern Ocean, 1997-2006. J. Geophys. Res. Ocean. 113, 1997–2006. https://doi.org/10.1029/2007JC004551

Arrigo, K.R., van Dijken, G.L., Bushinsky, S., 2008. Primary production in the Southern Ocean, 1997-2006. J. Geophys. Res. Ocean. 113, 1997–2006. https://doi.org/10.1029/2007JC004551

Arst, H., Nõges, P., Nõges, T., Kauer, T., Arst, G.E., 2012. Quantification of a Primary Production Model Using Two Versions of the Spectral Distribution of the Phytoplankton Absorption Coefficient. Environ. Model. Assess. 17, 431–440. https://doi.org/10.1007/s10666-011-9305-z

Arst, H., Nõges, T., Nõges, P., Paavel, B., 2008. In situ measurements and model calculations of primary production in turbid waters. Aquat. Biol. 3, 19–30. https://doi.org/10.3354/ab00059

Babin, M., Morel, A., Claustre, H., Bricaud, A., Kolber, Z., Falkowski, P.G., 1996. Nitrogen- and irradiance-dependent variations of the maximum quantum yield of carbon fixation in eutrophic, mesotrophic and oligotrophic marine systems. Deep Sea Res. Part I Oceanogr. Res. Pap. 43, 1241–1272.

Bailey, S.W., Werdell, P.J., 2006. A multi-sensor approach for the on-orbit validation of ocean color satellite data products. Remote Sens. Environ. 102, 12–23. https://doi.org/10.1016/j.rse.2006.01.015

Balasubramanian, S. V., Pahlevan, N., Smith, B., Binding, C., Schalles, J., Loisel, H., Gurlin, D., Greb, S., Alikas, K., Randla, M., Bunkei, M., Moses, W., Nguyễn, H., Lehmann, M.K., O'Donnell, D., Ondrusek, M., Han, T.H., Fichot, C.G., Moore, T., Boss, E., 2020. Robust algorithm for estimating total suspended solids (TSS) in inland and nearshore coastal waters. Remote Sens. Environ. 246, 111768. https://doi.org/10.1016/j.rse.2020.111768

Beckers, J.M., Rixen, M., 2003. EOF calculations and data filling from incomplete oceanographic datasets. J. Atmos. Ocean. Technol. 20, 1839–1856. https://doi.org/10.1175/1520-0426(2003)020<1839:ECADFF>2.0.CO;2

Behrenfeld, M.J., Falkowski, P.G., 1997a. A consumer's guide to phytoplankton primary productivity models. Limnol. Oceanogr. 42, 1479–1491. https://doi.org/10.4319/lo.1997.42.7.1479

Behrenfeld, M.J., Falkowski, P.G., 1997b. Photosynthetic rates derived from satellite-based chlorophyll concentration. Limnol. Oceanogr. 42, 1–20. https://doi.org/10.4319/lo.1997.42.1.0001

Behrenfeld, M.J., Hu, Y., Hostetler, C.A., Dall'Olmo, G., Rodier, S.D., Hair, J.W., Trepte, C.R., 2013. Space-based lidar measurements of global ocean carbon stocks. Geophys. Res. Lett. 40, 4355–4360. https://doi.org/10.1002/grl.50816

Behrenfeld, M.J., O'Malley, R.T., Boss, E.S., Westberry, T.K., Graff, J.R., Halsey, K.H., Milligan, A.J., Siegel, D.A., Brown, M.B., 2016. Revaluating ocean warming impacts on global phytoplankton. Nat. Clim. Chang. 6, 323–330. https://doi.org/10.1038/nclimate2838

Behrenfeld, M.J., O'Malley, R.T., Siegel, D.A., McClain, C.R., Sarmiento, J.L., Feldman, G.C., Milligan, A.J., Falkowski, P.G., Letelier, R.M., Boss, E.S., 2006. Climate-driven trends in contemporary ocean productivity. Nature 444, 752–755. https://doi.org/10.1038/nature05317

Behringer, D.W., Xue, Y., 2004. Evaluation of the global ocean data assimilation system at NCEP: The Pacific Ocean, in: Proc. Eighth Symp. on Integrated Observing and Assimilation Systems for Atmosphere, Oceans, and Land Surface.

Bender, M., Grande, K., Johnson, K., Marra, J., Williams, P.J.L.B., Sieburth, J., Pilson, M., Langdon, C., Hitchcock, G., Orchardo, J., Hunt, C., Donaghay, P., Heinemann, K., 1987. A comparison of four methods for determining planktonic community production. Limnol. Oceanogr. 32, 1085–1098. https://doi.org/10.4319/lo.1987.32.5.1085

Bergamino, N., Horion, S., Stenuite, S., Cornet, Y., Loiselle, S., Plisnier, P.D., Descy, J.P., 2010. Spatio-temporal dynamics of phytoplankton and primary production in Lake Tanganyika using a MODIS based bio-optical time series. Remote Sens. Environ. 114, 772–780. https://doi.org/10.1016/j.rse.2009.11.013

Bindoff, N.L., Cheung, W.W.L., Kairo, J.G., Arístegui, J., Guinder, V.A., Hallberg, R., Hilmi, N.J.M., Jiao, N., Karim, M.S., Levin, L., 2019. Changing ocean, marine ecosystems, and dependent communities. IPCC Spec. Rep. Ocean Cryosph. a Chang. Clim. 477–587.

Bouman, H., Platt, T., Sathyendranath, S., Stuart, V., 2005. Dependence of light-saturated photosynthesis on temperature and community structure. Deep. Res. Part I Oceanogr. Res. Pap. 52, 1284–1299. https://doi.org/10.1016/j.dsr.2005.01.008

Bouman, H.A., Jackson, T., Sathyendranath, S., Platt, T., 2020. Vertical structure in chlorophyll profiles: Influence on primary production in the Arctic Ocean: Vertical Structure in Arctic Chlorophyll. Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. 378. https://doi.org/10.1098/rsta.2019.0351

Bouman, H.A., Platt, T., Doblin, M., Figueiras, F.G., Gudmundsson, K., Gudfinnsson, H.G., Huang, B., Hickman, A., Hiscock, M., Jackson, T., Lutz, V.A., Mélin, F., Rey, F., Pepin, P., Segura, V., Tilstone, G.H., Van Dongen-Vogels, V., Sathyendranath, S., 2018. Photosynthesis-irradiance parameters of marine phytoplankton: Synthesis of a global data set. Earth Syst. Sci. Data 10, 251–266. https://doi.org/10.5194/essd-10-251-2018

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324

Brewin, R.J.W., Sathyendranath, S., Platt, T., Bouman, H., Ciavatta, S., Dall'Olmo, G., Dingle, J., Groom, S., Jönsson, B., Kostadinov, T.S., Kulk, G., Laine, M., Martínez-

Vicente, V., Psarra, S., Raitsos, D.E., Richardson, K., Rio, M.H., Rousseaux, C.S., Salisbury, J., Shutler, J.D., Walker, P., 2021. Sensing the ocean biological carbon pump from space: A review of capabilities, concepts, research gaps and future developments. Earth-Science Rev. 217. https://doi.org/10.1016/j.earscirev.2021.103604

Brewin, R.J.W., Tilstone, G.H., Jackson, T., Cain, T., Miller, P.I., Lange, P.K., Misra, A., Airs, R.L., 2017. Modelling size-fractionated primary production in the Atlantic Ocean from remote sensing. Prog. Oceanogr. 158, 130–149. https://doi.org/10.1016/j.pocean.2017.02.002

Bricaud, A., Morel, A., Babin, M., Allali, K., Claustre, H., 1998. Variations of light absorption by suspended particles with chlorophyll a concentration in oceanic (case 1) waters : Analysis and implications for bio-optical models Abstract . Spectral absorption coefficients of total particulate matter mg m-3). As pre. J. Geophys. Res. 103, 31033–31044.

Bricaud, A., Stramski, D., 1990. Spectral absorption coefficients of living phytoplankton and nonalgal biogenous matter: A comparison between the Peru upwelling areaand the Sargasso Sea. Limnol. Oceanogr. 35, 562–582. https://doi.org/10.4319/lo.1990.35.3.0562

Campbell, J.W., Antoine, D., Armstrong, R., Arrigo, K., Balch, W., Barber, R., Behrenfeld, M., Bidigare, R., Bishop, J., Carr, M.E., Esaias, W., Falkowski, P., Hoepffner, N., Iverson, R., Kiefer, D., Lohrenz, S., Marra, J., Morel, A., Ryan, J., Vedernikov, V., Waters, K., Yentsch, C., Yoder, J., 2002. Comparison of algorithms for estimating ocean primary production from surface chlorophyll, temperature, and irradiance. Global Biogeochem. Cycles 16. https://doi.org/10.1029/2001gb001444

Carr, M., Friedrichs, M.A.M., Schmeltz, M., Noguchi, M., Antoine, D., Arrigo, K.R., Asanuma, I., Aumont, O., Barber, R., Behrenfeld, M., Bidigare, R., Buitenhuis, E.T., Campbell, J., Ciotti, A., Dierssen, H., Dowell, M., Dunne, J., Esaias, W., Gentili, B., Gregg, W., Groom, S., Hoepffner, N., Ishizaka, J., Kameda, T., Que, C. Le, Reddy, T.E., Ryan, J., Scardi, M., Moore, K., Smyth, T., Turpie, K., Tilstone, G., Waters, K., Yamanaka, Y., 2006. A comparison of global estimates of marine primary production from ocean color. Deep. Res. Part II 53, 741–770. https://doi.org/10.1016/j.dsr2.2006.01.028

Carr, M.E., Friedrichs, M.A.M., Schmeltz, M., Noguchi Aita, M., Antoine, D., Arrigo, K.R., Asanuma, I., Aumont, O., Barber, R., Behrenfeld, M., Bidigare, R., Buitenhuis, E.T., Campbell, J., Ciotti, A., Dierssen, H., Dowell, M., Dunne, J., Esaias, W., Gentili, B., Gregg, W., Groom, S., Hoepffner, N., Ishizaka, J., Kameda, T., Le Quéré, C., Lohrenz, S., Marra, J., Mélin, F., Moore, K., Morel, A., Reddy, T.E., Ryan, J., Scardi, M., Smyth, T., Turpie, K., Tilstone, G., Waters, K., Yamanaka, Y., 2006. A comparison of global estimates of marine primary production from ocean color. Deep. Res. Part II Top. Stud. Oceanogr. 53, 741–770. https://doi.org/10.1016/j.dsr2.2006.01.028

Carrea, L., Merchant, C.J., 2019. GloboLakes: Lake Surface Water Temperature (LSWT) v4. 0 (1995-2016). Centre for Environmental Data Analysis.

Chavez, F.P., Messié, M., Pennington, J.T., 2011. Marine primary production in relation to climate variability and change. Ann. Rev. Mar. Sci. 3, 227–260. https://doi.org/10.1146/annurev.marine.010908.163917

Chen, S., Hu, C., Barnes, B.B., Wanninkhof, R., Cai, W.J., Barbero, L., Pierrot, D., 2019. A machine learning approach to estimate surface ocean pCO2 from satellite measurements. Remote Sens. Environ. 228, 203–226. https://doi.org/10.1016/j.rse.2019.04.019

Claustre, H., Antoine, D., Babin, M., Belviso, S., Begovic, M., Bianchi, M., Birdwhistell, S., Blanchot, J., Bricaud, A., Bruyant, F., Carlotti, F., Conan, P., Copin, C., Dolan, J., Garczarek, L., Genty, B., Guieu, C., Guillou, L., Gomez, F., Goutx, M., Gorsky, G., Hooker, S., Jacquet, S., Lanoiselle, J., Lantoine, F., Leblanc, K., Leblond, N., Lefèvre, D., Marie, D., Malara, G., Marty, J.-C., Merlivat, L., Migon, C., Momzikoff, A., Morel, A., Moutin, T., Nicolas, E., Oubelkheir, K., Panagiotopoulos, C., Partensky, F., Picheral, M., Prasil, O., Quéguiner, B., Raimbault, P., Ras, J., Ridame, C., Sciandra, A., Sempéré, R., Simon, N., Taillez, D., Van Wambeke, F., Vaulot, D., 2020. BIOGEOCHEMICAL dataset collected during the PROSOPE cruise. https://doi.org/10.17882/71723

Claustre, H., Babin, M., Merien, D., Ras, J., Prieur, L., Dallot, S., Prasil, O., Dousova, H., Moutin, T., 2005. Toward a taxon-specific parameterization of bio-optical models of primary production: A case study in the North Atlantic. J. Geophys. Res. C Ocean. 110, 1–17. https://doi.org/10.1029/2004JC002634

Claustre, H., Sciandra, A., Vaulot, D., 2008. Introduction to the special section bio-optical and biogeochemical conditions in the South East Pacific in late 2004: The BIOSOPE program. Biogeosciences 5, 679–691. https://doi.org/10.5194/bg-5-679-2008

Cloern, J.E., Foster, S.Q., Kleckner, A.E., 2014. Phytoplankton primary production in the world's estuarine-coastal ecosystems. Biogeosciences 11, 2477–2501. https://doi.org/10.5194/bg-11-2477-2014

Cole, J.J., Prairie, Y.T., Caraco, N.F., McDowell, W.H., Tranvik, L.J., Striegl, R.G., Duarte, C.M., Kortelainen, P., Downing, J.A., Middelburg, J.J., Melack, J., 2007. Plumbing the global carbon cycle: Integrating inland waters into the terrestrial carbon budget. Ecosystems 10, 171–184. https://doi.org/10.1007/s10021-006-9013-8

Dalpadado, P., Arrigo, K.R., van Dijken, G.L., Gunasekara, S.S., Ostrowski, M., Bianchi, G., Sperfeld, E., 2021. Warming of the Indian Ocean and its impact on temporal and spatial dynamics of primary production. Prog. Oceanogr. 198, 102688. https://doi.org/10.1016/j.pocean.2021.102688

Dean, W.E., Gorham, E., Dean, W.E., Survey, U.S.G., 1998. Magnitude and significance of carbon burial in lakes, reservoirs, and peatlands. Geology 26, 535–538. https://doi.org/10.1130/0091-7613(1998)026<0535

Di Lorenzo, E., Schneider, N., Cobb, K.M., Franks, P.J.S., Chhak, K., Miller, A.J., McWilliams, J.C., Bograd, S.J., Arango, H., Curchitser, E., Powell, T.M., Rivière, P., 2008. North Pacific Gyre Oscillation links ocean climate and ecosystem change. Geophys. Res. Lett. 35, 2–7. https://doi.org/10.1029/2007GL032838

Dunne, J.P., Sarmiento, J.L., Gnanadesikan, A., 2007. A synthesis of global particle export from the surface ocean and cycling through the ocean interior and on the seafloor. Global Biogeochem. Cycles 21, 1–16. https://doi.org/10.1029/2006GB002907

Eppley, R.W., 1972. Temperature and phytoplankton growth in the sea. Fish. Bull.

Eppley, R.W., Stewart, E., Abbott, M.R., Heyman, U., 1985. Estimating ocean primary production from satellite chlorophyll. Introduction to regional differences and statistics for the Southern California Bight. J. Plankton Res. 7, 57–70.

https://doi.org/10.1093/plankt/7.1.57
Falkowski, P., Scholes, R.J., Boyle, E., Canadell, J., Canfield, D., Elser, J., Gruber, N., Hibbard, K., Hogberg, P., Linder, S., Mackenzie, F.T., Moore, B., Pedersen, T., Rosental, Y., Seitzinger, S., Smetacek, V., Steffen, W., 2000. The global carbon cycle: A test of our knowledge of earth as a system. Science (80-.). 290, 291–296. https://doi.org/10.1126/science.290.5490.291
Falkowski, P.G., Barber, R.T., Smetacek, V., 1998. Biogeochemical controls and feedbacks on ocean primary production. Science (80-.). 281, 200–206. https://doi.org/10.1126/science.281.5374.200
Fragoso, G.M., Poulton, A.J., Yashayaev, I.M., Head, E.J.H., Purdie, D.A., 2017. Spring phytoplankton communities of the Labrador Sea (2005-2014): pigment signatures, photophysiology and elemental ratios. PANGAEA. https://doi.org/10.1594/PANGAEA.871872
Friedrichs, M.A.M., Carr, M.E., Barber, R.T., Scardi, M., Antoine, D., Armstrong, R.A., Asanuma, I., Behrenfeld, M.J., Buitenhuis, E.T., Chai, F., Christian, J.R., Ciotti, A.M., Doney, S.C., Dowell, M., Dunne, J., Gentili, B., Gregg, W., Hoepffner, N., Ishizaka, J., Kameda, T., Lima, I., Marra, J., Mélin, F., Moore, J.K., Morel, A., O'Malley, R.T., O'Reilly, J., Saba, V.S., Schmeltz, M., Smyth, T.J., Tjiputra, J., Waters, K., Westberry, T.K., Winguth, A., 2009. Assessing the uncertainties of model estimates of primary productivity in the tropical Pacific Ocean. J. Mar. Syst. 76, 113–133. https://doi.org/10.1016/j.jmarsys.2008.05.010
Frouin, R., Ramon, D., Boss, E., Jolivet, D., Compiègne, M., Tan, J., Bouman, H., Jackson, T., Franz, B., Platt, T., Sathyendranath, S., 2018. Satellite radiation products for ocean biology and biogeochemistry: Needs, state-of-the-art, gaps, development priorities, and opportunities. Front. Mar. Sci. 5, 1–20. https://doi.org/10.3389/fmars.2018.00003
Gallegos, C.L., 1992. Phytoplankton photosynthesis, productivity, and species composition in a eutrophic estuary: comparison of bloom and non-bloom assemblages. Mar. Ecol. Prog. Ser. 81, 257–267. https://doi.org/10.3354/meps081257
Gallegos, C.L., 2012. Phytoplankton photosynthetic capacity in a shallow estuary: Environmental correlates and interannual variation. Mar. Ecol. Prog. Ser. 463, 23–37. https://doi.org/10.3354/meps09850
Gilerson, A.A., Gitelson, A.A., Zhou, J., Gurlin, D., Moses, W., Ioannou, I., Ahmed, S.A., 2010. Algorithms for remote estimation of chlorophyll-a in coastal and inland waters using red and near infrared bands. Opt. Express 18, 24109. https://doi.org/10.1364/oe.18.024109
Gregg, W.W., Rousseaux, C.S., 2019. Global ocean primary production trends in the modern ocean color satellite record (1998-2015). Environ. Res. Lett. 14, 124011. https://doi.org/10.1088/1748-9326/ab4667
Gulev, S.K., P.W. Thorne, J. Ahn, F.J. Dentener, C.M. Domingues, S. Gerland, D. Gong, D.S. Kaufman, H.C. Nnamchi, J. Quaas, J.A. Rivera, S. Sathyendranath, S.L. Smith, B. Trewin, K. von Schuckmann, and R.S. Vose, 2021. Changing State of the Climate System, in: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N.

Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 287–422, https://doi.org/10.1017/9781009157896.004.

Guo, J., Lu, J., Zhang, Y., Zhou, C., Zhang, S., Wang, D., Lv, X., 2022. Variability of Chlorophyll-a and Secchi Disk Depth (1997–2019) in the Bohai Sea Based on Monthly Cloud-Free Satellite Data Reconstructions. Remote Sens. 14. https://doi.org/10.3390/rs14030639

Hama, T., Miyazaki, T., Ogawa, Y., Iwakuma, T., Takahashi, M., Otsuki, A., Ichimura, S., 1983. Measurement of Photosynthetic Production of a Marine Phytoplankton Population Using a Stable 13C Isotope. Mar. Biol. 73, 31–36.

Harding, L.W., Mallonee, M.E., Perry, E.S., 2002. Toward a predictive understanding of primary productivity in a temperate, partially stratified estuary. Estuar. Coast. Shelf Sci. 55, 437–463. https://doi.org/10.1006/ecss.2001.0917

Harrison, W.G., Platt, T., 1986. Photosynthesis-irradiance relationships in polar and temperate phytoplankton populations. Polar Biol. 5, 153–164. https://doi.org/10.1007/BF00441695

Hill, V.J., Matrai, P.A., Olson, E., Suttles, S., Steele, M., Codispoti, L.A., Zimmerman, R.C., 2013. Synthesis of integrated primary production in the Arctic Ocean: II. In situ and remotely sensed estimates. Prog. Oceanogr. 110, 107–125. https://doi.org/10.1016/j.pocean.2012.11.005

Hirawake, T., Shinmyo, K., Fujiwara, A., Saitoh, S., 2012. Satellite remote sensing of primary productivity in the Bering and Chukchi Seas using an absorption-based approach. ICES J. Mar. Sci. 69, 1194–1204.

Hirawake, T., Takao, S., Horimoto, N., 2011. A phytoplankton absorption-based primary productivity model for remote sensing in the Southern Ocean. Polar Biol. 34, 291–302. https://doi.org/10.1007/s00300-010-0949-y

Holm-Hansen, O., Lorenzen, C.J., Holmes, R.W., Strickland, J.D.H., 1965. Fluorometric determination of chlorophyll. ICES J. Mar. Sci. 30, 3–15. https://doi.org/10.1093/icesjms/30.1.3

Hu, C., Lee, Z., Franz, B., 2012. Chlorophyll a algorithms for oligotrophic oceans: A novel approach based on three-band reflectance difference. J. Geophys. Res. Ocean. 117, 1–25. https://doi.org/10.1029/2011JC007395

Huang, Y., Nicholson, D., Huang, B., Cassar, N., 2021. Global Estimates of Marine Gross Primary Production Based on Machine Learning Upscaling of Field Observations. Global Biogeochem. Cycles 35, 1–18. https://doi.org/10.1029/2020GB006718

IOCCG, 2014. Update of the Quasi-Analytical Algorithm (QAA_v6). Available online. http://www.ioccg.org/groups/Software_OCA/QAA_v6_2014209.pdf.

IOCCG Protocol Series, 2021. Aquatic Primary Productivity Field Protocols for Satellite Validation and Model Synthesis. Balch, W.M., Carranza, M., Cetinic, I., Chaves, J.E., Duhamel, S., Erickson, Z., Fassbender, A., Fernandez-Carrera, A., Ferrón, S., García-Martín, E., Goes, J., Gomes, H., Gorbunov, M., Gundersen, K., Halsey, K., Hirawake, T., Isada, T., Juranek, L., Kulk, G., Langdon, C., Letelier, R., López-Sandoval, D., Mannino, A., Marra, J., Neale, P., Nicholson, D., Silsbe, G., Stanley, R., Vandermeulen, R.A. IOCCG Ocean Optics and Biogeochemistry Protocols for Satellite Ocean Colour

Sensor Validation, Volume 7.0, edited by R.A. Vandermeulen, J. E. Chaves, IOCCG, Dartmouth, NS, Canada.

Jamet, C., Ibrahim, A., Ahmad, Z., Angelini, F., Babin, M., Behrenfeld, M.J., Boss, E., Cairns, B., Churnside, J., Chowdhary, J., Davis, A.B., Dionisi, D., Duforêt-Gaurier, L., Franz, B., Frouin, R., Gao, M., Gray, D., Hasekamp, O., He, X., Hostetler, C., Kalashnikova, O. V., Knobelspiesse, K., Lacour, L., Loisel, H., Martins, V., Rehm, E., Remer, L., Sanhaj, I., Stamnes, K., Stamnes, S., Victori, S., Werdell, J., Zhai, P.-W., 2019. Going Beyond Standard Ocean Color Observations: Lidar and Polarimetry. Front. Mar. Sci. 6. https://doi.org/10.3389/fmars.2019.00251

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., 1996. The NCEP/NCAR 40-year reanalysis project. Bull. Am. Meteorol. Soc. 77, 437–472.

Kameda, T., Ishizaka, J., 2005. Size-fractionated primary production estimated by a two-phytoplankton community model applicable to ocean color remote sensing. J. Oceanogr. 61, 663–672. https://doi.org/10.1007/s10872-005-0074-7

Kauer, T., Kutser, T., Arst, H., Danckaert, T., Nõges, T., 2015. Modelling primary production in shallow well mixed lakes based on MERIS satellite data. Remote Sens. Environ. 163, 253–261. https://doi.org/10.1016/j.rse.2015.03.023

Kauer, T., Kutser, T., Arst, H., Danckaert, T., Nõges, T., 2015. Modelling primary production in shallow well mixed lakes based on MERIS satellite data. Remote Sens. Environ. 163, 253–261. https://doi.org/10.1016/j.rse.2015.03.023

Kiefer, D.A., Mitchell, B.G., 1983. A simple, steady state description of phytoplankton growth based on absorption cross section and quantum efficiency. Limnol. Oceanogr. 28, 770–776. https://doi.org/10.4319/lo.1983.28.4.0770

Kishino, M., Takahashi, M., Okami, N., Ichimura, S., 1985. Estimation of the spectral absorption coefficients of phytoplankton in the sea. Bull. Mar. Sci. 37, 634–642.

Kug, J.S., Kang, I.S., Lee, J.Y., Jhun, J.G., 2004. A statistical approach to Indian Ocean sea surface temperature prediction using a dynamical ENSO prediction. Geophys. Res. Lett. 31, 1–5. https://doi.org/10.1029/2003GL019209

Kulk, G., Platt, T., Dingle, J., Jackson, T., Jönsson, B.F., Bouman, H.A., Babin, M., Brewin, R.J.W., Doblin, M., Estrada, M., Figueiras, F.G., Furuya, K., González-Benítez, N., Gudfinnsson, H.G., Gudmundsson, K., Huang, B., Isada, T., Kovač, Ž., Lutz, V.A., Marañón, E., Raman, M., Richardson, K., Rozema, P.D., van de Poll, W.H., Segura, V., Tilstone, G.H., Uitz, J., van Dongen-Vogels, V., Yoshikawa, T., Sathyendranath, S., 2021a. BICEP/NCEO: Monthly global Marine Phytoplankton Primary Production, between 1998-2020 at 9 km resolution (derived from the Ocean Colour Climate Change Initiative v4.2 dataset). NERC EDS Centre for Environmental Data Analysis, 22 December 2021. doi:10.5285/69b2c9c6c4714517ba10dab3515e4ee6.

Kulk, G., Platt, T., Dingle, J., Jackson, T., Jönsson, B.F., Bouman, H.A., Babin, M., Brewin, R.J.W., Doblin, M., Estrada, M., Figueiras, F.G., Furuya, K., González-Benítez, N., Gudfinnsson, H.G., Gudmundsson, K., Huang, B., Isada, T., Kovač, Ž., Lutz, V.A., Marañón, E., Raman, M., Richardson, K., Rozema, P.D., van de Poll, W.H., Segura, V., Tilstone, G.H., Uitz, J., van Dongen-Vogels, V., Yoshikawa, T., Sathyendranath, S., 2021b. Correction: Kulk et al. Primary production, an index of climate change in the

ocean: Satellite-based estimates over two decades (Remote Sens., (2020), 12, (826), 10.3390/rs12050826). Remote Sens. 13, 1–13. https://doi.org/10.3390/rs13173462

Kulk, G., Platt, T., Dingle, J., Jackson, T., Jönsson, B.F., Bouman, H.A., Babin, M., Brewin, R.J.W., Doblin, M., Estrada, M., Figueiras, F.G., Furuya, K., González-Benítez, N., Gudfinnsson, H.G., Gudmundsson, K., Huang, B., Isada, T., Kovač, Ž., Lutz, V.A., Marañón, E., Raman, M., Richardson, K., Rozema, P.D., van de Poll, W.H., Segura, V., Tilstone, G.H., Uitz, J., van Dongen-Vogels, V., Yoshikawa, T., Sathyendranath, S., 2020. Primary production, an index of climate change in the ocean: Satellite-based estimates over two decades. Remote Sens. 12. https://doi.org/10.3390/rs12050826

Lavergne, T., Macdonald Sørensen, A., Kern, S., Tonboe, R., Notz, D., Aaboe, S., Bell, L., Dybkjær, G., Eastwood, S., Gabarro, C., Heygster, G., Anne Killie, M., Brandt Kreiner, M., Lavelle, J., Saldo, R., Sandven, S., Pedersen, L.T., 2019. Version 2 of the EUMETSAT OSI SAF and ESA CCI sea-ice concentration climate data records. Cryosphere 13, 49–78. https://doi.org/10.5194/tc-13-49-2019

Lee, Y.J., Matrai, P.A., Friedrichs, M.A.M., Saba, V.S., Antoine, D., Ardyna, M., Asanuma, I., Babin, M., Belanger, S., Benoit-Gagne, M., Devred, E., Fernandez-Mendez, M., Gentili, B., Hirawake, T., Kang, S.-H., Kameda, T., Katlein, C., Lee, S.H., Lee, Z., Melin, F., Scardi, M., Smyth, T.J., Tang, S., Turpie, K.R., Waters, K.J., Westberry, T.K., 2015. An assessment of phytoplankton primary productivity in the Arctic Ocean from satellite ocean color/in situ chlorophyll-a based models. J. Geophys. Res. Ocean. Res. 120, 6508–6541. https://doi.org/10.1002/2015JC011018.Received

Lee, Z., Carder, K.L., Arnone, R.A., 2002. Deriving inherent optical properties from water color: a multiband quasi-analytical algorithm for optically deep waters. Appl. Opt. 41, 5755–5772. https://doi.org/10.1364/AO.41.005755

Lee, Z., Carder, K.L., Marra, J., Steward, R.G., Perry, M.J., 1996. Estimating primary production at depth from remote sensing. Appl. Opt. 35, 463. https://doi.org/10.1364/ao.35.000463

Lee, Z., Du, K., Arnone, R., Liew, S., Penta, B., 2005. Penetration of solar radiation in the upper ocean: A numerical model for oceanic and coastal waters. J. Geophys. Res. C Ocean. 110, 1–12. https://doi.org/10.1029/2004JC002780

Lee, Z., Lance, V.P., Shang, S., Vaillancourt, R., Freeman, S., Lubac, B., Hargreaves, B.R., Del Castillo, C., Miller, R., Twardowski, M., Wei, G., 2011. An assessment of optical properties and primary production derived from remote sensing in the Southern Ocean (SO GasEx). J. Geophys. Res. Ocean. 116, 1–15. https://doi.org/10.1029/2010JC006747

Lee, Z., Marra, J., Jane, M., Kahru, M., 2015. Estimating oceanic primary productivity from ocean color remote sensing: A strategic assessment. J. Mar. Syst. 149, 50–59. https://doi.org/10.1016/j.jmarsys.2014.11.015

Lee, Z., Weidemann, A., Kindle, J., Arnone, R., Carder, K.L., Davis, C., 2007. Euphotic zone depth: Its derivation and implication to ocean-color remote sensing. J. Geophys. Res. Ocean. 112, 1–11. https://doi.org/10.1029/2006JC003802

Levitus, S., 1982. Climatological atlas of the world ocean. US Department of Commerce, National Oceanic and Atmospheric Administration.

Lewis, K.M., Van Dijken, G.L., Arrigo, K.R., 2020. Changes in phytoplankton concentration now drive increased Arctic Ocean primary production. Science (80-.). 369, 198–202.

https://doi.org/10.1126/science.aay8380
Lewis, M., Smith, J., 1983. A small volume, short-incubation-time method for measurement of photosynthesis as a function of incident irradiance. Mar. Ecol. Prog. Ser. 13, 99–102. https://doi.org/10.3354/meps013099
Likens, G.E., 1975. Primary Production of Inland Aquatic Ecosystems 185–202. https://doi.org/10.1007/978-3-642-80913-2_9
Liu, X., Wang, M., 2018. Gap filling of missing data for VIIRS global ocean color products using the DINEOF Method. IEEE Trans. Geosci. Remote Sens. 56, 4464–4476. https://doi.org/10.1109/TGRS.2018.2820423
Liu, X., Wang, M., 2019. Filling the gaps of missing data in the merged VIIRS SNPP/NOAA-20 ocean color product using the DINEOF method. Remote Sens. 11. https://doi.org/10.3390/rs11020178
Lobanova, P., Tilstone, G.H., Bashmachnikov, I., Brotas, V., 2018. Accuracy assessment of primary production models with and without photoinhibition using Ocean-Colour climate change initiative data in the North East Atlantic Ocean. Remote Sens. 10, 1–24. https://doi.org/10.3390/rs10071116
Longhurst, A., Sathyendranath, S., Platt, T., Caverhill, C., 1995. An estimate of global primary production in the ocean from satellite radiometer data. J. Plankton Res. 17, 1245–1271.
Longhurst, A.R., 2007. Ecological geography of the sea, 2nd ed. Elsevier Academic Press: Cambridge, MA, USA
Ma, C., Zhao, J., Ai, B., Sun, S., 2021. Two-Decade Variability of Sea Surface Temperature and Chlorophyll-a in the Northern South China Sea as Revealed by Reconstructed Cloud-Free Satellite Data. IEEE Trans. Geosci. Remote Sens. 59, 9033–9046. https://doi.org/10.1109/TGRS.2021.3051025
MacCallum, S.N., Merchant, C.J., 2012. Surface water temperature observations of large lakes by optimal estimation. Can. J. Remote Sens. 38, 25–45. https://doi.org/10.5589/m12-010
Mackey, D.J., Parslow, J., Higgins, H.W., Griffiths, F.B., O'Sullivan, J.E., 1995. Plankton productivity and biomass in the western equatorial Pacific: Biological and physical controls. Deep. Res. Part II 42, 499–533. https://doi.org/10.1016/0967-0645(95)00038-R
Madhupratap, M., Prasanna Kumar, S., Bhattathiri, P.M.A., Dileep Kumar, M., Raghukumar, S., Nair, K.K.C., Ramaiah, N., 1996. Mechanism of the biological response to winter cooling in the northeastern Arabian Sea. Nature 384, 549–552. https://doi.org/10.1038/384549a0
Mann, H.B., 1945. Non-Parametric Test Against Trend. Econometrica 13, 245–259.
Marcel, B., Morel, A., Gagnon, R., 1994. An incubator designed for extensive and sensitive measurements of phytoplankton photosynthetic parameters. Limnol. Oceanogr. 39, 694–702. https://doi.org/10.4319/lo.1994.39.3.0694
Maritorena, S., d'Andon, O.H.F., Mangin, A., Siegel, D.A., 2010. Merged satellite ocean color data products using a bio-optical model: Characteristics, benefits and issues. Remote Sens. Environ. 114, 1791–1804. https://doi.org/10.1016/j.rse.2010.04.002
Marra, J., 2002. Approaches to the measurement of plankton production. Phytoplankt. Product. Carbon Assim. Mar. Freshw. Ecosyst. 78–108.

Marra, J., 2009. Net and gross productivity: Weighing in with 14C. Aquat. Microb. Ecol. 56, 123–131. https://doi.org/10.3354/ame01306

Marra, J., Ho, C., Trees, C., 2003. An alternative algorithm for the calculation of primary productivity from remote sensing data. Lamont Doherty Earth Obs. Tech. Rep.

Martinez, E., Antoine, D., D'Ortenzio, F., Gentili, B., 2009. Climate-driven basin-scale decadal oscillations of oceanic phytoplankton. Science (80-.). 326, 1253–1256. https://doi.org/10.1126/science.1177012

Massicotte, P., Amon, R.M.W., Antoine, D., Archambault, P., Balzano, S., Bélanger, S., Benner, R., Boeuf, D., Bricaud, A., Bruyant, F., Chaillou, G., Chami, M., Charrière, B., Chen, J., Claustre, H., Coupel, P., Delsaut, N., Doxaran, D., Ehn, J., Fichot, C., Forget, M.-H., Fu, P., Gagnon, J., Garcia, N., Gasser, B., Ghiglione, J.-F., Gorsky, G., Gosselin, M., Gourvil, P., Gratton, Y., Guillot, P., Heipieper, H.J., Heussner, S., Hooker, S.B., Huot, Y., Jeanthon, C., Jeffrey, W., Joux, F., Kawamura, K., Lansard, B., Leymarie, E., Link, H., Lovejoy, C., Marec, C., Marie, D., Martin, J., Martín, J., Massé, G., Matsuoka, A., McKague, V., Mignot, A., Miller, W.L., Miquel, J.-C., Mucci, A., Ono, K., Ortega-Retuerta, E., Panagiotopoulos, C., Papakyriakou, T., Picheral, M., Piepenburg, D., Prieur, L., Raimbault, P., Ras, J., Reynolds, R.A., Rochon, A., Rontani, J.-F., Schmechtig, C., Schmidt, S., Sempéré, R., Shen, Y., Song, G., Stramski, D., Tachibana, E., Thirouard, A., Tolosa, I., Tremblay, J.-É., Vaïtilingom, M., Vaulot, D., Vaultier, F., Volkman, J.K., Vonk, J.E., Xie, H., Zheng, G., Babin, M., 2009. The Malina oceanographic expedition: How do changes in ice cover, permafrost and UV radiation impact biodiversity and biogeochemical fluxes in the Arctic Ocean? https://doi.org/10.17882/75345

Merchant, C.J., Embury, O., Bulgin, C.E., Block, T., Corlett, G.K., Fiedler, E., Good, S.A., Mittaz, J., Rayner, N.A., Berry, D., Eastwood, S., Taylor, M., Tsushima, Y., Waterfall, A., Wilson, R., Donlon, C., 2019. Satellite-based time-series of sea-surface temperature since 1981 for climate applications. Sci. Data 6, 1–18. https://doi.org/10.1038/s41597-019-0236-x

Milligan, A.J., Halsey, K.H., Behrenfeld, M.J., 2015. Advancing interpretations of 14C-uptake measurements in the context of phytoplankton physiology and ecology. J. Plankton Res. 37, 692–698. https://doi.org/10.1093/plankt/fbv051

Mishra, S., Mishra, D.R., 2012. Normalized difference chlorophyll index: A novel model for remote estimation of chlorophyll-a concentration in turbid productive waters. Remote Sens. Environ. 117, 394–406. https://doi.org/10.1016/j.rse.2011.10.016

Morel, A., 1991. Light and marine photosynthesis: a spectral model with geochemical and climatological implications. Prog. Oceanogr. 26, 263–306. https://doi.org/10.1016/0079-6611(91)90004-6

Morel, A., Antoine, D., Babin, M., Dandonneau, Y., 1996. Measured and modeled primary production in the northeast Atlantic (EUMELI JGOFS program): The impact of natural variations in photosynthetic parameters on model predictive skill. Deep. Res. Part I Oceanogr. Res. Pap. 43, 1273–1304. https://doi.org/10.1016/0967-0637(96)00059-3

Morel, A., Huot, Y., Gentili, B., Werdell, P.J., Hooker, S.B., Franz, B.A., 2007. Examining the consistency of products derived from various ocean color sensors in open ocean (Case 1) waters in the perspective of a multi-sensor approach. Remote Sens. Environ.

111, 69–88. https://doi.org/10.1016/j.rse.2007.03.012
Nielsen, E.S., 1952. The use of radio-active carbon (c14) for measuring organic production in the sea. ICES J. Mar. Sci. 18, 117–140. https://doi.org/10.1093/icesjms/18.2.117
Pahlevan, N., Smith, B., Schalles, J., Binding, C., Cao, Z., Ma, R., Alikas, K., Kangro, K., Gurlin, D., Hà, N., Matsushita, B., Moses, W., Greb, S., Lehmann, M.K., Ondrusek, M., Oppelt, N., Stumpf, R., 2020. Seamless retrievals of chlorophyll-a from Sentinel-2 (MSI) and Sentinel-3 (OLCI) in inland and coastal waters: A machine-learning approach. Remote Sens. Environ. 240, 111604. https://doi.org/10.1016/j.rse.2019.111604
Palevsky, H.I., Quay, P.D., Lockwood, D.E., Nicholson, D.P., 2016. The annual cycle of gross primary production, net community production, and export efficiency across the North Pacific Ocean. Global Biogeochem. Cycles 30, 361–380. https://doi.org/10.1111/1462-2920.13280
Pei, S., Laws, E.A., 2013. Does the 14C method estimate net photosynthesis? Implications from batch and continuous culture studies of marine phytoplankton. Deep. Res. Part I Oceanogr. Res. Pap. 82, 1–9. https://doi.org/10.1016/j.dsr.2013.07.011
Perry, M., 2018. Primary productivity measurements from on-deck bottle incubations during R/V Knorr cruise KN193-03 and R/V Bjarni Saemundsson cruises B10-2008 and B4-2008 to the subpolar North Atlantic, Iceland Basin in 2008. https://doi.org/http://lod.bco-dmo.org/id/dataset/746215
Petrenko, D., Pozdnyakov, D., Johannessen, J., Counillon, F., Sychov, V., 2013. Satellite-derived multi-year trend in primary production in the Arctic Ocean. Int. J. Remote Sens. 34, 3903–3937. https://doi.org/10.1080/01431161.2012.762698
Picart, S.S., Sathyendranath, S., Dowell, M., Moore, T., Platt, T., 2014. Remote sensing of assimilation number for marine phytoplankton. Remote Sens. Environ. 146, 87–96.
Pinckney, J., Papa, R., Zingmark, R., 1994. Comparison of high-performance liquid chromatographic, spectrophotometric, and fluorometric methods for determining chlorophyll a concentrations in estuarine sediments. J. Microbiol. Methods 19, 59–66. https://doi.org/10.1016/0167-7012(94)90026-4
Pinkerton, M.H., Boyd, P.W., Deppeler, S., Hayward, A., Höfer, J., Moreau, S., 2021. Evidence for the Impact of Climate Change on Primary Producers in the Southern Ocean. Front. Ecol. Evol. 9, 1–19. https://doi.org/10.3389/fevo.2021.592027
Platt, T., Jassby, A.D., 1976. The relationship between photosynthesis and light for natural assemblages of coastal marine phytoplankton. J. Phycol. 12, 421–430.
Platt, T., Sathyendranath, S., 1988. Oceanic Primary Production: Estimation by Remote Sensing at Local and Regional Scales. Science (80). 241, 1613–1620.
Platt, T., Sathyendranath, S., 1999. Spatial structure of pelagic ecosystem processes in the global ocean. Ecosystems 2, 384–394. https://doi.org/10.1007/s100219900088
Platt, T., Sathyendranath, S., Forget, M.H., White, G.N., Caverhill, C., Bouman, H., Devred, E., Son, S.H., 2008. Operational estimation of primary production at large geographical scales. Remote Sens. Environ. 112, 3437–3448. https://doi.org/10.1016/j.rse.2007.11.018
Platt, T., Sathyendranath, S., Ravindran, P., 1990. Primary production by phytoplankton: Analytic solutions for daily rates per unit area of water surface. Proc. R. Soc. B Biol. Sci. 241, 101–111. https://doi.org/10.1098/rspb.1990.0072
Platt, T., Sathyendranath, S., White, G.N., Jackson, T., Picart, S.S., Bouman, H., 2017.

Primary production: Sensitivity to surface irradiance and implications for archiving data. Front. Mar. Sci. 4, 1–9. https://doi.org/10.3389/fmars.2017.00387

Prasad, M.B.K., Sapiano, M.R.P., Anderson, C.R., Long, W., Murtugudde, R., 2010. Long-Term Variability of Nutrients and Chlorophyll in the Chesapeake Bay: A Retrospective Analysis, 1985-2008. Estuaries and Coasts 33, 1128–1143. https://doi.org/10.1007/s12237-010-9325-y

Prentice, I.C., Farquhar, G.D., Fasham, M.J.R., Goulden, M.L., Heimann, M., Jaramillo, V.J., Kheshgi, H.S., Le Quéré, C., Scholes, R.J., Wallace, D.W.R., 2001. The carbon cycle and atmospheric CO2, in: Climate Change 2000: The Science of Climate Change. Contributions of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, pp. 183–237.

Qin, P., Simis, S.G.H., Tilstone, G.H., 2017. Radiometric validation of atmospheric correction for MERIS in the Baltic Sea based on continuous observations from ships and AERONET-OC. Remote Sens. Environ. 200, 263–280. https://doi.org/10.1016/j.rse.2017.08.024

Ras, J., Claustre, H., Uitz, J., 2008. Spatial variability of phytoplankton pigment distributions in the Subtropical South Pacific Ocean: Comparison between in situ and predicted data. Biogeosciences 5, 353–369. https://doi.org/10.5194/bg-5-353-2008

Raymond, P.A., Hartmann, J., Lauerwald, R., Sobek, S., McDonald, C., Hoover, M., Butman, D., Striegl, R., Mayorga, E., Humborg, C., Kortelainen, P., Dürr, H., Meybeck, M., Ciais, P., Guth, P., 2013. Global carbon dioxide emissions from inland waters. Nature 503, 355–359. https://doi.org/10.1038/nature12760

Regnier, P., Friedlingstein, P., Ciais, P., Mackenzie, F.T., Gruber, N., Janssens, I.A., Laruelle, G.G., Lauerwald, R., Luyssaert, S., Andersson, A.J., Arndt, S., Arnosti, C., Borges, A. V., Dale, A.W., Gallego-Sala, A., Goddéris, Y., Goossens, N., Hartmann, J., Heinze, C., Ilyina, T., Joos, F., Larowe, D.E., Leifeld, J., Meysman, F.J.R., Munhoven, G., Raymond, P.A., Spahni, R., Suntharalingam, P., Thullner, M., 2013. Anthropogenic perturbation of the carbon fluxes from land to ocean. Nat. Geosci. 6, 597–607. https://doi.org/10.1038/ngeo1830

Rousseaux, C.S., Gregg, W.W., 2014. Interannual variation in phytoplankton primary production at a global scale. Remote Sens. 6, 1–19. https://doi.org/10.3390/rs6010001

Roy, S., Sathyendranath, S., Platt, T., 2017. Size-partitioned phytoplankton carbon and carbon-to-chlorophyll ratio from ocean colour by an absorption-based bio-optical algorithm. Remote Sens. Environ. 194, 177–189. https://doi.org/10.1016/j.rse.2017.02.015

Saba, V.S., Friedrichs, M.A.M., Antoine, D., Armstrong, R.A., Asanuma, I., Behrenfeld, M.J., Ciotti, A.M., Dowell, M., Hoepffner, N., Hyde, K.J.W., Ishizaka, J., Kameda, T., Marra, J., Mlin, F., Morel, A., O'Reilly, J., Scardi, M., Smith, W.O., Smyth, T.J., Tang, S., Uitz, J., Waters, K., Westberry, T.K., 2011. An evaluation of ocean color model estimates of marine primary productivity in coastal and pelagic regions across the globe. Biogeosciences 8, 489–503. https://doi.org/10.5194/bg-8-489-2011

Saba, V.S., Friedrichs, M.A.M., Carr, M.E., Antoine, D., Armstrong, R.A., Asanuma, I., Aumont, O., Bates, N.R., Behrenfeld, M.J., Bennington, V., Bopp, L., Bruggeman, J., Buitenhuis, E.T., Church, M.J., Ciotti, A.M., Doney, S.C., Dowell, M., Dunne, J.,

Dutkiewicz, S., Gregg, W., Hoepffner, N., Hyde, K.J.W., Ishizaka, J., Kameda, T., Karl, D.M., Lima, I., Lomas, M.W., Marra, J., McKinley, G.A., Melin, F., Moore, J.K., Morel, A., O'Reilly, J., Salihoglu, B., Scardi, M., Smyth, T.J., Tang, S., Tjiputra, J., Uitz, J., Vichi, M., Waters, K., Westberry, T.K., Yool, A., 2010. Challenges of modeling depth-integrated marine primary productivity over multiple decades: A case study at BATS and HOT. Global Biogeochem. Cycles 24, 1–21. https://doi.org/10.1029/2009GB003655

Sathyendranath, S, Platt, T., Brewin, R.J., Jackson, T., 2019. Primary production distribution, in: Cochran, J.K., Bokuniewicz, H.J., Yager, P.L. (Eds.), Encyclopedia of Ocean Sciences. Academic Press, 635–640. https://doi.org/10.1016/B978-0-12-409548-9.04304-9

Sathyendranath, S., Longhurst, A., Caverhill, C.M., Platt, T., 1995. Regionally and seasonally differentiated primary production in the North Atlantic. Deep. Res. Part I 42, 1773–1802. https://doi.org/10.1016/0967-0637(95)00059-F

Sathyendranath, S., Platt, T., 2007. Spectral effects in bio-optical control on the ocean system. Oceanologia 49.

Sathyendranath, S., Platt, T., Kovač, Ž., Dingle, J., Jackson, T., Brewin, R.J.W., Franks, P., Marañón, E., Kulk, G., Bouman, H.A., 2020. Reconciling models of primary production and photoacclimation [Invited]. Appl. Opt. 59, C100. https://doi.org/10.1364/ao.386252

Sathyendranath, S., Stuart, V., Nair, A., Oka, K., Nakane, T., Bouman, H., Forget, M.H., Maass, H., Platt, T., 2009. Carbon-to-chlorophyll ratio and growth rate of phytoplankton in the sea. Mar. Ecol. Prog. Ser. 383, 73–84. https://doi.org/10.3354/meps07998

Sathyendranath, Shubha, Brewin, R.J.W., Brockmann, C., Brotas, V., Calton, B., Chuprin, A., Cipollini, P., Couto, A.B., Dingle, J., Doerffer, R., Donlon, C., Dowell, M., Farman, A., Grant, M., Groom, S., Horseman, A., Jackson, T., Krasemann, H., Lavender, S., Martinez-Vicente, V., Mazeran, C., Mélin, F., Moore, T.S., Müller, D., Regner, P., Roy, S., Steele, C.J., Steinmetz, F., Swinton, J., Taberner, M., Thompson, A., Valente, A., Zühlke, M., Brando, V.E., Feng, H., Feldman, G., Franz, B.A., Frouin, R., Gould, R.W., Hooker, S.B., Kahru, M., Kratzer, S., Mitchell, B.G., Muller-Karger, F.E., Sosik, H.M., Voss, K.J., Werdell, J., Platt, T., 2019. An ocean-colour time series for use in climate studies: The experience of the ocean-colour climate change initiative (OC-CCI). Sensors (Switzerland) 19. https://doi.org/10.3390/s19194285

Sayers, M., Bosse, K., Fahnenstiel, G., Shuchman, R., 2020. Carbon Fixation Trends in Eleven of the World's Largest Lakes: 2003–2018. Water 12, 2–16. https://doi.org/doi:10.3390/w12123500

Sayers, M.J., Fahnenstiel, G.L., Shuchman, R.A., Bosse, K.R., 2021. A new method to estimate global freshwater phytoplankton carbon fixation using satellite remote sensing: initial results. Int. J. Remote Sens. 42, 3708–3730. https://doi.org/10.1080/01431161.2021.1880661

Schuback, N., Tortell, P.D., Berman-Frank, I., Campbell, D.A., Ciotti, A., Courtecuisse, E., Erickson, Z.K., Fujiki, T., Halsey, K., Hickman, A.E., Huot, Y., Gorbunov, M.Y., Hughes, D.J., Kolber, Z.S., Moore, C.M., Oxborough, K., Prášil, O., Robinson, C.M., Ryan-Keogh, T.J., Silsbe, G., Simis, S., Suggett, D.J., Thomalla, S., Varkey, D.R., 2021. Single-Turnover Variable Chlorophyll Fluorescence as a Tool for Assessing Phytoplankton Photosynthesis and Primary Productivity: Opportunities, Caveats and

Recommendations. Front. Mar. Sci. 8. https://doi.org/10.3389/fmars.2021.690607
Sedwick, P.N., Blain, S., Quéguiner, B., Griffiths, F.B., Fiala, M., Bucciarelli, E., Denis, M., 2002. Resource limitation of phytoplankton growth in the Crozet Basin, Subantarctic Southern Ocean. Deep. Res. Part II Top. Stud. Oceanogr. 49, 3327–3349. https://doi.org/10.1016/S0967-0645(02)00086-3
Shen, M., Duan, H., Cao, Z., Xue, K., Qi, T., Ma, J., Liu, D., Song, K., Huang, C., Song, X., 2020. Sentinel-3 OLCI observations of water clarity in large lakes in eastern China: Implications for SDG 6.3.2 evaluation. Remote Sens. Environ. 247, 111950. https://doi.org/10.1016/j.rse.2020.111950
Silsbe, G.M., Behrenfeld, M.J., Halsey, K.H., Milligan, A.J., Westberry, T.K., 2016. The CAFE model: A net production model for global ocean phytoplankton. Glob. Biogeochem. Cycles 1756–1777. https://doi.org/10.1002/2016GB005521.Received
Smyth, T.J., Pemberton, K.L., Aiken, J., Geider, R.J., 2004. A methodology to determine primary production and phytoplankton photosynthetic parameters from Fast Repetition Rate Fluorometry. J. Plankton Res. 26, 1337–1350. https://doi.org/10.1093/plankt/fbh124
Smyth, T.J., Tilstone, G.H., Groom, S.B., 2005. Integration of radiative transfer into satellite models of ocean primary production. J. Geophys. Res. C Ocean. 110, 1–11. https://doi.org/10.1029/2004JC002784
Son, S.H., Campbell, J., Dowell, M., Yoo, S., Noh, J., 2005. Primary production in the Yellow Sea determined by ocean color remote sensing. Mar. Ecol. Prog. Ser. 303, 91–103. https://doi.org/10.3354/meps303091
Son, S.H., Wang, M., Harding, L.W., 2014. Satellite-measured net primary production in the Chesapeake Bay. Remote Sens. Environ. 144, 109–119. https://doi.org/10.1016/j.rse.2014.01.018
Soomets, T., Kutser, T., Wüest, A., Bouffard, D., 2019. Spatial and temporal changes of primary production in a deep peri-alpine lake. Inl. Waters 9, 49–60. https://doi.org/10.1080/20442041.2018.1530529
Steinmetz, F., Deschamps, P.-Y., Ramon, D., 2011. Atmospheric correction in presence of sun glint: application to MERIS. Opt. Express 19, 9783. https://doi.org/10.1364/oe.19.009783
Taboada, F.G., Barton, A.D., Stock, C.A., Dunne, J., John, J.G., 2019. Seasonal to interannual predictability of oceanic net primary production inferred from satellite observations. Prog. Oceanogr. 170, 28–39. https://doi.org/10.1016/j.pocean.2018.10.010
Takamura, N., Aizaki, M., 1991. Change in Primary Production in Lake Kasumigaura (1986-1989) accompanied by Transition of Dominant Species. Japanese J. Limnol. 52, 173–187. https://doi.org/10.3739/rikusui.52.173
Takamura, N., Iwakuma, T., Yasuno, M., 1987. Primary Production in Lake Kasumigaura, 1981-1985. Japanese J. Limnol. 48, 13–38. https://doi.org/10.3739/rikusui.48.Special_13
Takamura, N., Nakagawa, M., 2016. Photosynthesis and primary production in Lake Kasumigaura (Japan) monitored monthly since 1981. Ecol. Res. 31, 287. https://doi.org/10.1007/s11284-016-1347-x
Tang, S., Chen, C., Zhan, H., Zhang, T., 2008. Determination of ocean primary productivity using support vector machines. Int. J. Remote Sens. 29, 6227–6236.

https://doi.org/10.1080/01431160802175355
Thomalla, S.J., Ogunkoya, A.G., Vichi, M., Swart, S., 2017. Using optical sensors on gliders to estimate phytoplankton carbon concentrations and chlorophyll-to-carbon ratios in the Southern Ocean. Front. Mar. Sci. 4, 1–19. https://doi.org/10.3389/FMARS.2017.00034
Tilstone, G., Smyth, T., Poulton, A., Hutson, R., 2009. Measured and remotely sensed estimates of primary production in the Atlantic Ocean from 1998 to 2005. Deep. Res. Part II Top. Stud. Oceanogr. 56, 918–930. https://doi.org/10.1016/j.dsr2.2008.10.034
Tilstone, G.H., Miller, P.I., Brewin, R.J.W., Priede, I.G., 2014. Enhancement of primary production in the North Atlantic outside of the spring bloom, identified by remote sensing of ocean colour and temperature. Remote Sens. Environ. 146, 77–86. https://doi.org/10.1016/j.rse.2013.04.021
Tilstone, G.H., Pardo, S., Dall'Olmo, G., Brewin, R.J.W., Nencioli, F., Dessailly, D., Kwiatkowska, E., Casal, T., Donlon, C., 2021. Performance of Ocean Colour Chlorophyll a algorithms for Sentinel-3 OLCI, MODIS-Aqua and Suomi-VIIRS in open-ocean waters of the Atlantic. Remote Sens. Environ. 260, 112444. https://doi.org/10.1016/j.rse.2021.112444
Tranvik, L.J., Downing, J.A., Cotner, J.B., Loiselle, S.A., Striegl, R.G., Ballatore, T.J., Dillon, P., Finlay, K., Fortino, K., Knoll, L.B., Kortelainen, P.L., Kutser, T., Larsen, S., Laurion, I., Leech, D.M., Leigh McCallister, S., McKnight, D.M., Melack, J.M., Overholt, E., Porter, J.A., Prairie, Y., Renwick, W.H., Roland, F., Sherman, B.S., Schindler, D.W., Sobek, S., Tremblay, A., Vanni, M.J., Verschoor, A.M., Von Wachenfeldt, E., Weyhenmeyer, G.A., 2009. Lakes and reservoirs as regulators of carbon cycling and climate. Limnol. Oceanogr. 54, 2298–2314. https://doi.org/10.4319/lo.2009.54.6_part_2.2298
Tripathy, S.C., Ishizaka, J., Siswanto, E., Shibata, T., Mino, Y., 2012. Modification of the vertically generalized production model for the turbid waters of Ariake Bay, southwestern Japan. Estuar. Coast. Shelf Sci. 97, 66–77. https://doi.org/10.1016/j.ecss.2011.11.025
Uitz, J., Claustre, H., Gentili, B., Stramski, D., 2010. Phytoplankton class-specific primary production in the world's oceans: Seasonal and interannual variability from satellite observations. Global Biogeochem. Cycles 24, 1–19. https://doi.org/10.1029/2009GB003680
Uitz, J., Claustre, H., Griffiths, F.B., Ras, J., Garcia, N., Sandroni, V., 2009. A phytoplankton class-specific primary production model applied to the Kerguelen Islands region (Southern Ocean). Deep. Res. Part I Oceanogr. Res. Pap. 56, 541–560. https://doi.org/10.1016/j.dsr.2008.11.006
Uitz, J., Claustre, H., Morel, A., Hooker, S.B., 2006. Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll. J. Geophys. Res. Ocean. 111. https://doi.org/10.1029/2005JC003207
Uitz, J., Huot, Y., Bruyant, F., Babin, M., Claustre, H., 2008. Relating phytoplankton photophysiological properties to community structure on large scales. Limnol. Oceanogr. 53, 614–630. https://doi.org/10.4319/lo.2008.53.2.0614
van Dongen-Vogels, V., Everett, J.D., Doblin, M.A., 2015. Phytoplankton production, photosynthetic (P vs E) parameters, and particulate organic carbon and nitrogen within

distinct water masses of eastern Australian coastal and shelf waters between 29°S and 36°S during spring 2010. Univ. Technol. Sydney. https://doi.org/10.1594/PANGAEA.843554

Vantrepotte, V., Mélin, F., 2011. Inter-annual variations in the SeaWiFS global chlorophyll a concentration (1997-2007). Deep. Res. Part I Oceanogr. Res. Pap. 58, 429–441. https://doi.org/10.1016/j.dsr.2011.02.003

Vernet, M., 2004. Production vs Irradiance data from RVIB Nathaniel B. Palmer cruise NBP0103 in the Southern Ocean in 2001 (SOGLOBEC project). https://doi.org/10.1575/1912/bco-dmo.2375.1

Warren, M.A., Simis, S.G.H., Martinez-Vicente, V., Poser, K., Bresciani, M., Alikas, K., Spyrakos, E., Giardino, C., Ansper, A., 2019. Assessment of atmospheric correction algorithms for the Sentinel-2A MultiSpectral Imager over coastal and inland waters. Remote Sens. Environ. 225, 267–289. https://doi.org/10.1016/j.rse.2019.03.018

Wei, J., Lee, Z., Shang, S., 2016. A system to measure the data quality of spectral remote-sensing reflectance of aquatic environments. J. Geophys. Res. Ocean. 121, 1063–1084. https://doi.org/10.1002/2015JC010796.Received

Wei, J., Lee, Z.P., Shang, S., Yu, X., 2019. Semianalytical derivation of phytoplankton, CDOM, and detritus absorption coefficients from the Landsat 8/OLI reflectance in coastal waters. J. Geophys. Res. Ocean. 124, 3682–3699. https://doi.org/10.1029/2019JC015125

Wentz, F.J., 2015. A 17-yr climate record of environmental parameters derived from the Tropical Rainfall Measuring Mission (TRMM) microwave imager. J. Clim. 28, 6882–6902. https://doi.org/10.1175/JCLI-D-15-0155.1

Werdell, P.J., Behrenfeld, M.J., Bontempi, P.S., Boss, E., Cairns, B., Davis, G.T., Franz, B.A., Gliese, U.B., Gorman, E.T., Hasekamp, O., Knobelspiesse, K.D., Mannino, A., Martins, J.V., McClain, C., Meister, G., Remer, L.A., 2019. The plankton, aerosol, cloud, ocean ecosystem mission status, science, advances. Bull. Am. Meteorol. Soc. 100, 1775–1794. https://doi.org/10.1175/BAMS-D-18-0056.1

Westberry, T., Behrenfeld, M.J., Siegel, D.A., Boss, E., 2008. Carbon-based primary productivity modeling with vertically resolved photoacclimation. Global Biogeochem. Cycles 22, 1–18. https://doi.org/10.1029/2007GB003078

Wolny, J.L., Tomlinson, M.C., Schollaert Uz, S., Egerton, T.A., McKay, J.R., Meredith, A., Reece, K.S., Scott, G.P., Stumpf, R.P., 2020. Current and Future Remote Sensing of Harmful Algal Blooms in the Chesapeake Bay to Support the Shellfish Industry. Front. Mar. Sci. 7, 1–16. https://doi.org/10.3389/fmars.2020.00337

Xi, H., Losa, S.N., Mangin, A., Soppa, M.A., Garnesson, P., Demaria, J., Liu, Y., d'Andon, O.H.F., Bracher, A., 2020. Global retrieval of phytoplankton functional types based on empirical orthogonal functions using CMEMS GlobColour merged products and further extension to OLCI data. Remote Sens. Environ. 240, 111704. https://doi.org/10.1016/j.rse.2020.111704

Xie, Y., Huang, B., Lin, L., Laws, E.A., Wang, L., Shang, S., Zhang, T., Dai, M., 2015. Photosynthetic parameters in the northern South China Sea in relation to phytoplankton community structure. J. Geophys. Res. Ocean. 120, 4187–4204.

Xing, X., Boss, E., 2021. Chlorophyll-Based Model to Estimate Underwater

Photosynthetically Available Radiation for Modeling, In-Situ, and Remote-Sensing Applications. Geophys. Res. Lett. 48. https://doi.org/10.1029/2020GL092189

Xue, K., Ma, R., Duan, H., Shen, M., Boss, E., Cao, Z., 2019. Inversion of inherent optical properties in optically complex waters using sentinel-3A/OLCI images: A case study using China's three largest freshwater lakes. Remote Sens. Environ. 225, 328–346. https://doi.org/10.1016/j.rse.2019.03.006

Yang, W., Matsushita, B., Chen, J., Fukushima, T., 2011. Estimating constituent concentrations in case II waters from MERIS satellite data by semi-analytical model optimizing and look-up tables. Remote Sens. Environ. 115, 1247–1259. https://doi.org/10.1016/j.rse.2011.01.007

Yang, W., Matsushita, B., Chen, J., Fukushima, T., Ma, R., 2010. An enhanced three-band index for estimating chlorophyll-a in turbid case-II waters: case studies of Lake Kasumigaura, Japan, and Lake Dianchi, China. IEEE Geosci. Remote Sens. Lett. 7, 655–659.

Yang, W., Matsushita, B., Chen, J., Yoshimura, K., Fukushima, T., 2013. Retrieval of inherent optical properties for turbid inland waters from remote-sensing reflectance. IEEE Trans. Geosci. Remote Sens. 51, 3761–3773. https://doi.org/10.1109/TGRS.2012.2220147

Yoo, S., Kong, C.E., Son, Y.B., Ishizaka, J., 2019. A critical re-assessment of the primary productivity of the Yellow Sea, East China Sea and Sea of Japan/East Sea Large Marine Ecosystems. Deep. Res. Part II Top. Stud. Oceanogr. 163, 6–15. https://doi.org/10.1016/j.dsr2.2018.05.021

Yoshikawa, T., Furuya, K., 2008. Phytoplankton photosynthetic parameters and primary production in Japan Sea and the East China Sea: Toward improving primary production models. Cont. Shelf Res. 28, 962–976. https://doi.org/10.1016/j.csr.2008.01.016

Yu, S.Y., Fan, L., Zhang, Y., Zheng, X.T., Li, Z., 2021. Reexamining the Indian summer monsoon rainfall–ENSO relationship from its recovery in the 21st century: role of the Indian Ocean SST anomaly associated with types of ENSO evolution. Geophys. Res. Lett. 48. https://doi.org/10.1029/2021GL092873

Yu, X., Lee, Z., Wei, J., Shang, S., 2019. Impacts of pure seawater absorption coefficient on remotely sensed inherent optical properties in oligotrophic waters. Opt. Express 27, 34974. https://doi.org/10.1364/oe.27.034974

Zhang, H., Nettleton, D., Zhu, Z., 2019. Regression-enhanced random forests. arXiv Prepr. arXiv1904.10416.

Zoffoli, M.L., Lee, Z., Marra, J.F., 2018. Regionalization and dynamic parameterization of quantum yield of photosynthesis to improve the ocean primary production estimates from remote sensing. Front. Mar. Sci. 5, 1–17. https://doi.org/10.3389/fmars.2018.00446