Pittman Nicholas, A. (Orcid ID: 0000-0002-4632-7223)
Strutton Peter, G. (Orcid ID: 0000-0002-2395-9471)
Johnson Robert (Orcid ID: 0000-0002-5915-5416)
Matear Richard, J. (Orcid ID: 0000-0002-3225-0800)

An assessment and improvement of satellite ocean color algorithms for the tropical Pacific Ocean

Nicholas A. Pittman[1,2], Peter G. Strutton[1,2], Robert Johnson[1], Richard J. Matear[3]

[1]Institute for Marine and Antarctic Studies, University of Tasmania, Hobart, Tasmania, Australia

[2]Australian Research Council Center of Excellence for Climate Extremes, University of Tasmania, Hobart, Tasmania, Australia

[3]Commonwealth Scientific and Industrial Research Organisation, Hobart, Tasmania, Australia

Corresponding author: Peter Strutton (peter.strutton@utas.edu.au) and Nicholas Pittman (nic.pittman@utas.edu.au)

Key Points:

- Satellite ocean color measurements underestimate chlorophyll concentrations in the tropical Pacific Ocean, and we removed the systematic bias.
- Modifying algorithm coefficients and blending window size between algorithms can improve overall ocean color estimates.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1029/2019JC015498

- Improved chlorophyll estimates have implications for historical and future Pacific carbon cycle and net primary productivity estimates.

Abstract

The tropical Pacific Ocean is a globally significant region of climate-driven biogeochemical variability. Satellite ocean color algorithms have been used for over 20 years, providing a substantial historical record of global ocean chlorophyll-a variability. Current chlorophyll algorithms perform better in the tropical Pacific than for the globe. Nevertheless, improvements can be made to produce a robust historical record of chlorophyll variability, which is essential to accurately identify ocean-atmosphere carbon fluxes and long-term trends in ocean productivity. We use a large in situ chlorophyll database to tune empirical ocean color algorithms to reduce bias in the equatorial Pacific. Traditional band ratio chlorophyll algorithms (OCx) perform adequately, but exhibit errors at low chlorophyll concentrations. A new algorithm, the ocean Color Index (OCI; *Hu et al.,* 2012), is more effective at calculating low chlorophyll concentrations in the mesotrophic tropical Pacific Ocean. Existing ocean color algorithms underestimate tropical Pacific chlorophyll by 5.8%, 14% and 2% for three satellite ocean color sensors: SeaWiFS, MODIS-Aqua and MERIS. In this paper, we develop regionally tuned sensor specific coefficients and blending windows between the OCI to OCx algorithms to reduce systematic biases in the tropical Pacific. We assess cross-sensor consistency to produce robust 21-year time series trends. These updated estimates increase chlorophyll concentrations in open water and decrease around island and warm-pool regions, with implications for our understanding of El Niño-Southern Oscillation driven carbon fluxes and net primary productivity.

Plain Language Summary

The tropical Pacific Ocean is a globally significant region where climate drives the variability of phytoplankton, nutrient availability and primary productivity. Satellite-based chlorophyll observations provide the most comprehensive large-scale estimate of phytoplankton abundance in the upper ocean. With over two decades of observations available from satellites, it is essential that these satellite measurements are as accurate as possible in order for us to document chlorophyll variability and trends accurately. In this paper, we evaluate the performance of satellite chlorophyll observations in the tropical Pacific Ocean and suggest algorithm improvements. Reduced errors in chlorophyll estimates will provide essential insights into critical processes like primary productivity and biologically driven CO_2 transport. No significant long-term trends are found for the tropical Pacific Ocean so far. However, decreased error in our estimates will allow future climate studies to more accurately describe phytoplankton variability in this region.

1 Introduction

The tropical Pacific Ocean spans one-third of Earth's circumference, represents 18% of global oceanic new production (*Barber et al.,* 1996) and is the origin of the most globally influential mode of climate variability; the El Niño-Southern Oscillation (ENSO; *McPhaden et al.,* 2006). The tropical Pacific is defined here as 10°N to 10°S and 150°E to 90°W. It is a dynamic and complex region integral to the global carbon cycle, because it is the most significant oceanic source of atmospheric CO_2 (*Feely et al.*, 2002; *McClain et al.,* 2002; *Takahashi et al.,* 1997). Physical changes in stratification, upwelling, trade winds and surface nutrient supply due to ENSO force perturbations in chlorophyll-*a* (chl in mg m^{-3}) concentrations, primary productivity and ecosystem dynamics (*Gierach et al.,* 2012). The phase of ENSO strongly influences tropical Pacific biological productivity and CO_2

outgassing, with both typically enhanced during La Niña and inhibited during El Niño (*Sutton et al.,* 2014). To understand future trajectories of ocean ecology, carbon sources and sinks, it is essential to further elucidate the response of the ocean system to large scale natural perturbations.

Developments in satellite ocean color sensors have significantly increased our understanding of the variability in ocean primary production, chl concentrations and ecosystem dynamics at global and regional scales (*McClain,* 2009). Global satellite chl products from NASA have typically used empirical band-ratio algorithms, called OCx. These are based on a polynomial of the spectral ratio of remote-sensing reflectance (R_{rs}; *O'Reilly et al.,* 1998). For SeaWiFS, the maximum band ratio (MBR) is the ratio of the maximum R_{rs} from the blue wavelengths ($\lambda 443 > 490 > 510$) to R_{rs} in the green wavelength ($\lambda 555$; Equation 1a). A fourth-order polynomial estimates Chl_{OCx} and is derived by matching in situ chl ($chl_{in\ situ}$) and MBR (Equation 1b).

Equation 1a:

$$\chi = \log_{10}(MBR); \text{ where MBR} = \frac{\max[R_{rs}(\lambda_{blue})]}{R_{rs}(\lambda_{green})}$$

Equation 1b:

$$\text{Chl}_{\text{OCx}} = 10^{a_0+a_1\chi+a_2\chi^2+a_3\chi^3+a_4\chi^4}$$

Current empirical algorithms for the Sea-viewing Wide Field-of-view Sensor (SeaWiFS, 1997-2010), MEdium Resolution Imaging Spectrometer (MERIS, 2002-2012) and MODerate resolution Imaging Spectroradiometer (MODIS-Aqua, 2002-present) were developed using a database of $chl_{in\ situ}$ ranging from oligotrophic subtropical gyres to highly productive coastal upwelling systems (*O'Reilly et al.,* 2000; Werdell and Bailey, 2005). An improved low chl (< 0.2mg m^{-3}) hybrid band difference algorithm, the Color Index (CI), was proposed by *Hu et al.* (2012). Chl_{CI} uses a linear regression model that compares $chl_{in\ situ}$ with the difference

between green R_{rs555} and a linear reference formed between blue R_{rs443} and red R_{rs670}. The CI calculation is presented in equation 2a. A linear, first-order polynomial is fit through the CI and taken to its power of ten to calculate chl_{CI} (Equation 2c). Figure 1 is an illustration of the relative magnitude of the relevant CI wavelengths and a description of the equation. The linear reference between blue and the red baseline creates a background from which the relative height of green (λ555) can derive chl more accurately at low concentrations (*Hu et al.,* 2012). Chl_{CI} may perform marginally better at low chl compared to chl_{OCx} by being more responsive to changes in chl specific backscatter and non-phytoplankton absorption (*Hu et al.,* 2019). Its major advantage is that Chl_{CI} is less sensitive to errors induced by instrument noise and imperfect corrections for the atmosphere, sun-glint or whitecaps, leading to cleaner imagery and improved cross-sensor consistency (*Hu et al.,* 2012).

Equation 2a:

$$\mathrm{CI} = \mathrm{R_{rs}}(\lambda_{\mathrm{green}}) - \frac{\mathrm{R_{rs}}(\lambda_{\mathrm{blue}}) + (\lambda_{\mathrm{green}} - \lambda_{\mathrm{blue}})}{(\lambda_{\mathrm{red}} - \lambda_{\mathrm{blue}}) \times (\mathrm{R_{rs}}(\lambda_{\mathrm{red}}) - \mathrm{R_{rs}}(\lambda_{\mathrm{blue}}))}$$

Equation 2c:

$$\mathrm{Chl_{CI}} = 10^{b_0 + b_1 \mathrm{CI}}$$

Chl_{CI} is merged with chl_{OCx} using an algorithm called the Ocean Color Index (OCI) which uses a blending window to best represent open ocean, case 1 waters (*Morel and Prieur,* 1977). Chl_{OCI} is calculated by using chl_{CI} at low values < 0.15mg m^{-3}, a linear blending (Equation 3a) with chl_{OCx} between the lower cutoff *l* and higher cutoff *h* (*l* = 0.15 and *h* = 0.2mg m^{-3} by default), with only chl_{OCx} used above the *h* cutoff (Equation 3b; *Hu et al.,* 2012). A single set of CI coefficients is provided for all NASA ocean color sensors (SeaWiFS, MERIS, MODIS-Aqua, VIIRS, OCTS, CZCS, OLI/Landsat 8). However, an

updated set of CI coefficients (OCI2) and blending window (0.25 to 0.4mg m^{-3}) was recently provided for the global ocean (*Hu et al.,* 2019). NASA produces and distributes a final global chl OCI product under the name chlor_a.

Equation 3a:

$$\alpha = \frac{\mathrm{Chl_{CI}} - \mathrm{l}}{\mathrm{h} - \mathrm{l}}, \beta = \frac{\mathrm{h} - \mathrm{Chl_{CI}}}{\mathrm{h} - \mathrm{l}}$$

Equation 3b:

$$\mathrm{Chl_{OCI}} = \mathrm{Chl_{CI}}[\text{for } \mathrm{Chl_{CI}} \leq \mathrm{l}]$$

$$\mathrm{Chl_{OCI}} = \mathrm{Chl_{OCx}}[\text{for } \mathrm{Chl_{CI}} > \mathrm{h}]$$

$$\mathrm{Chl_{OCI}} = \alpha \times \mathrm{Chl_{OCx}} + \beta \times \mathrm{Chl_{CI}}[\text{for } \mathrm{l} < \mathrm{Chl_{CI}} \leq \mathrm{h}]$$

Regional ocean color algorithms have typically focused on areas of complex optical properties and poor chl estimates. Numerous studies have assessed the accuracy of traditional OCx algorithms from global to regional scales, observing significant differences between ocean basins (*Gregg and Casey,* 2004; *O'Reilly et al.,* 2000; *Szeto et al.,* 2011). Regional OCx coefficients have been provided for basins including the Southern Ocean (*Garcia et al.,* 2005; *Johnson et al.,* 2013; *Szeto et al.,* 2011), Pacific and Atlantic Ocean (*Szeto et al.,* 2011) and the Mediterranean Sea (*Antoine et al.,* 2008; *Gómez Jakobsen et al.,* 2018; *Volpe et al.,* 2007). Compared to other ocean basins, the Pacific is well represented in global OCx parameterization datasets. Approximately one-third of the global $\mathrm{chl}_{\text{in situ}}$ was collected from this region and it generally lacks the optical complexities found in other basins (*Gregg and Casey,* 2004). Nonetheless, it continues to exhibit a systematic underestimation of about 15% (*Szeto et al.,* 2011). This underestimation illustrates the need to examine the scope for regional improvements in areas where global coefficients have historically performed relatively well. Even where algorithms do perform well, minor improvements may enhance the accuracy of long-term trends and internal variability.

In this paper, we assess current satellite ocean color algorithms for the tropical Pacific. We remove systematic bias, improve the accuracy and dynamic range of chl estimates in the region and improve cross sensor consistency. Ultimately, this will improve our understanding of the role of ENSO and climate change in biogeochemical variability and trends. This manuscript has the following goals:

1. Evaluate the accuracy of current satellite chl algorithms for SeaWiFS, MODIS-Aqua and MERIS in the tropical Pacific.
2. Improve existing chl algorithms to better represent the mesotrophic tropical Pacific.
3. Identify factors contributing towards the most accurate chl algorithm retrievals in this region.
4. Quantify the impact of algorithm choice on the detection of chl variability and long-term change.

2 Materials and Methods

2.1 Satellite data

SeaWiFS, MODIS-Aqua and MERIS data were downloaded (10-25th September 2018) from the NASA ocean color website (https://oceandata.sci.gsfc.nasa.gov/). For each sensor, daily, 9km, level 3 mapped (L3m) chl data were obtained for the chl_{OCx} and chl_{OCI} (chlor_a) products. R_{rs} data at the same resolution for the relevant wavelengths for each sensor were also downloaded (Table 1).

2.2 In situ data

The largest tropical Pacific chl database to date was compiled from three data sources, with a total of 4641 surface $chl_{in\ situ}$ observations in the region 10°N to 10°S, 150°E to 90°W. The spatial distribution of $chl_{in\ situ}$ matchups and sensor specific satellite climatologies are

presented in Figure 2. These observations were averaged over the upper 20m where CTD profiles of depth-resolved chl exist. This represents the approximate first optical depth, the depth observed by satellite ocean color sensors. This calculation of the first optical depth (*Gordon and McCluney,* 1975), was based on a MODIS-Aqua mission average diffuse attenuation coefficient at 490nm (Kd_{490}) of approximately 0.05 m^{-1} in the tropical Pacific. We did not weight the measurements in the upper 20m according to Kd_{490} because chlorophyll in the tropical Pacific is usually uniform over the upper 20m (See Fig 4 in Strutton and Chavez, 2000), and the 20m criterion was used mainly to make sure that we did not miss the shallowest chlorophyll sample, which may be at around 10m. $Chl_{in\ situ}$ was measured either fluorometrically (*Chavez et al.,* 1995) or from high performance liquid chromatography (*Mock and Hoch,* 2005). There was a temporal bias, with more observations in the first half of the record for all sensors (Figure 3; a, b, c). This non-uniformity was variable before 2005 with a significant drop in observations after 2008, however no seasonal bias was present. Chl was log-normal and positively skewed, with a peak at the data mean around 0.15mg m^{-3} (Figure 3; d, e, f). The data was sourced from several databases, requiring duplicate observations to be averaged due to rounding inconsistencies. For each observation, latitude and longitude were rounded to the nearest 0.1° (10km). The mean chl was stored for any observations in the same location and day. This 10km rounding was approximately the same resolution as the 9km satellite pixels. The data sources compiled in this database are described below.

2.2.1 Tropical Atmosphere Ocean (TAO) mooring maintenance cruises

The TAO-TRITON array is a network of moorings across the tropical Pacific that monitor El Niño. During the late 1990s and 2000s, the NOAA ship Ka'imimoana, and occasionally other NOAA vessels, maintained the US portion of TAO-TRITON. This period corresponded with

sustained satellite ocean color observations from SeaWiFS, MODIS-Aqua and MERIS. During mooring maintenance operations, ships spanned the entire basin twice yearly from the Galapagos to the dateline. CTDs were deployed at every degree of latitude between 8°N and 8°S, and every 0.5° between 2°N and 2°S along meridional mooring lines, resulting in approximately 300 CTDs per year. This portion of the dataset contained 1377 surface $chl_{in\ situ}$ samples for potential satellite matchups from the beginning of SeaWiFS (4 October 1997) to 9 August 2004. This is the same dataset used by *Strutton et al.* (2008), provided by Francisco Chavez, Monterey Bay Aquarium Research Institute.

2.2.2 Pangaea global bio-optical in situ data

An additional 2224 tropical Pacific surface $chl_{in\ situ}$ observations were sourced from *Valente et al.* (2016; https://doi.org/10.1594/PANGAEA.854832), with 1640 fluorometric readings (7 October 1997-27 December 2009) and 584 HPLC observations (10 December 1997 - 4 September 2010). HPLC observations were used in preference over coincidental fluorometric observations as they were of higher quality (*O'Reilly et al.,* 1998). If more than one HPLC observation was present in a 0.1° box on the same day, $chl_{in\ situ}$ values were averaged. The inclusion of HPLC observations allowed the final $chl_{in\ situ}$ matchup dataset to be as large and representative as possible.

2.2.3 World Ocean Database

A further 1040 surface fluorometric observations were accessed from the Word Ocean Database 2018 (WOD; *Boyer et al.,* 2018; https://www.nodc.noaa.gov/OC5/WOD/pr_wod.html). Only Ocean Station Data samples have been used (8 October 1997-24 June 2016), where water samples were collected and analyzed fluorometrically. We omitted CTD, float and glider in vivo fluorometric

observations. As above, WOD data were averaged over the upper 20m. The eastern boundary was limited to 100°W rather than 90°W to remove very high values (~35mg m^{-3}) around the Galapagos Islands and Papagayo upwelling system.

2.3 Model performance metrics

The performance of ocean color algorithms in estimating $chl_{in\ situ}$ is sensitive to the performance metrics used to rank them. Using recommendations from *Seegers et al.* (2018), four diagnostics have been used here for model evaluation. In our description of methods, we use the terms 'model' and 'algorithm' to refer to the equation or system of equations that convert R_{rs} to chl. The notation median (M-O) means the median difference between modelled (satellite) chl and observed $chl_{in\ situ}$.

2.3.1 Pairwise comparison

Percent wins, also known as point-by-point accuracy or pairwise comparison, was calculated for new models as compared to the current NASA chlor_a product for each satellite. For each matchup, model chl and $chl_{in\ situ}$ were differenced, and the model with the smallest residual won that matchup (residual = model - observation). The number of wins per model was tabulated, with the total percent wins calculated for each model (*Seegers et al.,* 2018).

2.3.2 Bias and absolute error

Bias and absolute error are useful diagnostics to identify systematic over or underestimation and overall error. Log-transformed diagnostics convert metrics from linear to multiplicative space; for example, a bias of 0.8 means an underestimation of 20% (*Seegers et al.,* 2018). Here, median was the primary diagnostic. It is more robust than mean when chl is positively skewed, non-Gaussian and log-normal, as in this case (Figure 3; d, e, f). The log-transformed

median bias distinguishes overall systematic bias (bias; Equation 4). The median absolute error (MAE; Equation 5) is the absolute value of the bias. Absolute error (AE; Equation 6) was also used as a single observation diagnostic to identify outliers. Bias, MAE and percent wins were the primary diagnostics used to assess algorithm performance.

Equation 4:

$$\text{median bias} = 10^{\text{median}\left(\log_{10}(M_i)-\log_{10}(O_i)\right)}$$

Equation 5:

$$\text{median AE} = 10^{\text{median}}$$

Equation 6:

$$\text{AE}_\text{i} = 10^{\log_{10}(M_i)-\log_{10}(O_i)\vee}$$

2.3.3 Traditional diagnostics

The slope, intercept and coefficient of determination (R^2) of satellite vs $chl_{in\ situ}$ were recorded, despite these statistics being easily misinterpreted due to systematic biases in poorly performing models (*Seegers et al.,* 2018). The optimized model slope was sometimes confounded by systematic low and high chl bias in the models. Despite this, these traditional diagnostics proved effective when used in conjunction with the multi-metric approach of bias, MAE, pairwise comparison metrics and visual identification to identify optimal models.

2.3.4 Moving MAE Window

A moving MAE window was calculated to assess empirical and blended model performance and sensitivity through chl space. The MAE was calculated in 0.1mg m^{-3} bins for each sensor and empirical model estimate. This diagnostic presents the median error at different chl concentrations.

2.4 Satellite matchups and climate indices

To evaluate and develop satellite algorithms, it is necessary to identify coincident $chl_{in\ situ}$ and satellite observations. Temporal and spatial averaging helps to increase the number of matchups that might otherwise be lost due to gaps in the satellite data. Clouds, or the space between overpasses can cause these gaps. In order to identify the ideal temporal and spatial averaging window for matchups, a range of combinations were evaluated. In space, windows of 1, 3x3 and 5x5 pixels were evaluated, with the $chl_{in\ situ}$ in the center pixel of each grid. In time, averaging periods of 1, 3 (±1 day) and 5 days (±2 days) were evaluated centered around the time of the $chl_{in\ situ}$ measurement. For each combination assessed, the total number and percent of matches were recorded (Table 2). The slope, R^2, bias and MAE of satellite vs $chl_{in\ situ}$ was calculated for each combination. Different spatial and temporal averages result in a tradeoff between the number and quality of successful matchups. Any pixel averaging window with a standard deviation > 0.05mg m^{-3} calculated from the valid pixels was flagged as an outlier and removed to ensure spatial uniformity. For SeaWiFS, 30 matchups were rejected on this criterion. Similarly, if the AE of a match compared to $chl_{in\ situ}$ was above the 68% confidence interval (~> 3mg m^{-3}), it was rejected. This is a stricter outlier criterion than the 95% confidence interval used by *Gómez Jakobsen et al.* (2018). For SeaWiFS, this resulted in an additional 17 matchups being rejected as outliers.

For the optimal matchup criteria, R_{rs} values described in Table 1 were stored alongside the NASA chlor_a product. A database was created with $chl_{in\ situ}$, R_{rs}, NASA chlor_a estimates, date, latitude, longitude and MEI (Multivariate ENSO Index). The monthly MEI was obtained from https://www.esrl.noaa.gov/psd/enso/mei/ as an indicator for ENSO state

(Accessed: December 2018). ENSO here has been split into its three phases using MEI; where El Niño is defined as MEI >= 1, La Niña is MEI <= -1 and Neutral is -1 < MEI < 1.

2.5 Independent evaluation

The matchups for each sensor were randomly split into two groups: training and validation. Each dataset had the same number of observations, similar distributions in location, chl concentration and over time (Figure 3). The training datasets were used to train new models during the optimization process. The validation dataset was then used to evaluate model performance. This cross-validation methodology ensures that overfitting of models to training $chl_{in\ situ}$ can be identified and prevented by inferior validation diagnostics.

2.6 Creating new models

A 'least squares' minimization routine, similar to the optimization used by *Johnson et al.* (2013) was used to tune our OCx algorithms. The Python module *scipy.optimise.least_squares* was used for this purpose. Log_{10}(MBR), denoted χ, and $chl_{in\ situ}$ from the training dataset were fed into the algorithm with starting polynomial coefficients, from either the original NASA OCx or the Szeto Pacific polynomial (Table 1). Model chl was calculated and a linear regression was performed on the resulting chl_{OCx} estimates against $chl_{in\ situ}$. The slope, intercept and R^2 were returned to the optimizer in order to improve the OCx polynomials by iterating for perfect diagnostics (slope=1, intercept=0, R^2=1). The MAE, bias and percent wins were calculated and used in some optimization runs in place of the default statistics slope, intercept and R^2. The default blending window of 0.15 to 0.2mg m^{-3} was kept during optimization which only optimized chl_{OCx} for values > 0.15mg m^{-3}. By including the default blending window, chl_{CI} estimates below the lower cutoff were retained. This routine was repeated for each sensor until no more improvements could be made to the

polynomials. A linear regression through the training $chl_{in\ situ}$ to CI improved the CI coefficients and proved sufficient for use here.

2.7 Selecting models

After several OCx and CI coefficients were created using the optimization method on training data for each sensor, the OCI blending window was assessed for its impact on model performance. Using a model combination iterator, we cycled over four sets of variables; 1) OCx polynomial coefficients, 2) CI polynomial coefficients, 3) upper OCI blending cutoffs and, 4) lower OCI blending cutoffs. Seven OCx polynomials were assessed. These were the NASA algorithms for SeaWiFS, MODIS-Aqua and MERIS, the *Szeto et al.* (2011) Pacific Ocean algorithm, an optimized third-order polynomial and two optimized fourth-order polynomials. Six CI first-order polynomials were assessed. These were the original *Hu et al.* (2012), the updated *Hu et al.* (2019) and the two separate polynomials that best fit the training data for each of the three sensors. Finally, 29 different OCI blending window combinations (values of *l* and *h*) were assessed: 0mg m^{-3} to 0.5 at 0.05 increments, 0 to 0.6 at 0.1 increments, 0 to 0.6 at 0.15 increments, 0 to 0.6 at 0.2 increments, 0 to 1 at 0.5 increments, 0 to 0, 0 to 1, 1 to 2 and 0.25 to 0.4mg m^{-3}. The 0 to 0 and 1 to 2mg m^{-3} cases were used to simulate chl_{OCx} and chl_{CI} only combinations. In total, this experiment produced 1218 different algorithm combinations for each sensor.

Final models were ranked by percent wins, but all other metrics (bias, MAE, R^2 and slope) were tracked. The model with the most percent wins and bias closest to 1 was selected for our proposed tropical Pacific chlorophyll algorithm (TPCA). The TPCA model for each sensor was applied to the full sensor record for analysis of variability and trends, and ultimately for dissemination to potential users.

2.8 Cross-sensor consistency

SeaWiFS, MODIS-Aqua and MERIS overlap between 2002 to 2010. To compare algorithms, climatologies were created for each sensor for this period, using both the NASA and TPCA algorithms. Sensor-to-sensor bias (Equation 4), mean chl and percentage difference was calculated. All sensor combinations were assessed to report how algorithm choice impacts long-term cross sensor consistency.

2.9 Time series and spatial analysis

The selected models were applied to the entire time series, creating TPCA data products for SeaWiFS and MODIS-Aqua. Daily files were averaged through time into sensor climatologies, and through space to provide sensor time series. The climatologies were split into El Niño, La Niña and neutral phases as discussed in section 2.4. These climatologies were differenced between the NASA and TPCA implementations to identify spatial differences between the algorithms dependent on ENSO mode. The basin was split into three portions, the western (165°-180°E, ±10°N-S), central (170°-155°W, ±10°N-S) and eastern (140°-115°W, ±10°N-S), similar to *Strutton et al.* (2008). A linear regression was applied to these regional time series to identify how the TPCA algorithm impacts long-term trends through space and time.

3 Results

3.1 In situ to satellite matchups

Our $chl_{in\ situ}$ data ranged from 0.005mg m^{-3} to 3.867mg m^{-3}, with a median of 0.132mg m^{-3} and a mean of 0.154mg m^{-3}. Many of the largest $chl_{in\ situ}$ values were located on the eastern side of the basin in the Gulfs of Tehuantepec and Papagayo, near the Galapagos Islands or in

the Peruvian upwelling system (Figure 2). Table 2 presents the number of matchups for each temporal and spatial bin size assessed. The matchup window of 5 (±2) day and 9 (3x3) pixel was selected for its balance between high match rates of 68.8%, 64.8% and 72.1% for SeaWiFS, MODIS-Aqua and MERIS respectively, with favorable bias, MAE, slope and R^2. The spatial decorrelation timescale of chl in the central tropical Pacific is 2 days (*Strutton and Chavez,* 2003), consistent with the ±2 day matchup window. This tradeoff between the number of matches and performance was selected because the diagnostics were comparable to those for the ideal 1-pixel 1-day scenario, which had < 20% matches. SeaWiFS had the most matchups of 2551 with 151 outliers = 2400 viable matches, followed by MODIS-Aqua 948 with 48 outliers = 900 and MERIS 928 matches with 38 outliers = 890. All three sensors' default NASA algorithms systematically underestimated $chl_{in\ situ}$. MERIS had the best overall performance with a median bias of 0.973, a 2.7% underestimation compared to SeaWiFS' bias of 0.92 (an underestimation of 7.6%), and MODIS-Aqua which underestimated by 10.6%, particularly at high chl concentrations (> 0.3mg m^{-3}; Figure 4). The slopes for satellite vs $chl_{in\ situ}$ were 0.54, 0.56 and 0.59 for SeaWiFS, MODIS-Aqua and MERIS respectively. The R^2 for each of the sensors was 0.76. 0.80 and 0.81 respectively, meaning that the models generally capture variance well.

3.2 Developing new models

By using the training dataset to optimize bias and % wins, one third-order and two forth-order OCx polynomial coefficients were produced for each sensor. Updated OCx and CI polynomial coefficients and blending windows were compared to identify optimal algorithms for each sensor. The total number of combinations of OCx and CI polynomials, lower and upper OCI blending cutoffs was 1218. Of these, 428 SeaWiFS, 956 MODIS-Aqua and 44 MERIS models showed improvements over the existing NASA algorithms. Table 3 outlines

10 of the best models for each sensor. The selected models (most wins and bias closest to 1) for the TPCA are highlighted in bold. The OCI window has more influence on model performance than CI or OCx polynomial coefficients. The best blending window for MODIS-Aqua and MERIS was 0 to 0.2mg m^{-3} and 0 to 0.5mg m^{-3} for SeaWiFS.

Sensor specific OCx polynomials for $chl_{in\ situ}$ as a function of MBR are presented in Figure 5. Values below the default OCI blending window cutoff of 0.15mg m^{-3} are semi-transparent gray crosses. The underestimation of $chl_{in\ situ}$ by the original algorithms is represented by most of our optimized estimates (red lines) falling above the traditional algorithm (blue lines) in Figure 5. On average between a MBR of 2 to 4, approximately equal to 0.5mg m^{-3} to 0.15mg m^{-3}, the new OCx algorithms increase estimates by 0.033mg m^{-3}, 0.039mg m^{-3} and 0.015mg m^{-3} for SeaWiFS, MODIS-Aqua and MERIS. The *Szeto et al.* (2011) Pacific polynomial was reasonably effective for both SeaWiFS and MODIS-Aqua but not MERIS. The current NASA MODIS-Aqua polynomial is the worst default performer of the three sensors, while MERIS retrieves the most accurate estimates.

Figure 6 presents the CI linear estimates for the original OCI, updated OCI2 and our best fit regression through the training data. Our regression coefficients are most similar to the original *Hu et al.* (2012) version for SeaWiFS and MODIS-Aqua. The original CI coefficients resulted in slight underestimation for all sensors, with more data points located above than below the blue *Hu et al.* (2012) line. MERIS' best fit is closest to the updated *Hu et al.* (2019) coefficients; however the CI for MERIS is the least useful of any sensor (Table 3). The OCI linear fit was among the best performing CI coefficients for all three sensors during the model combination iterator. The OCI coefficients typically outperformed both the updated OCI2 and our best fit coefficients.

3.3 OCI blending window

The chl_{CI} to chl_{OCx} OCI blending window had the largest influence on algorithm performance. Figure 7a presents the spectrum of 1218 model combinations of percent wins verses overall bias for SeaWiFS (black), MODIS-Aqua (blue) and MERIS (orange). The selected models (large solid circles) were chosen based on a combination of least bias and highest percent wins. For SeaWiFS, this was the 2nd ranked percent wins model, for MERIS it was the 5th ranked, however for MODIS, the 21st model was selected by these criteria (Table 3). For MODIS-Aqua, high percent wins did not correspond with improved bias and error statistics. The highest percent wins for MODIS-Aqua was 65.6%, but this algorithm had a median bias of 0.909, only slightly better than the existing NASA median bias of 0.860. For this reason, it was not selected as the best new MODIS-Aqua algorithm in preference for better bias values offered by the 21st model. MODIS-Aqua has more models with high percent wins and more options to select a final model, making it the easiest sensor to improve. However, many of these models were skewed by negative bias (Figure 7a). The NASA MERIS algorithm was the best performing sensor, but many model combinations are associated with the largest median biases and fewest % wins (Figure 7a).

The optimized and blended models from the training data were not always the most effective on the validation dataset. Overfit models were recognized by poor validation diagnostics and were discarded. Ten of the best models for each sensor are shown in Table 3. Selected TPCA models are highlighted in bold and presented in detail in Table 4. In all cases for the tropical Pacific, the default OCI window of 0.15 to 0.2mg m^{-3} performed poorly (Table 3). For MODIS-Aqua, the default 0.15 to 0.2mg m^{-3} blending window was the worst performer of the 29 window combinations assessed. The optimal SeaWiFS combination used a large OCI

blending window of 0 to 0.5mg m^{-3}, the default MERIS polynomial and the *Hu et al.* (2012) CI. The optimal MERIS model used an optimized OCx polynomial, the *Hu et al.* (2012) CI and a blending of 0 to 0.2mg m^{-3}. The best percent wins MODIS-Aqua models had large OCI windows, but remain highly biased (blue dots towards the bottom right in Figure 7a). The selected MODIS-Aqua model was the 21st ranked in percent wins and used the SeaWiFS OCx, the *Hu et al.* (2012) CI and an OCI blending of 0 to 0.2mg m^{-3} (Table 4). TPCA models for SeaWiFS and MODIS-Aqua had a binomial probability of $< 10^{-6}$, rejecting the null hypothesis that these percent wins could happen by chance. MERIS had the highest yet still insignificant odds of reaching these probability levels by chance, with a binomial probability of 6×10^{-6}, due to the small sample size of 890.

MAE was calculated over 0.1mg m^{-3} moving windows of chl values for the most improved empirical model estimates for each sensor (Figure 7; b, c, d). A MAE minimum was observed near the whole-of-dataset mean of 0.15mg m^{-3}. This minimum was associated with a crossover in MAE from chl_{CI} (orange) estimates to the TPCA chl_{OCx} polynomial estimates (green). CI estimates perform relatively poorly above 0.3mg m^{-3} for all sensors (orange). The most significant improvements between old and new algorithms are between current NASA chl_{OCx} (blue) and TPCA chl_{OCx} estimates (green). Our new blending window and algorithm (black) is improved through most of chl space compared to the current global implementation (dotted black). All empirical estimates for MERIS perform similarly for chl < 0.25mg m^{-3} with less CI performance gains than for either SeaWiFS or MODIS. Algorithms selected from the model combination iterator removed overall bias and decreased MAE. However, only minor improvements were made through chl space, especially for MERIS (Figure 7; b, c, d).

Selected TPCA algorithms, diagnostics, coefficients and blending windows are presented in Table 4. The existing NASA MERIS algorithm had excellent validation diagnostics, subjectively better than that of the selected TPCA model. The bias of the existing MERIS algorithm was 1.003, and the TPCA retrieved an excellent but inferior bias of 1.007. The updated MERIS algorithm showed improvements and 56.9% wins over NASA. However, these improvements were minor, primarily through chl space, seen in Figure 7d. The use of chl_{CI} for MERIS was significantly less effective than the other sensors, and chl_{OCx} performed well. For these reasons and to retain the best algorithm diagnostics and cross-sensor consistency, we have chosen to discard our solution and retain the existing NASA MERIS algorithm for use in the TPCA.

Figure 8 shows the final selected models (closed circles) applied to the validation dataset compared to the NASA implementation (open circles). The solid black line is the new slope, the thin dotted line is the old slope, and the dashed $x=y$ is the ideal fit. On the validation dataset, we improve the slope from 0.53 to 0.54 for SeaWiFS and 0.56 to 0.70 for MODIS-Aqua. MERIS remains with a slope of 0.61. MODIS-Aqua shows the greatest improvements. The increased slope means that updated chl estimates are slightly higher for all sensors than the current implementation, thus reducing systematic bias.

3.4 Cross-sensor consistency

NASA and TPCA algorithms were compared for SeaWiFS, MODIS-Aqua and MERIS, but for MERIS, TCPA is the same as the existing NASA chl implementation (see above). Comparing NASA SeaWiFS to MODIS-Aqua results in a bias of 1.031, which means that SeaWiFS overestimates MODIS-Aqua by 0.0048mg m^{-3}. The new TPCA SeaWiFS to MODIS-Aqua bias is 0.958, meaning MODIS-Aqua now overestimates SeaWiFS by

0.0089mg m-3. MERIS overestimates NASA SeaWiFS and MODIS-Aqua by 0.0082mg m^{-3} (bias=0.950) and 0.0129mg m^{-3} (bias=0.921) respectively. MERIS now overestimates SeaWiFS by 0.0036mg m^{-3} (bias=0.982) and MODIS-Aqua overestimates MERIS by 0.0051mg m^{-3} (bias=1.026). That is, the TPCA algorithm has improved the consistency between SeaWiFS and MODIS-Aqua to MERIS, however not to each other. These biases change through time and space, but they are smaller than most in-situ measurement uncertainties and detection limits. The largest discrepancies occurred during the 2010-2011 La Niña. The eastern portion of the tropical Pacific basin presents more significant cross-sensor differences than the central or western regions.

3.5 Spatial Distribution

Spatial and temporal variability in the satellite chl underestimation might lead to differences in the expression of ENSO-induced biological patterns. Figure 9 shows the tropical Pacific chl distribution for (a) El Niño, (b) La Niña and (c) neutral conditions for our new TPCA SeaWiFS algorithm. The SeaWiFS TPCA to NASA difference for the three ENSO phases is presented in Figure 9(d), (e) and (f). SeaWiFS is used here as an example, and a positive anomaly indicates that the TPCA chl estimate was higher than the original NASA algorithm. There is very little difference across ENSO phases, except perhaps for the MODIS representation of El Niño, where the difference between the TPCA and the original NASA algorithm was smaller (9.5%) than for La Niña (13.1%) and neutral (11.0%) conditions.

3.6 Time series analysis

The full 21-year, sensor merged ocean color time series is presented in Figure 10, for the current NASA algorithm (Figure 10; a, b, c) and the new TPCA (with NASA MERIS; Figure 10; d, e, f). Results are calculated for the western (a, d; 165°-180°E, ±10°N-S), central (b, e;

170°-155°W, ±10°N-S) and eastern (c, f; 140°-115°W, ±10°N-S) boxes in the tropical Pacific, similar to that of *Strutton et al.* (2008). The averages here were lower than the $chl_{in\ situ}$ basin average of 0.15mg m^{-3} as the eastern box was limited to 115°W, omitting the high chl values in the Gulfs of Tehuantepec and Papagayo and around the Galapagos Islands. There was a small but statistically significant decrease in chl over time for all regions and algorithms considered when the three sensors were combined. For NASA standard algorithms, we found % change trends and average chl concentrations of -0.79% $decade^{-1}$ (0.081mg m^{-3}), -1.30% $decade^{-1}$ (0.094mg m^{-3}) and -1.14% $decade^{-1}$ (0.123mg m^{-3}) for the western, central and eastern Pacific, respectively. The average NASA trend in these boxes was -1.07% $decade^{-1}$. Our new combined TPCA processing retrieves trends half that of the global NASA algorithms, with an overall tropical Pacific chl trend of -0.49% $decade^{-1}$ since 1997. Regionally, these chl trends and averages are -0.40% $decade^{-1}$ (0.084mg m^{-3}), -0.73% $decade^{-1}$ (0.098mg m^{-3}) and -0.33% $decade^{-1}$ (0.131mg m^{-3}) for the western, central and eastern Pacific respectively. The NASA basin averages (150°E to 90°W, ±10°N-S) are 0.154mg m^{-3} for SeaWiFS, 0.142mg m^{-3} for MODIS-Aqua and 0.153mg m^{-3} for MERIS. With TPCA, these values increased to 0.159mg m^{-3} for SeaWiFS and 0.160mg m^{-3} for MODIS-Aqua. The NASA MERIS algorithm average of 0.153mg m^{-3} was retained. These values are slightly higher than the tropical Pacific mean of 0.154mg m^{-3}. However, this is expected due to the increased sampling rate and resolution of observations.

4 Discussion

Satellite ocean color sensors performed better in the tropical Pacific Ocean than the global average. However, the chl concentrations in the training dataset were underestimated by 5.8%, 14%, and 2% for SeaWiFS, MODIS, and MERIS respectively (Table 2, Figure 5). A regionally tailored implementation of the OCI algorithm is essential to correct this

underestimation and to accurately resolve low chl concentrations. We developed updated empirical algorithms for this region which reduced the systematic underestimation to 1.4% and 0% for SeaWiFS and MODIS, but retained the default MERIS algorithm with a validation matchups overestimation of 0.3%. The MAE for chl was reduced, and the new algorithms outperformed NASA when compared point-for-point (56.5% and 56.9% wins respectively; Figure 7a and Table 3). The blending window size and bounds were more critical than the CI and OCx polynomial coefficients. The removal of systematic bias in this physically complex region has implications for understanding variability in productivity and carbon flux due to ENSO and climate change. The global accuracy goal for NASA's chl algorithms is 35% (*Bailey and Werdell,* 2006; *Hooker et al.,* 1992). Here, we have improved the median accuracy from 28.4% to 27% for SeaWiFS and 30.9% to 26.3% for MODIS-Aqua (Table 2). Most importantly, we reduced the validation dataset bias from -5.8% and -14% for SeaWiFS and MODIS-Aqua, to less than ±2% for all sensors, effectively eliminating the underestimation problem in existing tropical Pacific chl estimates.

4.1 OCI blending window size is essential

The most significant driver of OCI algorithm performance was the $chl_{CI\ to}$ chl_{OCx} blending window in chl space. The moving MAE is a useful diagnostic to compare chl_{OCx}, chl_{CI} and chl_{OCI} model performance through chl space (Figure 7; b, c, d). Modifying the coefficients for chl_{CI} and chl_{OCx} models had a smaller impact on algorithm accuracy compared to adjusting the blending window size. There is strong evidence that merged modelling approaches, using the best of several algorithms to complement each other, results in the highest retrieval accuracy over the whole distribution (*Hu et al.,* 2012, 2019; *Wang and Son,* 2016). Most regional chl algorithms report overall performance, rather than effectiveness across a range of chl values (*Cota,* 2004; *Johnson et al.,* 2013; *Szeto et al.,* 2011), but algorithm performance

can be more subtle than a single diagnostic, and may depend on chl distribution. All models assessed here performed best between 0.1 and 0.2mg m^{-3}, around the mean tropical Pacific chl value of 0.15mg m^{-3}. This error minimum is associated with a crossover between chl_{CI} and chl_{OCx} model performance for all sensors except MERIS. The difference in chl_{OCx} and chl_{CI} performance for MERIS was much lower than for either of the NASA sensors, making it the most difficult to improve (Figure 7d). For SeaWiFS, chl_{CI} performed better than chl_{OCx} below the crossover of 0.18mg m^{-3}, and chl_{OCx} outperformed chl_{CI} above the crossover. This crossover location appears ideal for the original blending window of 0.15 to 0.2mg m^{-3}. However, the original small blending window in conjunction with our improved OCx performed substantially worse than the optimized large 0 to 0.5mg m^{-3} blending window. For example, the red dotted line in Figure 7b shows the 0.15 to 0.2mg m^{-3} blending window, where the new OCx polynomial had a larger error in the crossover range. Larger blending windows allow for a more gradual transition of weighting from chl_{CI} to chl_{OCx}. The lower limit of 0 mg m^{-3} is likely important because it mitigates the sensitivity of the OCx algorithm. That is, the shape of the 4^{th} order OCx makes it very sensitive to small changes in the MBR at low chl. In contrast, the CI is linear so it exhibits reduced sensitivity in this same way. For the tropical Pacific, where much of the ocean is between 0 and 0.2 mg m^{-3}, the optimum lies somewhere between the CI and OCx, hence the success of the blending at these values.

4.2 Polynomial evaluation

The original *Hu et al.* (2012) CI coefficients performed better for all sensors in the tropical Pacific than the updated *Hu et al.* (2019). Larger blending windows of 0 to 0.2mg m^{-3} for MODIS-Aqua and MERIS and 0 to 0.5mg m^{-3} for SeaWiFS provided optimal chl retrievals in the tropical Pacific. These larger blending windows indicate that neither chl_{CI} or chl_{OCx} models truly describe the R_{rs} to $chl_{in\ situ}$ relationship, even at very low values. The CI method

performs poorly for MERIS, as shown by a lack of crossover in Figure 7d. Our best fit line for MERIS CI was closer to the updated *Hu et al.* (2019) coefficients, however due to this poor performance, the existing NASA MERIS algorithm was retained. The *Johnson et al.* (2013) optimization routine, optimizing for percent wins and bias, had little improvement in chl_{OCx} accuracy here. Optimizing for slope could retrieve a slope of 1, but resulted in poor error diagnostics and increased bias. The coefficient optimization method was not as successful in the tropical Pacific as it was in the Southern Ocean (*Johnson et al.,* 2013), likely due to the added complication of the OCI blending method, low mean chl concentrations and small range of chl concentrations. Of the proposed TPCA algorithms, all use NASA polynomials, however SeaWiFS uses MERIS and MODIS-Aqua uses SeaWiFS' OCx coefficients respectively. These coefficient swaps are due to our changes in blending window size. This indicates that both CI and OCx polynomials which best fit training data are not a major factor in overall algorithm skill. The *Szeto et al.* (2011) Pacific polynomial performed relatively well. It featured among the top results for both SeaWiFS and MODIS-Aqua but not for MERIS (Table 3). The *Szeto et al.* (2011) model was trained on the entire Pacific basin and is not constrained to the equatorial region used here. Despite performing well, ranking #1 for MODIS-Aqua and #13 for SeaWiFS percent wins, these models were heavily biased compared to the selected winners (Table 3). *Szeto et al.* (2011) may be successfully applied to the tropical Pacific when used alone and at relatively high chl concentrations, but it is not suitable when used with the OCI blending algorithm. Hence, it appears that we are reaching the effective limit of improvement for existing empirical algorithms.

4.3 Possible blending method improvements

A possible approach to improve the blending window is to base it on reflectances. This method of model blending based on the ratio between blue $R_{rs}(443)$ and green $R_{rs}(551)$, or

nearest wavelengths was developed to remove transition discontinuities for the Visible Infrared Imaging Radiometer Suite (VIIRS; *Wang and Son,* 2016). This method was structured around earlier reflectance ratio based merging approaches for $K_d(490)$ and K_d(PAR; *Wang et al.,* 2009). The MBR blending window of 2.0 < MBR < 4.0 proposed by *Wang and Son* (2016) for the global oceans barely intersected the MBR in the tropical Pacific. We attempted a MBR blending where chl_{OCx} was used for MBR < 3.5, chl_{CI} used for MBR > 4, and a linear blending was used between 3.5 and 4.0. These reflectance blended models performed better than the original NASA algorithm, but did not improve our tuned empirical models, which essentially chose the blending window through trial and error. None of the reflectance-based blending methods assessed were improved over our proposed TPCA. Several studies have used artificial neural networks to derive ocean color products (*Gross et al.,* 1999; *Ioannou et al.,* 2013; *Sauzède et al.,* 2015). These evolving techniques may provide new methods of chl retrieval from R_{rs} or other geophysical variables, or potentially create non-linear blending methods for chl_{CI} and chl_{OCx} estimates.

4.4 Cross-sensor consistency

Extending the ocean chl record by combining multiple ocean color sensors is difficult due to discontinuities introduced by different space agencies, calibration strategies, wavelengths and bandwidths, atmospheric correction and data processing (*Djavidnia et al.,* 2010; *Hammond et al.,* 2018; *Maritorena et al.,* 2010; *Mélin,* 2016). These factors make climate-scale monitoring of ocean color very difficult with the currently available data. As we have attempted here, uncertainty must reduced if we are to identify a robust climate change signal in tropical Pacific ocean color retrievals. Drift in ocean color sensors must be less than 1% $decade^{-1}$ to identify a climate change signal clearly and discontinuities between sensors must be minimized (*Dutkiewicz et al.,* 2019; *Hammond et al.,* 2018).

The two longest operating ocean color missions to date are SeaWiFS at 13 years of operation and MODIS at 17. Combining multiple ocean color sensors into a single robust record is essential to detect a long-term climate change signal. Modelling has indicated that potential trends in this region are small and may be indistinguishable from natural variability until at least 2069 (*Henson et al.,* 2010) or even 2100 (*Dutkiewicz et al.,* 2019). Therefore, a climate change signature in the tropical Pacific will not be robustly separable from interannual variability for at least several more decades. This is further confounded by Pacific Decadal Oscillation impacting tropical Pacific chl and ecology at time scales longer than individual El Niño events (*Chavez,* 2003; *Schollaert Uz et al.,* 2017).

The new TPCA algorithm improved the consistency between both SeaWiFS and MODIS-Aqua to MERIS, but slightly reduced the consistency between SeaWiFS and MODIS-Aqua. MERIS was the best performing sensor in this region with the least bias, and to reflect this we retained the NASA implementation. Larger cross-sensor variability in chl occurred in the eastern Pacific than anywhere else in the basin (Figure 10). Satellite chl algorithms are highly sensitive to optical properties (*Hu et al.,* 2012). Given the larger natural variability in in this region, it is unsurprising to identify larger sensitivities in the eastern Pacific than in the stratified, low chl, low variability western portion.

4.5 ENSO and long-term trends

A 21-year trend analysis for the 1997-2018 tropical Pacific ocean color record is presented in Figure 10. The TPCA reprocessing halved the magnitude of observed trends, from an average of -1.07% decade^{-1} to -0.49% decade^{-1}. The most substantial trend change was found in the eastern tropical Pacific, where chl concentration and variability is largest. These differences

were likely caused by the poor performance of MODIS-Aqua, which was the sensor with the most substantial chl increases under the TPCA algorithm compared to NASA. MODIS-Aqua was the only sensor active at the end of our satellite record, so the NASA underestimation may have amplified negative trends.

Basin-wide positive anomalies are seen in Figure 9. Small areas of negative anomalies are found around the Galapagos islands, Kiribati islands, Marquesas islands and warm-pool regions. These Pacific islands have been linked with high productivity caused by island mass effects (*Messié et al.,* 2006; *Signorini et al.,* 1999; *Palacios,* 2002). The negative anomalies in Figure 9; d, e, f indicate that these seasonal island mass effect blooms have been historically overestimated. This is especially likely during El Niño for Kiribati and La Niña for the Marquesas and the Galapagos islands. These region- and sensor-specific differences infer spatial changes in optical properties and phytoplankton community structure east-west throughout the tropical Pacific.

Primary productivity in the tropical Pacific is strongly influenced by ENSO events (*Behrenfeld et al.,* 2006; *Chavez et al.,* 1999; *Radenac et al.,* 2012; *Stramma et al.,* 2016; *Strutton et al.,* 2008). Long-term chl trends are masked by high levels of variability as a result of ENSO (*Boyce et al.,* 2010; *Mélin et al.,* 2017). For example, events such as the 1997-1998 and 2015-2016 El Niño, drove a strong chl decrease in the central and eastern Pacific, and positive trends in the western Pacific (*Chavez et al.,* 1999; *Mélin et al.,* 2017). During the twentieth century, the canonical El Niño has become less frequent, while El Niño Modoki, also known as central Pacific El Niño events, have become more common (*Yeh et al.,* 2009). Extreme El Niño events are forecast to intensify under a changing climate (*Cai et al.,* 2014).

5 Conclusions

The tropical Pacific Ocean is a region of large scale biogeochemical variability and it is integral to global carbon cycles. A regionally tuned tropical Pacific chlorophyll algorithm (TPCA; Table 4) removed systematic underestimation in existing satellite ocean color algorithms. NASA estimates typically performed well in this region but were influenced by systematic negative biases. The blending window in the OCI algorithm is the most significant factor in chl algorithm effectiveness. Larger blending windows were more effective, likely due to the non-linearity intrinsic to these algorithms. The updated OCI2 *Hu et al.* (2019) coefficients and blending window were less effective in this region than that currently distributed by NASA, proposed by *Hu et al.* (2012).

The TPCA increased chl retrievals in the tropical Pacific, with the most significant increase for MODIS-Aqua, and no change for MERIS. Chl increased from 0.154 to 0.159mg m^{-3} for SeaWiFS, 0.142 to 0.160mg m^{-3} for MODIS, and MERIS remained at 0.153mg m^{-3}, with some spatial and temporal variability dependent on ENSO state. High chl due to island mass effects around the Kiribati, Marquesas and Galapagos Islands were historically overestimated. A chl gradient was present west-east across the tropical Pacific, with the TPCA presenting a steeper gradient than NASA for El Niño and neutral conditions. The proposed TPCA algorithms halve basin-wide long-term trends from -1.07% $decade^{-1}$ to -0.49% $decade^{-1}$. No significant long-term changes can be found for this highly variable region with the current data record. However, improving the accuracy of estimates in this region will allow us to further discern climate-driven impacts on phytoplankton.

Future work might include a further assessment of the blending window between chl_{OCx} and chl_{CI}, perhaps further exploring MBR blending rather than chl_{CI} (*Wang and Son,* 2016). The

possibility of blending more than two algorithms could also be explored. Machine learning and artificial neural network methods for chl retrieval and blending may also be a promising solution. As these technologies mature, they may provide increasingly accurate chl retrievals from R_{rs} and other geophysical variables. Recently, O'Reilly and Werdell (2019) published an update for the OCx algorithm coefficients. If our analysis were to be updated and expanded, it would be essential to include these updated versions, and also compare their coefficient derivation approach with ours.

Acknowledgments

N.P. and P.S. acknowledge support from the ARC Centre of Excellence for Climate Extremes (CE170100023). This research was undertaken with the assistance of resources and services from the National Computational Infrastructure (NCI), which is supported by the Australian Government. We would like to thank NASA/Goddard Space Flight Center for freely providing satellite chlorophyll data for SeaWiFS, MODIS-Aqua and MERIS. Python 3.7 was used here for analysis, with packages including xarray, Numpy, Scipy, Pandas, and Matplotlib.

The TPCA algorithm Python source code, matchup database and TPCA satellite reprocessed satellite data are available as: Pittman, N., 2019. Tropical Pacific Chlorophyll Algorithm (TPCA): Reprocessing v2019_01. NCI Data Collection doi: 10.25914/5dccbd3b64bdc.

References

Antoine, D., d'Ortenzio, F., Hooker, S.B., Bécu, G., Gentili, B., Tailliez, D., and Scott, A.J. (2008). Assessment of uncertainty in the ocean reflectance determined by three

satellite ocean color sensors (MERIS, SeaWiFS and MODIS-A) at an offshore site in the Mediterranean Sea (BOUSSOLE project). Journal of Geophysical Research *113*.

Bailey, S.W., and Werdell, P.J. (2006). A multi-sensor approach for the on-orbit validation of ocean color satellite data products. Remote Sensing of Environment *102*, 12–23.

Barber, R.T., Sanderson, M.P., Lindley, S.T., Chai, F., Newton, J., Trees, C.C., Foley, D.G., and Chavez, F.P. (1996). Primary productivity and its regulation in the equatorial Pacific during and following the 1991–1992 El Nino. Deep Sea Research Part II: Topical Studies in Oceanography *43*, 933–969.

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., and Boss, E.S. (2006). Climate-driven trends in contemporary ocean productivity. Nature *444*, 752–755.

Boyce, D.G., Lewis, M.R., and Worm, B. (2010). Global phytoplankton decline over the past century. Nature *466*, 591.

Boyer, T.P., Baranova, O.K., Coleman, C., Garcia, H.E., Grodsky, A., Locarnini, R.A., Mishonov, A.V., Paver, C.R., Reagan, J.R., Seidov, D., et al. (2018). World Ocean Database 2018. A. V. Mishonov, Technical Editor, NOAA Atlas NESDIS 87.

Cai, W., Borlace, S., Lengaigne, M., van Rensch, P., Collins, M., Vecchi, G., Timmermann, A., Santoso, A., McPhaden, M.J., Wu, L., et al. (2014). Increasing frequency of extreme El Niño events due to greenhouse warming. Nature Climate Change *4*, 111–116.

Chavez, F., Strutton, P., Friederich, G., Feely, R., Feldman, G., Foley, D., and McPhaden, M. (1999). Biological and chemical response of the equatorial Pacific Ocean to the 1997-98 El Niño. Science *286*, 2126–2131.

Chavez, F.P., Buck, K.R., Bidigare, R.R., Karl, D.M., Hebel, D., Latasa, M., Campbell, L., and Newton, J. (1995). On the chlorophyll a retention properties of glass-fiber GF/F filters. Limnology and Oceanography *40*, 428–433.

Chavez, F.P., Buck, K.R., Service, S.K., Newton, J., and Barber, R.T. (1996). Phytoplankton variability in the central and eastern tropical Pacific. Deep Sea Research Part II: Topical Studies in Oceanography *43*, 835–870.

Chavez, F.P. (2003). From Anchovies to Sardines and Back: Multidecadal Change in the Pacific Ocean. Science *299*, 217–221.

Cota, G. (2004). Transformation of global satellite chlorophyll retrievals with a regionally tuned algorithm. Remote Sensing of Environment *90*, 373–377.

Djavidnia, S., Mélin, F., and Hoepffner, N. (2010). Comparison of global ocean colour data records. Ocean Science *6*, 61–76.

Dutkiewicz, S., Hickman, A.E., Jahn, O., Henson, S., Beaulieu, C., and Monier, E. (2019). Ocean colour signature of climate change. Nature Communications *10*.

Feely, R.A., Boutin, J., Cosca, C.E., Dandonneau, Y., Etcheto, J., Inoue, H.Y., Ishii, M., Quéré, C.L., Mackey, D.J., McPhaden, M., et al. (2002). Seasonal and interannual variability of CO2 in the equatorial Pacific. Deep Sea Research Part II: Topical Studies in Oceanography *49*, 2443–2469.

Garcia, C.A.E., Garcia, V.M.T., and McClain, C.R. (2005). Evaluation of SeaWiFS chlorophyll algorithms in the Southwestern Atlantic and Southern Oceans. Remote Sensing of Environment *95*, 125–137.

Gierach, M.M., Lee, T., Turk, D., and McPhaden, M.J. (2012). Biological response to the 1997–98 and 2009–10 El Niño events in the equatorial Pacific Ocean. Geophysical Research Letters *39*.

Gómez Jakobsen, F.J., Mercado, J.M., Tovar-Salvador, M.L., Cortés, D., Yebra, L., Salles, S., Sánchez, A., Valcárcel-Pérez, N., and Alonso, A. (2018). New algorithms for estimating chlorophyll-a in the Spanish waters of the Western Mediterranean Sea from multiplatform imagery. International Journal of Remote Sensing 1–22.

Gordon, H.R., and McCluney, W.R. (1975). Estimation of the Depth of Sunlight Penetration in the Sea for Remote Sensing. Appl. Opt. *14*, 413.

Gregg, W.W., and Casey, N.W. (2004). Global and regional evaluation of the SeaWiFS chlorophyll data set. Remote Sensing of Environment *93*, 463–479.

Gross, L., Thiria, S., and Frouin, R. (1999). Applying artificial neural network methodology to ocean color remote sensing. Ecological Modelling *120*, 237–246.

Hammond, M.L., Beaulieu, C., Henson, S.A., and Sahu, S.K. (2018). Assessing the Presence of Discontinuities in the Ocean Color Satellite Record and Their Effects on Chlorophyll Trends and Their Uncertainties. Geophysical Research Letters *45*, 7654–7662.

Henson, S.A., Sarmiento, J.L., Dunne, J.P., Bopp, L., Lima, I.D., Doney, S.C., John, J.G., and Beaulieu, C. (2010). Detection of anthropogenic climate change in satellite records of ocean chlorophyll and productivity. Biogeosciences 7, 621-640.

Hooker, S.B., Firestone, E.R., Esaias, W.E., Feldman, G.C., Gregg, W.W., and Mcclain, C.R. (1992). SeaWiFS technical report series. Volume 1: An overview of SeaWiFS and ocean color.

Hu, C., Lee, Z., and Franz, B. (2012). Chlorophyll a algorithms for oligotrophic oceans: A novel approach based on three-band reflectance difference. Journal of Geophysical Research: Oceans *117*.

Hu, C., Feng, L., Lee, Z., Franz, B.A., Bailey, S.W., Werdell, P.J., and Proctor, C.W. (2019). Improving Satellite Global Chlorophyll a Data Products Through Algorithm

Refinement and Data Recovery. Journal of Geophysical Research: Oceans, *124*(3), 1524–1543.

Ioannou, I., Gilerson, A., Gross, B., Moshary, F., and Ahmed, S. (2013). Deriving ocean color products using neural networks. Remote Sensing of Environment *134*, 78–91.

Johnson, R., Strutton, P.G., Wright, S.W., McMinn, A., and Meiners, K.M. (2013). Three improved satellite chlorophyll algorithms for the Southern Ocean. Journal of Geophysical Research: Oceans *118*, 3694–3703.

Maritorena, S., d'Andon, O.H.F., Mangin, A., and Siegel, D.A. (2010). Merged satellite ocean color data products using a bio-optical model: Characteristics, benefits and issues. Remote Sensing of Environment *114*, 1791–1804.

McClain, C.R. (2009). A Decade of Satellite Ocean Color Observations. Annual Review of Marine Science *1*, 19–42.

McClain, C.R., Christian, J.R., Signorini, S.R., Lewis, M.R., Asanuma, I., Turk, D., and Dupouy-Douchement, C. (2002). Satellite ocean-color observations of the tropical Pacific Ocean. Deep Sea Research Part II: Topical Studies in Oceanography *49*, 2533–2560.

McPhaden, M.J., Zebiak, S.E., and Glantz, M.H. (2006). ENSO as an Integrating Concept in Earth Science. Science *314*, 1740–1745.

Mélin, F. (2016). Impact of inter-mission differences and drifts on chlorophyll- a trend estimates. International Journal of Remote Sensing *37*, 2233–2251.

Mélin, F., Vantrepotte, V., Chuprin, A., Grant, M., Jackson, T., and Sathyendranath, S. (2017). Assessing the fitness-for-purpose of satellite multi-mission ocean color climate data records: A protocol applied to OC-CCI chlorophyll- a data. Remote Sensing of Environment *203*, 139–151.

Messié, M., Radenac, M.-H., Lefèvre, J., and Marchesiello, P. (2006). Chlorophyll bloom in the western Pacific at the end of the 1997–1998 El Niño: The role of the Kiribati Islands. Geophys. Res. Lett. *33*, L14601.

Mock, T., and Hoch, N. (2005). Long-Term Temperature Acclimation of Photosynthesis in Steady-State Cultures of the Polar Diatom Fragilariopsis cylindrus. Photosynthesis Research *85*, 307–317.

Morel, A., and Prieur, L. (1977). Analysis of variations in ocean color 1. Limnology and Oceanography *22*, 709–722.

O'Reilly, J.E., Maritorena, S., Mitchell, B.G., Siegel, D.A., Carder, K.L., Garver, S.A., Kahru, M., and McClain, C. (1998). Ocean color chlorophyll algorithms for SeaWiFS. Journal of Geophysical Research: Oceans *103*, 24937–24953.

O'Reilly, J.E., Maritorena, S., O'Brien, M.C., Siegel, D.A., Toole, D., Menzies, D., Smith, R.C., Mueller, J.L., Mitchell, B.G., Kahru, M., et al. (2000). Volume 11, SeaWiFS Postlaunch Calibration and Validation Analyses, Part 3. 51.

Palacios, D.M. (2002). Factors influencing the island-mass effect of the Galápagos Archipelago. Geophys. Res. Lett. *29*, 49-1-49–4.

Radenac, M.H., Léger, F., Singh, A., and Delcroix, T. (2012). Sea surface chlorophyll signature in the tropical Pacific during eastern and central Pacific ENSO events. Journal of Geophysical Research: Oceans *117*.

Sauzède, R., Claustre, H., Jamet, C., Uitz, J., Ras, J., Mignot, A., and D'Ortenzio, F. (2015). Retrieving the vertical distribution of chlorophyll a concentration and phytoplankton community composition from in situ fluorescence profiles: A method based on a neural network with potential for global-scale applications. Journal of Geophysical Research: Oceans *120*, 451–470.

Schollaert Uz, S., Busalacchi, A.J., Smith, T.M., Evans, M.N., Brown, C.W., and Hackert, E.C. (2017). Interannual and Decadal Variability in Tropical Pacific Chlorophyll from a Statistical Reconstruction: 1958–2008. Journal of Climate *30*, 7293–7315.

Seegers, B.N., Stumpf, R.P., Schaeffer, B.A., Loftin, K.A., and Werdell, P.J. (2018). Performance metrics for the assessment of satellite data products: an ocean color case study. Optics Express *26*, 7404.

Signorini, S.R., McClain, C.R., and Dandonneau, Y. (1999). Mixing and phytoplankton bloom in the wake of the Marquesas Islands. Geophysical Research Letters *26*, 3121–3124.

Stramma, L., Fischer, T., Grundle, D.S., Krahmann, G., Bange, H.W., and Marandino, C.A. (2016). Observed El Niño conditions in the eastern tropical Pacific in October 2015. Ocean Science *12*, 861–873.

Strutton, P.G. and Chavez, F.P. (2000). Primary productivity in the equatorial Pacific during the 1997-98 El Niño. Journal of Geophysical Research 105(C11): 26,089-026,101.

Strutton, P.G., and Chavez, F.P. (2003). Scales of biological-physical coupling in the equatorial Pacific. Handbook of Scaling Methods in Aquatic Ecology. CRC Press, Boca Raton, FL.

Strutton, P.G., Evans, W., and Chavez, F.P. (2008). Equatorial Pacific chemical and biological variability, 1997–2003. Global Biogeochemical Cycles *22(2)*.

Sutton, A.J., Feely, R.A., Sabine, C.L., McPhaden, M.J., Takahashi, T., Chavez, F.P., Friederich, G.E., and Mathis, J.T. (2014). Natural variability and anthropogenic change in equatorial Pacific surface ocean pCO2 and pH. Global Biogeochemical Cycles *28*, 131–145.

Szeto, M., Werdell, P.J., Moore, T.S., and Campbell, J.W. (2011). Are the world's oceans optically different? Journal of Geophysical Research *116*.

Takahashi, T., Feely, R.A., Weiss, R.F., Wanninkhof, R.H., Chipman, D.W., and Sutherland, S.C. (1997). Global air-sea flux of CO2: An estimate based on measurements of sea-air pCO2 difference. Proceedings of the National Academy of Sciences *94*, 8292–8299.

Valente, A., Sathyendranath, S., Brotas, V., Groom, S., Grant, M., Taberner, M., Antoine, D., Arnone, R., Balch, W.M., Barker, K., et al. (2016). A compilation of global bio-optical in situ data for ocean-colour satellite applications. Earth System Science Data 18.

Volpe, G., Santoleri, R., Vellucci, V., Ribera d'Alcalà, M., Marullo, S., and D'Ortenzio, F. (2007). The colour of the Mediterranean Sea: Global versus regional bio-optical algorithms evaluation and implication for satellite chlorophyll estimates. Remote Sensing of Environment *107*, 625–638.

Wang, M., and Son, S. (2016). VIIRS-derived chlorophyll-a using the ocean color index method. Remote Sensing of Environment *182*, 141–149.

Wang, M., Son, S., and Harding, L.W. (2009). Retrieval of diffuse attenuation coefficient in the Chesapeake Bay and turbid ocean regions for satellite ocean color applications. Journal of Geophysical Research *114*.

Werdell, P.J. and Bailey, S.W. (2005). An improved in-situ bio-optical data set for ocean color algorithm development and satellite data product validation. Remote Sensing of Environment *98*: 122-140.

Yeh, S.-W., Kug, J.-S., Dewitte, B., Kwon, M.-H., Kirtman, B.P., and Jin, F.-F. (2009). El Niño in a changing climate. Nature *461*, 511–514.

Table 1: Sensor specific ocean color OCx and CI algorithm coefficients. Note that sensors have different red and green wavelengths. The OCx blue wavelength is the maximum of the bands nearest to 443, 490 and 510nm. The CI uses blue, green and red bands closest to 443, 555 and 670nm respectively. The CI coefficient provided for all sensors is from *Hu et al.* (2012). *Hu et al.* (2019) provide an updated set of sensor independent coefficients (OCI2).

Sensor	Blue (nm)	Green (nm)	Red (nm)	OC4 a_0	OC4 a_1	OC4 a_2	OC4 a_3	OC4 a_4	CI b_0	CI b_1
SeaWiFS(2018.0)	443>490>510	555	670	0.3272	-2.9940	2.7218	-1.2259	-0.5683	-0.4900	191.6590
MODIS-Aqua (2018.0)	443>488	547	667	0.2424	-2.7423	1.8017	0.0015	-1.2280	-0.4900	191.6590
MERIS (2012.1)	443>490>510	560	665	0.3255	-2.7677	2.4409	-1.1288	-0.4990	-0.4900	191.6590
Szeto et al. (2011) SeaWiFS Pacific	443>490>510	555	-	0.5109	-3.0871	1.1427	0.7416	-0.5230	-	-
OCI2 (*Hu et al.,* 2019)	-	-	-	-	-	-	-	-	-0.4287	230.4700

Table 2: Summary of SeaWiFS, MODIS-Aqua and MERIS temporal and spatial bin sizes from the matchup process. The selected matchup windows of 5 (±2) days and 9 (3x3) pixels are highlighted in bold.

SeaWiFS	**Total concurrent observations: 3426**				**Traditional diagnostics**			**Median log**	
Temporal average	Pixel average	Match %	# Matches	# Outliers	R^2	Slope	Intercept	MAE	MLB
Daily	1	19.00	651	11	0.763	0.604	0.054	1.205	0.969
Daily	9	28.26	968	37	0.767	0.579	0.053	1.211	0.942
Daily	25	33.48	1147	60	0.783	0.543	0.056	1.217	0.930
3 Day	1	45.80	1569	30	0.735	0.542	0.060	1.233	0.939
3 Day	9	61.21	2097	93	0.761	0.545	0.056	1.225	0.926
3 Day	25	68.71	2354	146	0.768	0.537	0.056	1.224	0.925
5 Day	1	60.01	2056	60	0.732	0.537	0.059	1.234	0.936
5 Day	**9**	**74.46**	**2551**	**151**	**0.756**	**0.536**	**0.057**	**1.232**	**0.924**
5 Day	25	80.85	2770	210	0.750	0.532	0.057	1.230	0.926
MODIS-Aqua	**Total concurrent observations: 1232**				**Traditional diagnostics**			**Median log**	
Temporal average	Pixel average	Match %	# Matches	# Outliers	R^2	Slope	Intercept	MAE	MLB
Daily	1	16.64	205	10	0.82	0.568	0.051	1.209	0.903
Daily	9	27.52	339	19	0.794	0.550	0.049	1.232	0.882
Daily	25	34.50	425	19	0.772	0.512	0.052	1.226	0.880
3 Day	1	39.45	486	28	0.803	0.581	0.045	1.209	0.907
3 Day	9	57.79	712	40	0.802	0.576	0.044	1.229	0.889
3 Day	25	67.13	827	42	0.795	0.542	0.047	1.222	0.885
5 Day	1	56.98	702	36	0.818	0.574	0.046	1.223	0.907
5 Day	**9**	**76.95**	**948**	**48**	**0.802**	**0.556**	**0.047**	**1.230**	**0.894**
5 Day	25	84.98	1047	52	0.800	0.549	0.047	1.229	0.896
MERIS	**Total concurrent observations: 1304**				**Traditional diagnostics**			**Median log**	

Temporal average	Pixel average	Match %	# Matches	# Outliers	R^2	Slope	Intercept	MAE	MLB
Daily	1	17.87	233	3	0.813	0.589	0.056	1.205	0.956
Daily	9	21.09	275	6	0.817	0.595	0.052	1.202	0.965
Daily	25	23.24	303	11	0.798	0.59	0.053	1.208	0.964
3 Day	1	50.69	661	14	0.789	0.600	0.055	1.207	0.972
3 Day	9	59.36	774	26	0.804	0.594	0.054	1.208	0.982
3 Day	25	64.11	836	30	0.803	0.592	0.055	1.219	0.982
5 Day	1	64.49	841	23	0.797	0.596	0.055	1.210	0.961
5 Day	**9**	**71.17**	**928**	**38**	**0.805**	**0.594**	**0.054**	**1.212**	**0.973**
5 Day	25	77.15	1006	44	0.799	0.592	0.054	1.218	0.981

Table 3: Top 10 OCI model combinations for each sensor; SeaWiFS (sw), MODIS-Aqua (modis) and MERIS (meris) on the training dataset. Rank is ordered by descending percent wins. Selected TPCA models with the most percent wins and least bias are in bold and are presented in further detail in Table 4. The last two listed models for each sensor are the current NASA algorithm (*Hu et al.,* 2012) and the proposed new OCI2 algorithm with updated CI coefficients and blending window (*Hu et al.,* 2019).

	Rank	OCx Polynomial	CI Polynomial	l	h	% wins	Bias	MAE	Slope	Intercept	R^2
SeaWiFS	1	sw_bestfit1	hu2012	0	1	56.7	0.958	1.222	0.553	0.058	0.788
	2	meris_nasa	hu2012	0	0.5	56.5	0.986	1.215	0.556	0.061	0.785
	3	sw_bestfit2	hu2012	0	1	55.8	0.943	1.222	0.539	0.057	0.789
	4	sw_3rd_order	hu2019	0	0.5	55.8	0.968	1.223	0.581	0.056	0.786
	5	sw_bestfit2	hu2012	0	0.5	55.6	0.966	1.226	0.609	0.051	0.784
	6	sw_3rd_order	meris_bestfit	0	0.5	55.5	0.969	1.222	0.573	0.057	0.786
	7	sw_nasa	sw_bestfit	0	0.5	55.3	0.997	1.214	0.568	0.061	0.785
	8	meris_nasa	hu2012	0	1	55.3	0.957	1.223	0.513	0.062	0.787
	9	modis_nasa	sw_bestfit	0	1	55.3	0.975	1.216	0.537	0.062	0.786
	13	szeto2011	sw_bestfit	0	0.5	54.8	1.004	1.215	0.569	0.061	0.781
	429	sw_nasa	hu2012	0.15	0.2	51.1	0.942	1.231	0.543	0.057	0.761
	909	sw_nasa	hu2019	0.25	0.4	44.6	0.897	1.241	0.556	0.049	0.787
MODIS-Aqua	1	szeto2011	hu2012	0	0.5	65.6	0.909	1.216	0.592	0.045	0.806
	2	sw_nasa	hu2012	0	0.5	65.3	0.910	1.224	0.591	0.044	0.812
	3	modis_bestfit1	hu2012	0	1	63.6	0.904	1.226	0.581	0.045	0.811
	4	modis_nasa	sw_bestfit	0	0.5	62.7	0.901	1.227	0.595	0.043	0.813
	5	3rd_order	hu2012	0	1	62.4	0.909	1.225	0.555	0.048	0.812
	6	meris_nasa	hu2012	0	0.5	61.6	0.953	1.202	0.634	0.044	0.811
	7	modis_bestfit1	hu2012	0	0.5	61.3	0.964	1.196	0.646	0.044	0.808
	8	modis_nasa	modis_bestfit	0	0.5	60.9	0.959	1.208	0.560	0.053	0.813
	9	szeto2011	meris_bestfit	0	0.2	60.9	0.967	1.208	0.724	0.035	0.799
	21	sw_nasa	hu2012	0	0.2	58.9	0.991	1.197	0.690	0.043	0.810
	957	modis_nasa	hu2012	0.15	0.2	33.6	0.860	1.255	0.561	0.042	0.804
	1007	modis_nasa	hu2019	0.25	0.4	31.8	0.811	1.306	0.588	0.032	0.808

MERIS	1	meris_nasa	hu2019	0	0.2	59.0	0.945	1.198	0.572	0.054	0.806
	2	meris_nasa	meris_bestfit	0	0.2	59.0	0.946	1.198	0.570	0.055	0.806
	3	meris_nasa	hu2012	0	0.2	58.7	0.965	1.198	0.552	0.060	0.807
	4	meris_nasa	hu2012	0.1	0.2	57.0	0.981	1.199	0.532	0.063	0.796
	5	meris_best_fit1	hu2012	0	0.2	57.0	0.999	1.194	0.585	0.060	0.807
	6	meris_best_fit1	hu2012	0.1	0.2	56.7	1.003	1.188	0.566	0.061	0.796
	7	meris_nasa	sw_bestfit	0	0.2	56.5	0.982	1.198	0.551	0.061	0.806
	8	meris_best_fit1	hu2019	0	0.1	56.3	0.999	1.207	0.612	0.055	0.807
	9	meris_best_fit1	meris_bestfit	0	0.1	56.3	0.999	1.207	0.611	0.056	0.807
	10	meris_best_fit1	meris_bestfit	0.1	0.2	56.3	0.960	1.186	0.595	0.053	0.794
	45	meris_nasa	hu2012	0.15	0.2	54.0	0.983	1.202	0.520	0.064	0.789
	254	meris_nasa	hu2019	0.25	0.4	51.1	0.948	1.201	0.534	0.059	0.780

Table 4: The selected tropical Pacific chl algorithm (TPCA) polynomial coefficients, blending window and validation dataset diagnostics for each sensor.

Sensor	SeaWiFS TPCA	MODIS-Aqua TPCA	MERIS TPCA (Discarded)	MERIS NASA
% Wins	56.50%	58.90%	56.90%	N/A
Median Bias	0.986	0.990	0.991	1.003
% Wins Rank	2	21	4	45
Chl_{OCx} Polynomial [a_0, a_1, a_2, a_3, a_4]	MERIS NASA [0.3255,-2.7677, 2.4409,-1.1288,-0.4990]	SeaWiFS NASA [0.3272,-2.9940, 2.7218,-1.2259,-0.5683]	Optimized [0.3863, -2.9664, 2.7350,-1.2195,-0.5651]	MERIS NASA [0.3255,-2.7677, 2.4409,-1.1288,-0.4990]
Chl_{CI} Polynomial	*Hu et al.,* (2012) [-0.4909, 191.6590]	*Hu et al.,* (2012) [-0.4909, 191.6590]	*Hu et al.,* (2012) [-0.4909, 191.6590]	*Hu et al.,* (2012) [-0.4909, 191.6590]
OCI Blending Window	0 to 0.5mg m^{-3}	0 to 0.2mg m^{-3}	0 to 0.2mg m^{-3}	0.15 to 0.2mg m^{-3}

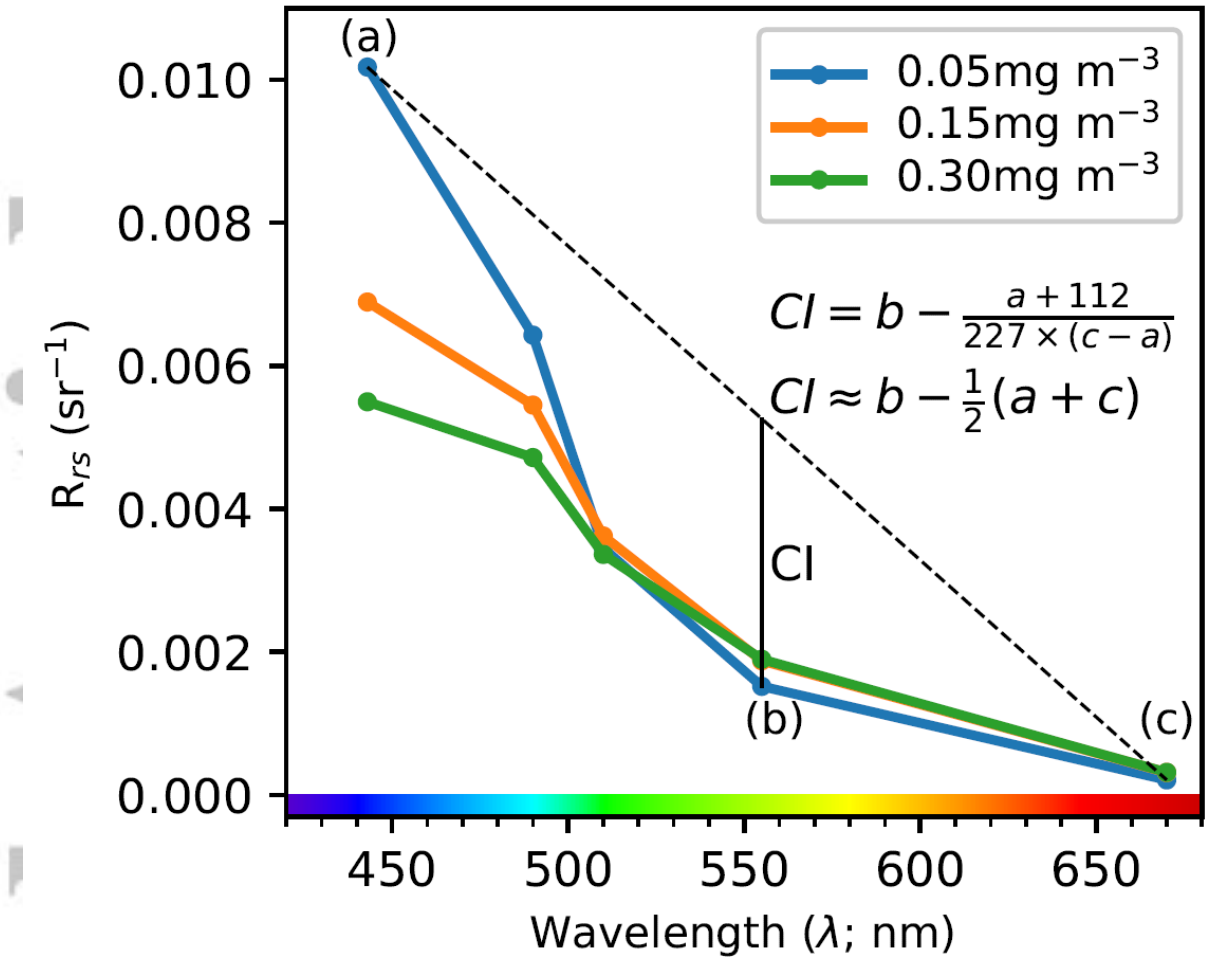

Figure 1: Illustration of reflectance values for the ocean color specific wavelengths observed by SeaWiFS; R_{rs} ([a] 443, 490, 510, [b] 555 and [c] 670) and a conceptual description of the CI algorithm, adapted from *Hu et al.* (2012). The three lines were created by identifying all pixels of a given chlorophyll concentration (0.05 (blue), 0.15 (orange) and 0.30mg m^{-3} (green) ±0.005mg m^{-3}), extracting the R_{rs} values for those pixels and averaging them to make a spectrum. Low chl concentrations reflect high blue and low green and red, compared to high chl with low blue, higher green and similar red background levels. The inclusion of red at R_{rs}(670) acts as a non-zero baseline, where the length of the vertical line indicates CI, a function for chl. The CI equation is derived from equation 2a, with the values 112 and 227 calculated by the difference between the green-blue and red-blue wavelengths respectively. An approximate color spectrum for the relevant wavelengths is shown at the bottom.

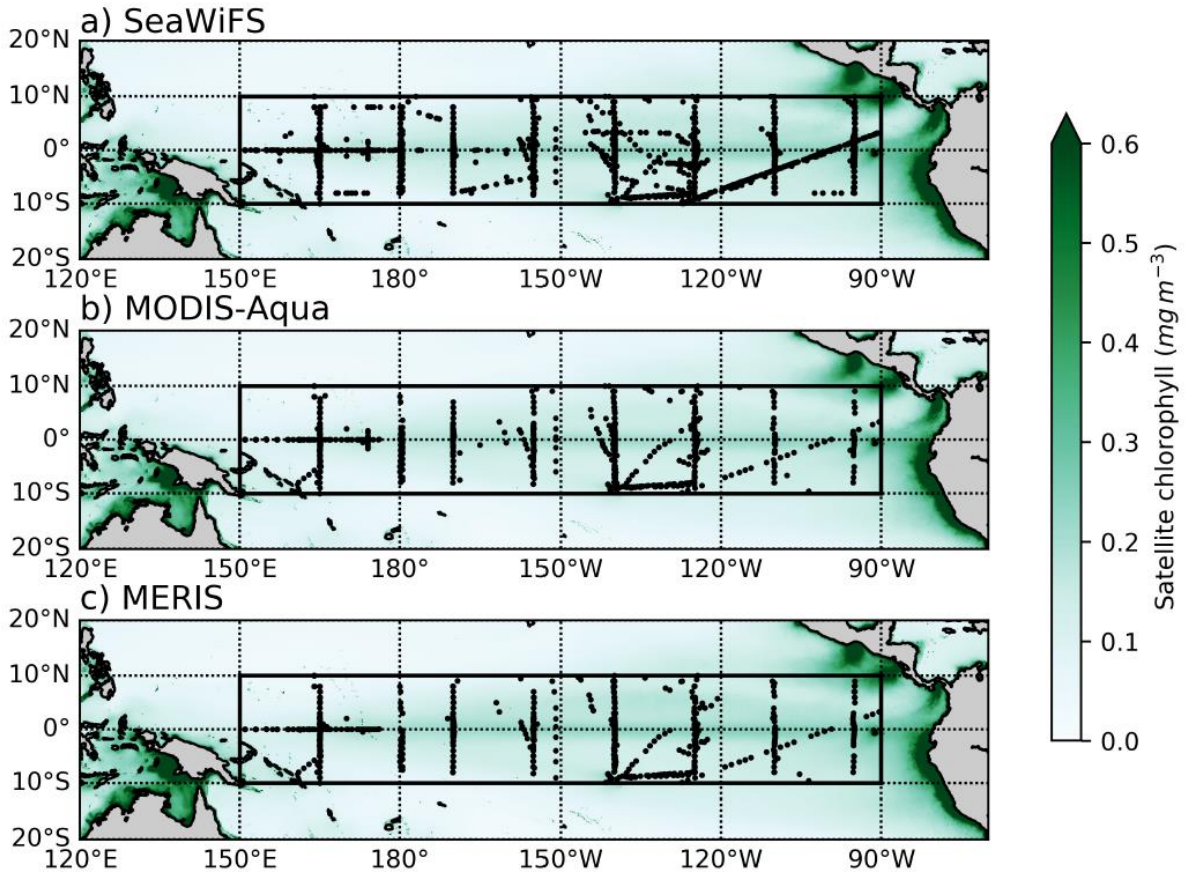

Figure 2: Map of the chlorophyll climatology and distribution of $chl_{in\ situ}$ locations for the satellite matchups (solid circles). The whole-of-mission chlorophyll climatology for each sensor's lifespan is shown (green shading).

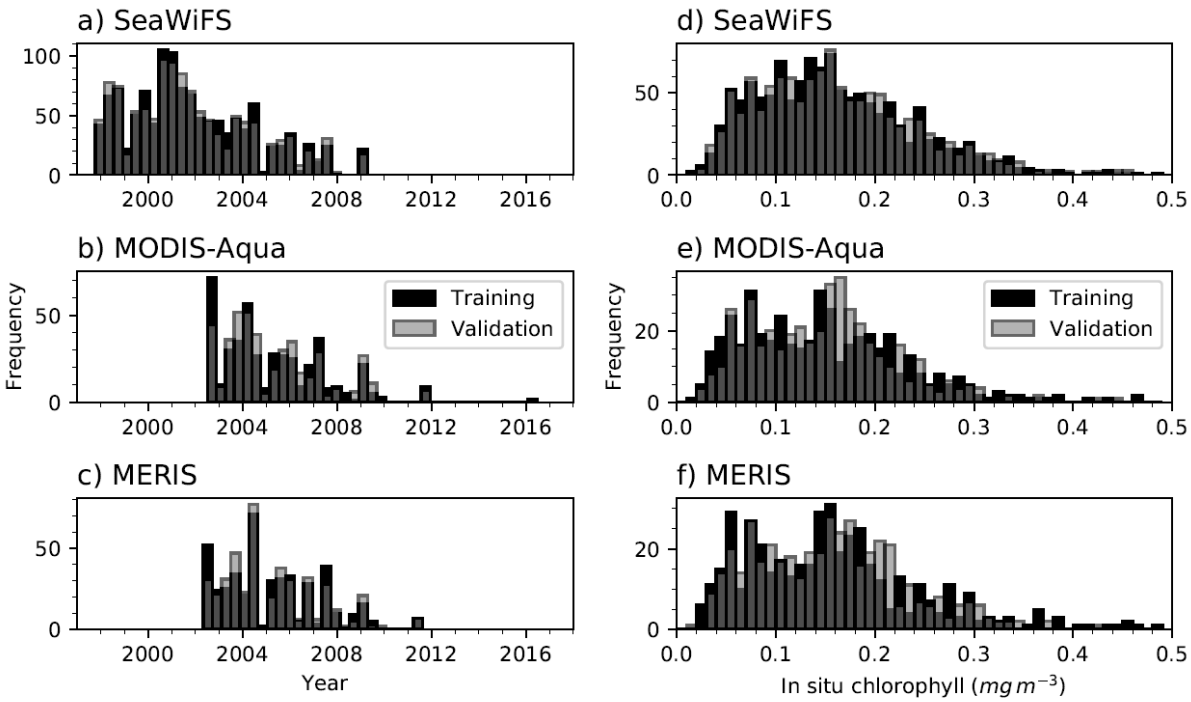

Figure 3: The temporal and concentration distribution of $chl_{in\ situ}$. a, b, c show that chl observations are not consistently distributed throughout time, with data clustered in the first half of the satellite record. The distribution between training and validation sets is relatively even. d, e, f shows the log-normal distribution of $chl_{in\ situ}$. The distribution is approximately equal between training and validation datasets.

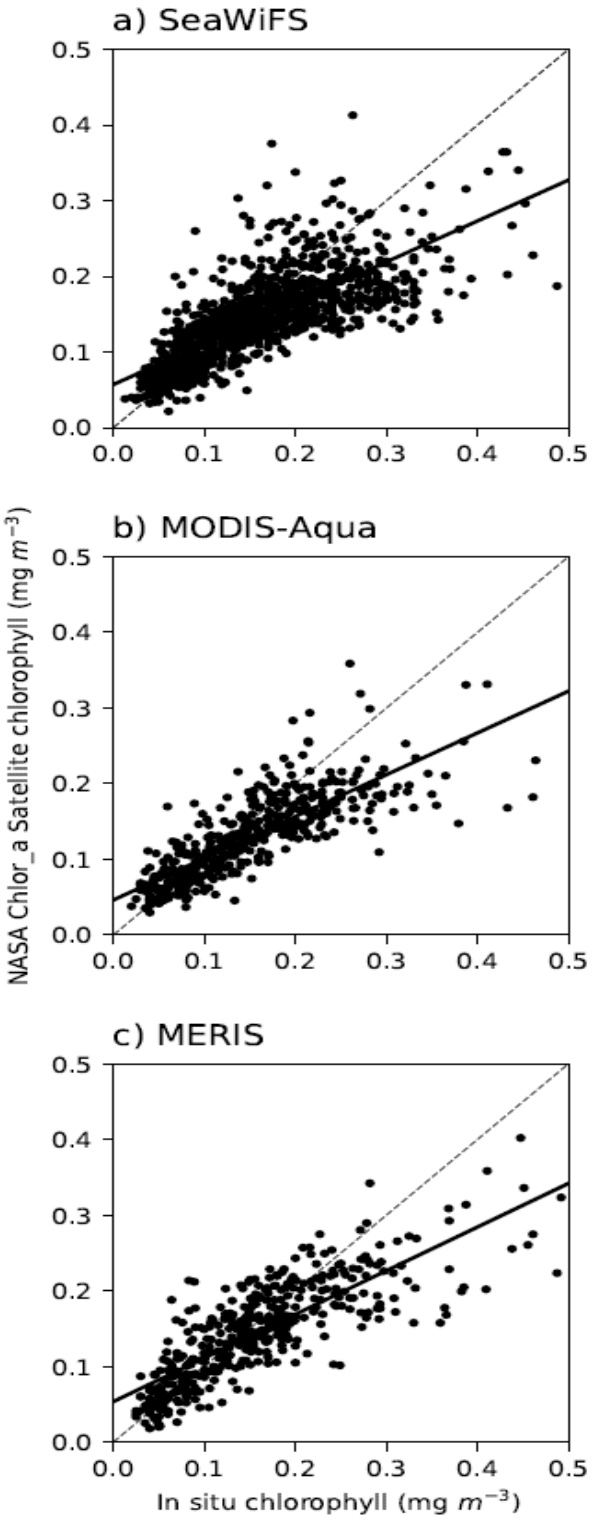

Figure 4: The current NASA chlor_a (OCI) satellite chl products verses the training $chl_{in\ situ}$ dataset for each satellite sensor (a) SeaWiFS, (b) MODIS-Aqua and (c) MERIS. The thin dashed line is $x=y$, and solid black line is the slope for each sensor.

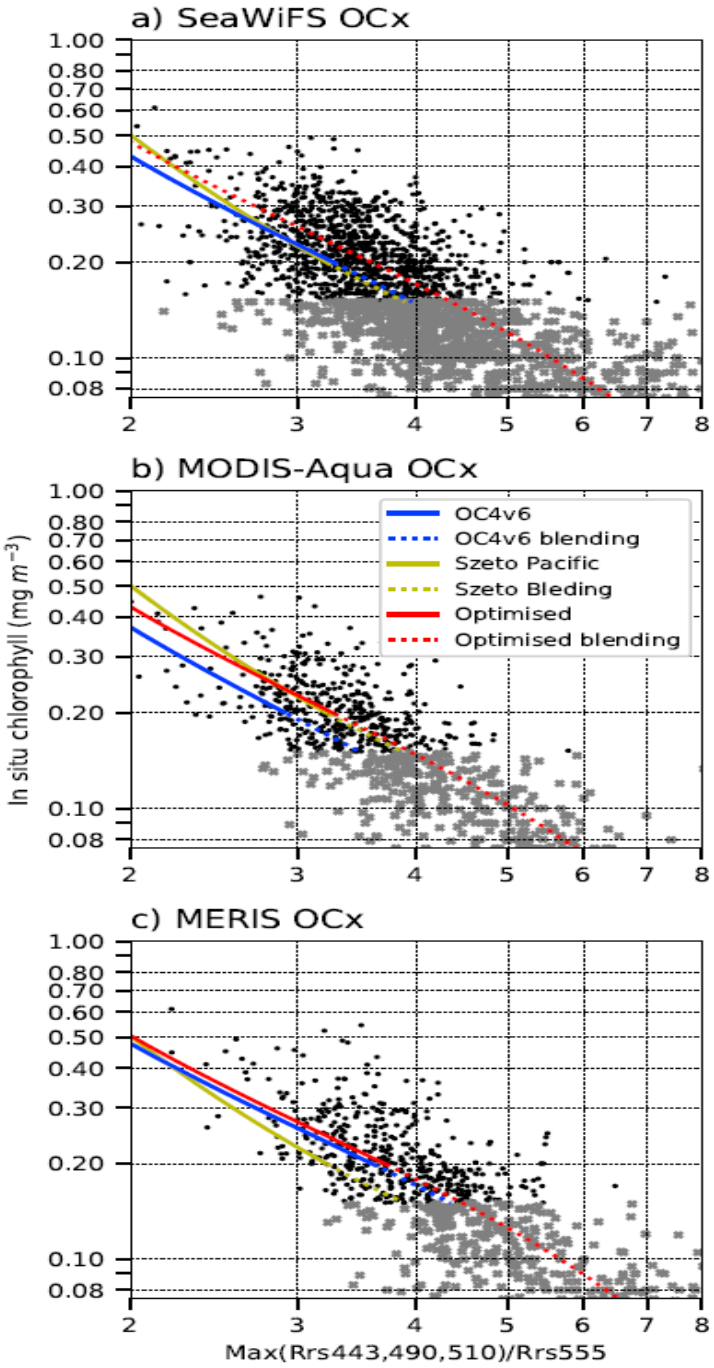

Figure 5: Default and optimized OCx polynomials for each sensor (a) SeaWiFS, (b) MODIS-Aqua and (c) MERIS. Blue lines are the current chlor_a products, yellow is *Szeto et al.* (2011) Pacific coefficient and red are our optimized models. The x axis is Max Band Ratio (MBR). Dotted lines indicate the blending zone. Gray data points are those located below the 0.15 blending cutoff, and use NASA chl_{CI}.

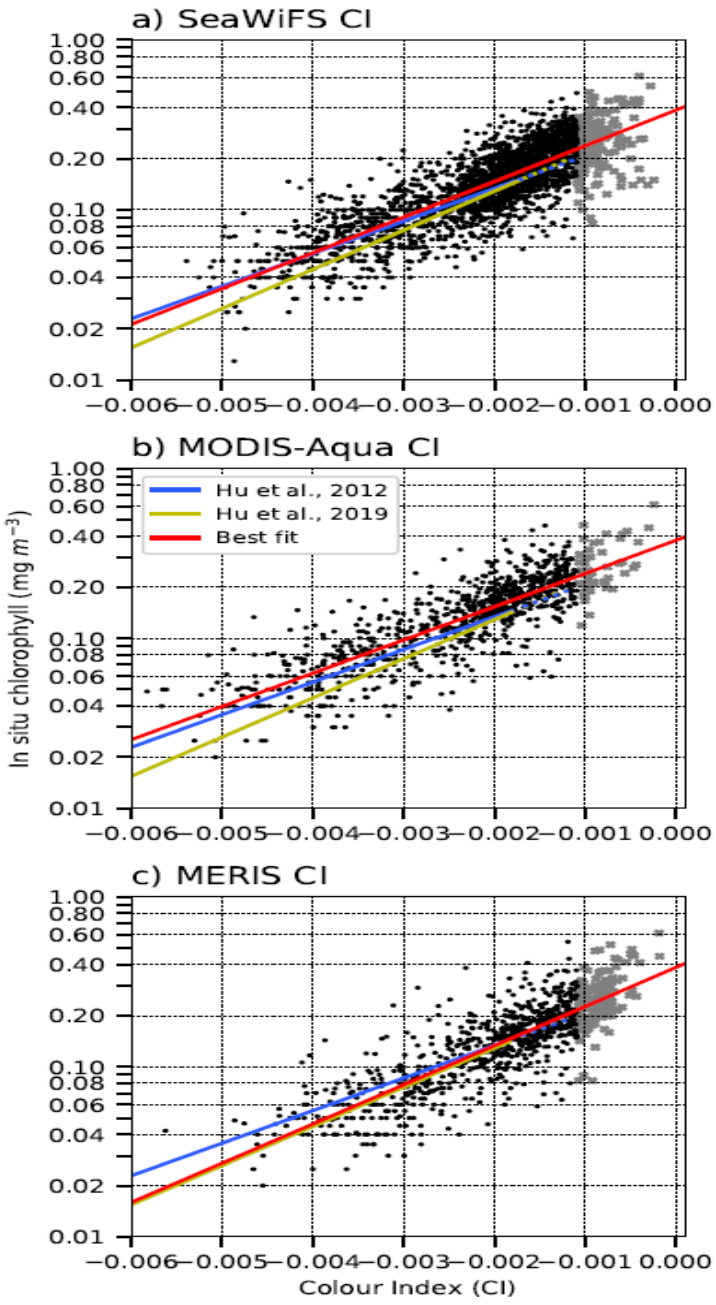

Figure 6: Default and optimized CI coefficients for each sensor (a) SeaWiFS, (b) MODIS-Aqua and (c) MERIS. The x axis is the Color Index (CI). Blue line is the coefficients proposed by *Hu et al.* (2012), and is the best fit for every sensor. Yellow lines are new updated CI coefficients for the next global reprocessing (*Hu et al.,* 2019). Red line is the best fit regression through the training dataset. Dotted lines are the blending windows and grey data points are that outside of the original blending zone and use chl_{OCx} (i.e. above 0.2mg m^{-3}).

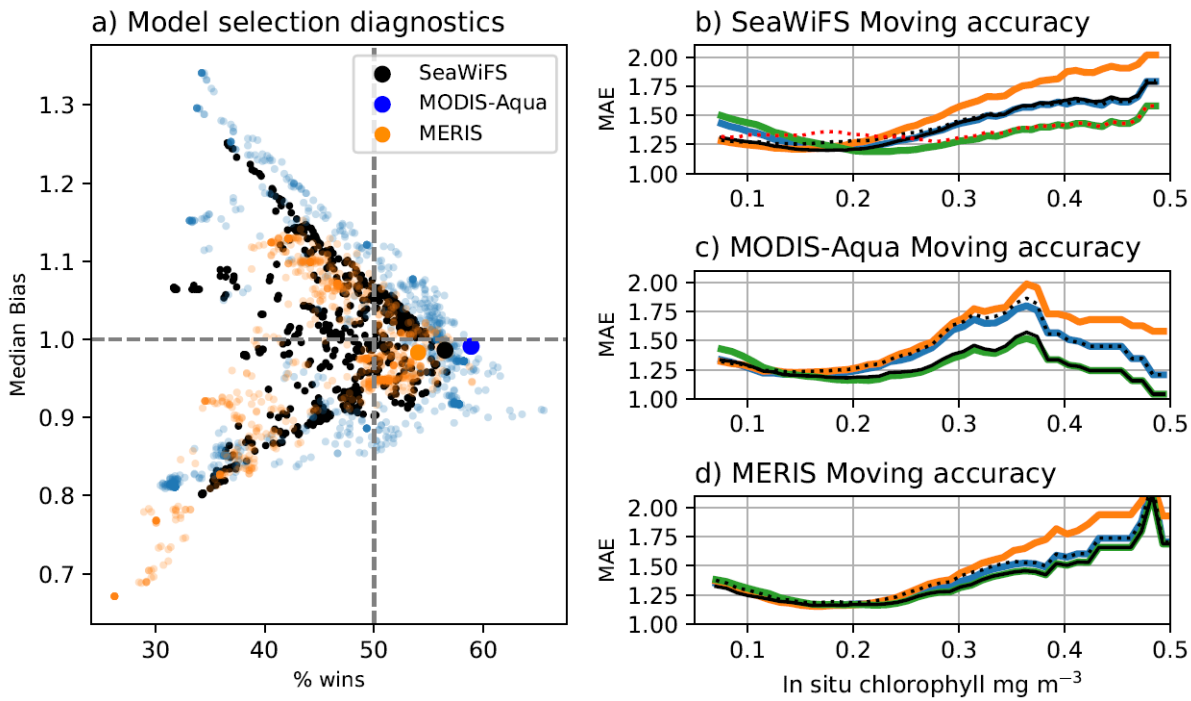

Figure 7: Rationale for TPCA model choice (a) for SeaWiFS (black), MODIS-Aqua (blue) and MERIS (NASA algorithm; orange). This is an integration of the median bias and point-for-point percent wins for 1218 algorithm combinations of OCx, CI and blending window sizes and locations. Large circles indicate selected models with the highest percent wins and least bias. Moving MAE accuracy is presented for each sensor (b) SeaWiFS, (c) MODIS-Aqua and (d) MERIS at 0.1mg m^{-3} bins. Orange lines are chl_{CI} estimates, blue the NASA chl_{OCx}, green the best performing chl_{OCx} polynomial coefficients, thin black line the new optimized model, and dotted black line the NASA chl_{OCI} estimates. The red dotted line in (b) is the default 0.15 to 0.2mg m^{-3} blending with the TPCA coefficients, resulting in increased uncertainty over the blending window at 0.18mg m^{-3}.

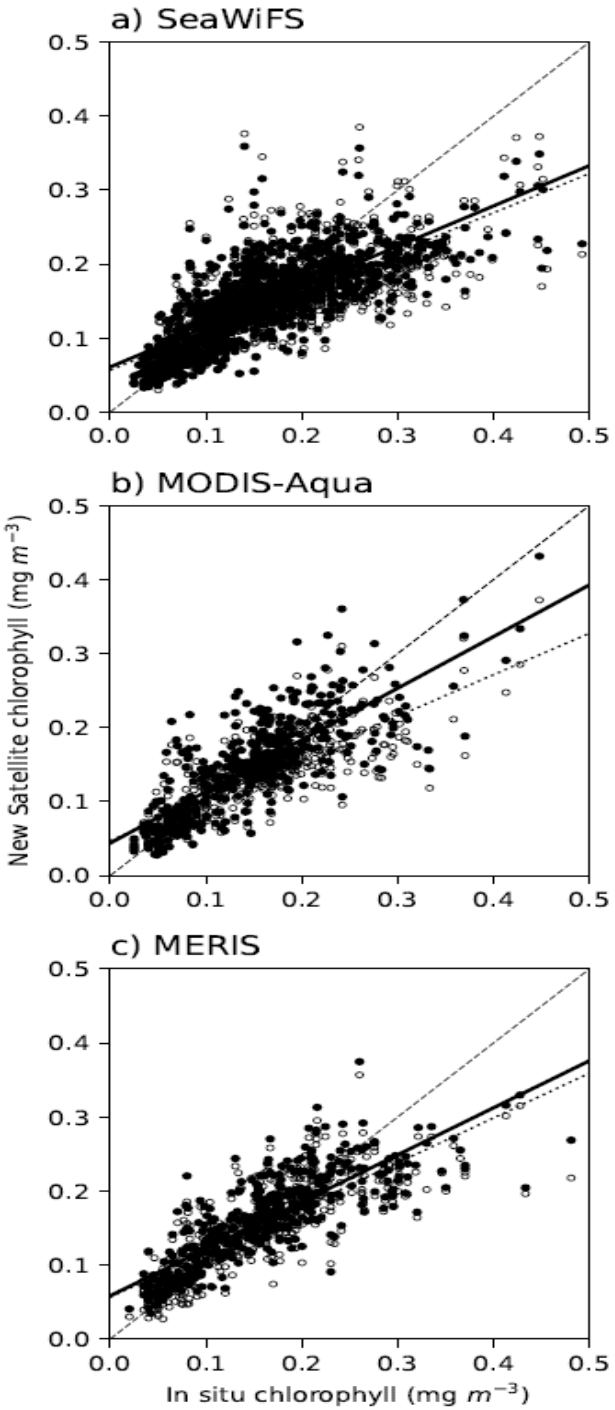

Figure 8: Comparison of NASA chlor_a (OCI) products (open circles) to the optimized satellite chl algorithm (closed circles) verses validation in situ observations. (a) SeaWiFS, (b) MODIS-Aqua and (c) MERIS. The black line is optimized slope, thin dotted line is the chlor_a slope and thick dashed line is $x=y$.

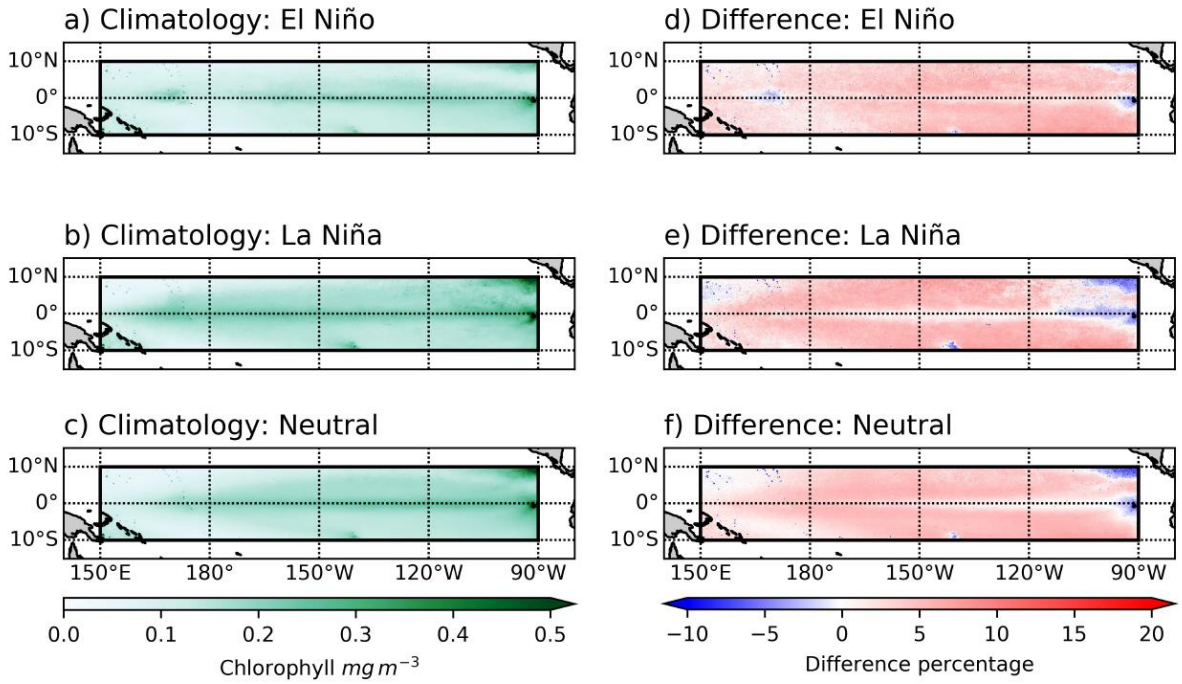

Figure 9: SeaWiFS spatial climatology (a,b,c) and differences (d,e,f) for El Niño (a, d) , La Niña (b, e), Neutral MEI (c, f) indexes across the tropical Pacific ocean. Difference is defined as (TPCA – NASA) / NASA) × 100. Red indicates TPCA estimates are higher than NASA. The algorithm is only applied within the target area of 10N-10S, 150E to 90W.

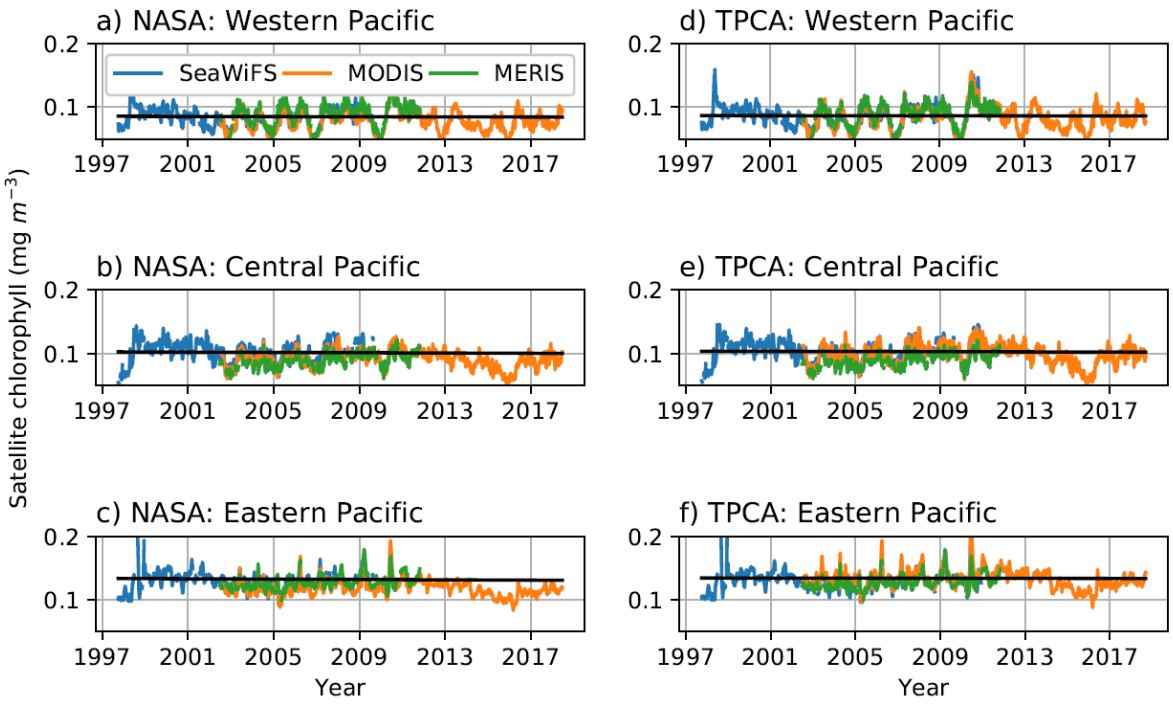

Figure 10: Time series of western (a, d; 165°-180°E, ±10 N and S), central (b, e; 170°-155°W, ±10 N and S) and eastern (c, f; 140°-115°W, ±10 N and S) Pacific boxes adapted from *Strutton et al.* (2008). Left panels (a, b, c) are NASA products and right side (d, e, f) are the updated TPCA algorithms. MERIS uses the original NASA implementation. Blue line is SeaWiFS, green is MERIS and orange is MODIS. The full 21 year ocean color time series is presented from September 1997 to September 2018. Black line is a linear regression showing the trend over this period. Average decadal change and average chlorophyll for each product and box is described in text (Section 3.4).