Recommendations for the correction of non-photochemical quenching of chlorophyll fluorescence in the Southern Ocean

## Introduction

Southern Ocean deployments of in-situ fluorometers are plagued by the wide-spread effects of day-time non-photochemical quenching (NPQ) (1;2;3). Many methods have been suggested to correct vertical profiles from in-situ fluorometers for the suppression of surface fluorescence in high light environments (4;5;6;7;1;2;3). These corrections attempt to restore the in-situ chlorophyll fluorescence measurements to chlorophyll-a concentrations, accounting for changing light conditions.

NPQ is expected to be widely prevalent in the iron limited Southern Ocean during summertime when light levels are high. The suppression of chlorophyll fluorescence by NPQ is initiated by marine phytoplankton when under extreme light stress. This response prevents the build-up of harmful reactive oxidants when nutrients are severely lacking for photosynthesis, by dissipating energy away from photosynthetic apparatuses. Iron limitation in Southern Ocean phytoplankton communities has been shown to exacerbate the magnitude of NPQ. As a result, innovative proxies are emerging for iron limitation based on NPQ measurements from in-situ fluorometers (8;9;10;11).

It is important to assess the impact of NPQ corrections on the accuracy of chlorophyll-a estimates from chlorophyll fluorescence, as datasets from deployments of in-situ fluorometers find use in large scale studies and the validation of biogeochemical models and remote sensing products (12;**Haentjens2017?**). Current assessments of NPQ corrections compare day-time measurements to a night-time reference to assess the relative accuracy of NPQ corrections. These assessments have shown preliminary evidence that performance is variable in the presence of changing oceanic conditions, indicated by the presence of subsurface chlorophyll-a maxima (SCMs) (1;2). Exploring these effects further in the Southern Ocean is essential, due to its high prevalence of both NPQ and SCMs in in the high-light environment of austral summer (**Baldry2020?**).

This work will be constrained to in-situ fluorometers deployed on ship-based rosettes and biogeochemical Argo (BGC-Argo) floats, for the purposes of characterising the vertical distribution of phytoplankton. We assess the performance of eight NPQ correction methods in the Southern Ocean, by comparing NPQ corrected fluorescence to chlorophyll measurements within the BIO-MATE ship data compilation (**BIOMATE?**). We calculate a residual sum of square (RSS), chlorophyll biases throughout the profile and the success of detecting a SCM and high surface chlorophyll (HSC) after NPQ correction.

From this wholistic approach, we find that the X12 and P18 methods perform best, even when considering methods supplemented by beam attenuation or backscatter data. The widely-used S08 method had a more varied performance between profiles and its application introduced on average up to 3% more RSS error.

To compile datasets from fluorescence profiles that are minimally impacted by the effects of NPQ in the Southern Ocean, we recommend:

* A deeper reference depth of 20 m is used to calculate MLD for NPQ correction methods.
* That B15, S15 and S15eu methods are not used in large-scale applications.
* In cases where the MLD is shallower than the SFM and a TML isn’t detected, that a new method X23 be applied to adjust for shallow mixed layers.
* In cases where the MLD is deeper than the SFM or a FWL exists, the X12 or P18 method is applied.
* In cases where a TML is detected, the X12/S08 method be applied. Caution is advised in this case when interpreting SCMs and surface chlorophyll, after a significant correction for NPQ effects (X % maximum fluorescence), particularly if there is evidence of complex mixing and multiple SFMs.
* Users note that high surface chlorophyll conditions cannot be reproduced when profiles are significantly quenched, and surface estimates of chlorophyll from chlorophyll fluorescence will be underestimated.
* That smoothed, uncalibrated beam attenuation derived from transmissometers can be used in place of backscatter in the S08 method for ship-based data.
* When applying the S08 method to BGC-Argo floats, only well understood, smoothed and quality-controlled backscatter and beam attenuation data are used. If a user is in doubt of this, similar, if not better results can be obtained using the X12 and P18 methods.
* caution in interpreting results when a NPQ signal > X% is observed.
* Most, significant SCMs form below Zeu(0.1).
* TML – X12 better for SCMS. Some SCMs still errerous.

users note a clear caveat in that NPQ affected fluorescence profiles are unable to reproduce the high surface chlorophyll (HSC) conditions observed….where?.

## Data Description and Methods

#### Ship dataset

To assess different NPQ correction methods, coincident profiles of chlorophyll concentration and chlorophyll fluorescence from shipboard systems were used. A common ship board system consists of an instrument package lowered to depth via a vessel winch. The instruments typically include a conductivity, temperature and depth (CTD) sensor, a sampling rosette loaded with Niskin bottles to collect water samples and a fluorometer.

Ship data was accessed using the BIO-MATE data product (v1.0;**BIOMATE?**). BIO-MATE contains open access ship data, converted to a standard format for the easy use of coincident data collected using different methods.

This study uses data collected on 59 oceanographic voyages (??), where both an *in-situ* fluorometer was deployed on a sampling rosette and chlorophyll-a concentration was determined by laboratory analysis of discrete water samples. 1977 profiles were selected for analysis which had at least four chlorophyll measurements shallower than the ecological mixed layer depth (EMLD, ??), one chlorophyll measurement shallower than 20 m, and one measurement below the EMLD. 1306 BIO-MATE profiles from 23 oceanographic voyages were unsuitable for our study based on these criteria. In cases where chlorophyll measurements were collected using multiple methods, high-performance liquid chromatography was the preferred method over fluorometry. Chlorophyll data marked as “not detected” was set to 0 mg/m3.

Fluorescence profiles were smoothed using an abstraction method. To remain comparable to chlorophyll measurements, depths for fluorescence measurements were calculated from pressure using the swDepth function in the R package *seacarb*. Zero fluorescence profiles with a high variability were removed from analysis. We define a high level of variability as profiles that have a residual standard error that is greater than 10% of the range of the profile, when compared to the smoothed profile.

To provide further quality assurance on the data, each profile was visually inspected. From this inspection, questionable fluorescence profiles and chlorophyll measurements were removed. Profiles that didn’t sample the entire ecological layer were also excluded. A list of the data removed in this process is in the supplementary material, which included ‘r length(which(QA$fluor\_profile == "Y" | QA$chl\_q == “all”))’ profiles and 146 chlorophyll measurements.

After post-processing and quality assurance we identified 1154 profiles to be included in analysis, with only 1154 profiles that include optical backscatter and 840 profiles that included beam attenuation. We worked with only the shallowest 500 m of profiles. Profiles locations spanned the Southern Ocean, including the Polar Frontal Zone, the continental zone of Antarctica and the sea-ice zone. An area corresponding to the Palmer Long-Term Ecological Record (PAL-LTER) was disproportionately sampled compared to other regions. Profiles were collected between 1992 and 2017 and mostly sampled during the day in austural summer (??).

#### Abstraction and calibration of ocean profiles

Fluorescence, backscatter, beam attenuation, density and temperature profiles were abstracted to remove fine scale variability, sensor noise and to remain comparable to chlorophyll profiles. The abstraction method reduces a profile to a piecewise linear fit. The benefits of abstraction, against more widely accepted smoothing methods including non-linear fits (13;14) and smoothing windows (i.e. moving means,15), are that it minimises the suppression of a thin SCM feature, is not influenced by outlying spikes and can be applied to a wide range of profile shapes. The abstraction method was built using the segmented.lm function from the R package *segmented*. The method is highly adaptable, and its performance is not impacted by sampling resolution. Full code for the fit\_segmented\_bp function is available at github.com/KimBaldry/Ocean\_code/tree/main/cleaning/BSM.

The abstraction method fits a piece wise linear model, optimising the placement of a number of break points. We apply the segmented.lm function iteratively, starting at one break point. After an additional break point is added to the model, it is checked that the addition of a break point results in lower residual standard error and Bayesian information criterion of the fit. The method stops adding break points when five sequential break points fail this check. When the final number of break points are found, outliers outside three standard deviations of the residuals are removed and the model is re-fit. This yields the final abstraction model for the depth profile.

The fit\_segmented\_bp function was applied to smooth density profiles in one step. A two-step method was used to smooth fluorescence, backscatter and beam attenuation profiles, by applying the function to a top portion of the profile (75 m below the deepest measurement in the 80th percentile), and a bottom portion of the profile with little variability (up to 500 m). We found this abstraction method was able to capture SCMs well and rarely suppressed their magnitude, even when the SCM was thin (3-10 m). Abstracted fluorescence and backscatter depth profiles were then adjusted using a deep offset, calculated using the minimum value of a linear fit to the bottom portion of the profile. If backscatter profiles were less than 3m resolution, they were first despiked using a 5-point running median.

After NPQ correction methods were applied, abstracted fluorescence profiles were calibrated to chlorophyll measurements. This was done by first calculating average fluorescence over a 2 m window spanning the chlorophyll sampling depths. This buffer attempts to partially account for heave and the potential 1-2 m difference between the fluorometer and Niskin bottle. These fluorescence measurements are then adjusted by a scalar that minimises a weighted residual sum of squares of corrected fluorescence against chlorophyll measurements. To prevent an increase in error from deeper samples with chlorophyll close to zero, the residual sum of squares was inversely weighted to the square of chlorophyll values.

#### Correcting for NPQ using coincident chlorophyll-a data

To assess the performance of NPQ methods we developed a method to correct chlorophyll fluorescence profiles for NPQ with coincident chlorophyll-a observations. The resulting chlorophyll fluorescence profile, referred to as chlorophyll-informed fluorescence herein, is used as a “truth” basis to assess NPQ correction methods. Chlorophyll-informed fluorescence were derived using an optimized NPQ depth and a constant upward correction according to X12 (CF\_X12) or a variable upward correction according to S08 with beam attenuation (CF\_S08).

To derive an optimized NPQ depth, first a chlorophyll fluorescence calibration scalar and a NPQ depth was optimised by minimising a weighted (1/(maximum chlorophyll value)2) residual sum of squares. Once an optimal result was determined, NPQ depth was further optimised while fixing the calibration scalar to its optimized value. Constraints were put in that the NPQ depth had to be shallower than the EMLD, or 5 m above the SCM if one was present (and in the case of constant upward correction shallower or equal to the SFM), and such that the corrected values were always greater than or equal to the uncorrected values.

To assess the robustness of the optimized NPQ depth, paired observations of chlorophyll concentration and fluorescence were used to observed a NPQ depth. The NPQ depth was observed by finding the deepest measurement where the chlorophyll to fluorescence ratio consecutively decreased from the surface. The accuracy of this observation is reliant on the resolution of the chlorophyll depth profile.

#### NPQ correction methods

Of the ten existing methods to correct fluorescence profiles for NPQ effects, we assess six methods. The remaining five methods either require day and night sampling or data from radiometers. Methods which require day and night sampling have already been adequately assessed (1), and are not widely suitable for biogeochemical Argo floats and ships. These platforms rarely have the capacity to perform day and night sampling, and when they do we refer the reader to 1. Similarly, methods requiring radiometers are not explored here as only X% of floats have been historically equipped with radiometers.

Three methods, S08, S15 and S15eu, are only suitable when optical backscatter sensors are deployed in tandem with fluorometers, which is the case for the biogeochemical Argo network and is becoming more common on ships. Historically, ships have been equipped with beam attenuation sensors, rather than optical backscatter sensors, so we also test variations of S08, S15, S15eu with the application of beam attenuation in the place of backscatter.

#### Assessment of NPQ correction performance

To assess the performance of different NPQ methods, corrected fluorescence profiles were compared to chlorophyll measurments. For uncorrected, chlorophyll-informed and NPQ corrected fluorescence profiles, we calculated the RSS relative to chlorophyll-a measurements and a bias at the surface and at the SCM relative to chlorophyll-a measurements. Significant differences in RSS, surface bias and SCM bias between uncorrected fluorescence and chlorophyll-informed corrected fluorescence were detected using a two-sided pair-wise Wilcoxon robust test (for continuous variables) or a McNemar’s chi-squared test for symmetry (binary variables) to a 95% confidence level. The ability of a fluorescence profile to reproduce profile shape was determined by comparing a SCM test to a SFM test and a HSC test to a HSF test.

We use the chlorophyll-informed fluorescence as a baseline to assess the performance of NPQ correction methods. A pair-wise relative comparison allowed the inherent variability of the dataset to be removed when testing performance. Comparisons were made across error measures, calculated NPQ depth and calculated SFM depth. Significant differences in error measures between chlorophyll-informed corrected fluorescence and tested NPQ methods were detected using a pair-wise Wilcoxon robust test (for continuous variables) or a McNemar’s chi-squared test for symmetry (binary variables) to a 95% confidence level.

The effects of significant NPQ, the presence of a SCM and the presence of HSC were analysed by comparing error measures across groups. Significant effects were detected using a Kolmogorov–Smirnov test (continuous variables) or a Pearsons chi-squared test (binary variables) for differences in distribution between treatment groups with a confidence level of 95 %. The effect of different ocean regimes was also analysed by comparing measures across groups, however to detect significant differences between groups a Kruskal-Wallis test was use in place of the Kolmogorov–Smirnov test

#### Identifying NPQ

To identify significant NPQ in fluorescence data, we compared chlorophyll-informed fluorescence using a constant upward correction (CF\_X12) to uncorrected fluorescence. For surface chlorophyll-a less than 1 mg/m3, uncorrected fluorescence had to be less than chlorophyll-informed fluorescence by at least 0.1 mg/m3. For surface chlorophyll-a concentrations higher than 1mg/m3, uncorrected fluorescence had to be at least 10% less than chlorophyll-informed fluorescence.

#### Characterising chlorophyll profile features

The effect of NPQ and NPQ correction methods were explored over two common chlorophyll profile features; the presence of a subsurface chlorophyll maximum (SCM) and the presence of high surface chlorophyll (HSC).

SCMs were detected by identifying when the maximum chlorophyll concentration was not at the surface (< 20 m). To minimise the number of false detections due to measurement uncertainty, SCMs were only classified as present if they met one of two criteria based on chlorophyll-a concentration. For surface chlorophyll-a less than 1 mg/m3, chlorophyll-a concentrations at the SCM were required to be greater than the surface by 0.2 mg/m3. For surface chlorophyll-a concentrations higher than 1mg/m3, chlorophyll-a concentrations at the SCM had to be greater than 1.2 times the surface chlorophyll-a concentrations. A subsurface fluorescence maximum (SFM) was detected with the same criteria as a SCM, using chlorophyll fluorescence in place of chlorophyll-a.

The presence or absence of HSC was detected if chlorophyll-a averaged over 10 m layers decreased from surface to 100 m (16), and surface chlorophyll-a was greater than 0.5 mg/m3. High surface fluorescence (HSF) was detected with the same criteria as HSC, using chlorophyll fluorescence in place of chlorophyll-a.

#### Characterisation of ocean regimes

To understand the impact of ocean conditions on the performance of NPQ corrections, we classified sampling stations into four different ocean regimes.

1. The mixed layer depth (MLD) was shallower than the SFM
2. The MLD was deeper than the SFM
3. A fresh water layer existed (FWL) likely due to recent ice-melt; characterised by a water column structure where salinity and temperature is at a minimum above the MLD
4. A temperature minimum layer exiisted (TML), characterised by a temperature minimum, at least 0.2°C cooler than the temperature at 10 m and occuring above the ecological mixed layer depth (EMLD), which is common in the Southern Ocean around Antarctica (**Holm-Hansen2005?**)

For this purpose, MLD was defined using a 0.03 kg/m3 increase compared to 10 m and the EMLD is defined as in Carvalho et al. (2017).

## Results

#### Chlorophyll-informed fluorescence

The chlorophyll-informed corrections for non-photochemical quenching performed well ~~against uncorrected data, decreasing errors when~~ compared to chlorophyll measurements(??, TableS1). By choosing an NPQ depth that is informed by chlorophyll data, corrections decreased RSS, surface bias and SCM bias significatly across X12dd, S08dd and BCPdd methods. Visual inspection revealed that the chlorophyll informed methods were able to determine an ideal NPQ depth and produce a corrected fluorescence profile that aligned with chlorophyll measurements (Supplementary figures).

The chlorophyll-informed correction was able to calibrate fluorescence to chlorophyll and correct for NPQ in 90% of fluorescence profiles to within a 37% RSS and a 25% surface bias. The performance of X12dd, S08dd and BCPdd were comparable as indicated by RSSs of 8.78 (+/- 7.62) %, 8.02 (+/-6.94) % and 9.1 (+/- 7.91) % respectively, and surface biases of 0.61 (+/- 8.53) %, 0.46 (+/- 5.58) % and 0.14 (+/- 8.76) %, improved from an RSS of 19.02 +/- 18.18) % and a surface bias of 12.43 (+/- 87.74) % in uncorrected data. SCM bias was also significantly improved, with the X12dd and BCPdd methods on average underestimating the SCM by -10.9 (+/-12.87) % and -10.49 (+/-12.62) % compared to a SCM bias of 65.3 (+/- 115.44) % in uncorrected data. The S08dd method did not demonstrate significant improvement as measured by the Wilcoxon test, but it appeared to have decreased SCM bias on average at -7.05 (+/-9.12) %.

The chlorophyll-informed methods significantly increased the success of replicating a SCM, as measured by the XXX test. The X12dd, S08dd and BCPdd methods replicated 81.72 %, 77.04 % and 78.45 % of SCMs compared to 66.46% in uncorrected data. Similarly, the chlorophyll-informed methods significantly increased the success of replicating HSCs. The X12dd, S08dd and BCPdd methods replicated 71.49 %, 66.67 % and 67.26 % of HSCs compared to 60.57% in uncorrected data. Compared to the observed measure of NPQ depth (NPQdobs), the X12dd, S08dd and BCPdd methods chose significantly different NPQ depths, on average chosing a shallower depth with a mean bias of -7.3 (+/-13.91) m, -7.63 (+/-15.00) m and -8.02 (+/-13.37) m respectively.

Compared to unquenched profiles, uncorrected fluorescence profiles affected by NPQ displayed significantly higher RSS, underestimated surface chlorophyll, and had a larger SCM bias and a lower rate of SCM detection. No differences were observed across quenched and unquenched profiles for the success of detecting a HSC, detecting a SCM or the NPQ depth bias. All chlorophyll-informed corrections removed this effect of NPQ on RSS, so that comparable RSS was observed across quenched and unquenched fluorescence profiles. Significantly differences were still observed in surface bias between quenched and unquenched profiles corrected using X12dd, S08dd, and BCPdd methods, and SCM bias using the X12dd method. However, these differences were magnitudes less compared to differences observed within uncorrected data (??).

Some differences in the RSS, surface bias and ability to detect a HSC or SCM were observed when considering the performance of chlorophyll methods in the presence of HSC and SCM conditions (??). No differences were observed When a SCM was present, higher RSS and surface bias were observed within data-driven corrected profiles. All chlorophyll informed methods detected the presence or absence similarly, whether a real SCM existed or not. However, when HSC was observed, chlorophyll-informed corrected profiles displayed a higher surface error, but a decreased RSS. It appears that in the presence of HSC, the BCPdd performs slightly better than the X12dd method, with a surface error of 1.3 (+/-10.64) compared to -0.33 (+/- 10.09). Both the X12dd and BCPdd methods were significantly better at detecting a HSC absence compared to a presence. Upon visual inspection, we found that a data-driven correction using only fluorescence data was unable to reproduce surface chlorophyll trends in 12 profiles (5.0 %) with HSC. Additionally a data-driven correction using only fluorescence data led to an overestimation of surface chlorophyll in 7 profiles (3.1%) with a strong SCM, by calculating an NPQ depth that was too deep.

Although the four ocean regimes displayed significantly different RSS across both uncorrected and chlorophyll-informed methods, these effects were not marked. Similarly the X12dd method displayed a significant effect for SCM bias, but this effect was not marked. No significant differences were observed in surface biases and NPQ depth biases across ocean regimes. The SCM detection success appeared to be lower in the MLD > SFM regime, but similar across X12dd, S08dd and BCPdd methods. Conversely, the detection success of a HSC appeared to be higher the MLD > SFM regime a trend that was only similar across X12dd and BCPdd methods.

#### Optimal thresholds for NPQ\_dmax

Different sensitivities were observed between the four ocean regiems when applying a range of light, temperature and density thresholds to calculate NPQ\_dmax. This was assessed by defining the success of a method to correctly define NPQ\_dmax between the chlorophyll-informed NPQ depth and 5 m shallower than the SCM. Marked differences in success rates were oberved for density and temperature thresholds, but less so for light thresholds.

Low density thresholds had a higher success rate when the MLD was deeper than the SFM (optimal density value of 0.010 kg/m3 and 0.004 kg/m3 for X12max and BCPmax respectively) or a fresh water layer was detected (optimal density value of 0.008 kg/m3 and 0.002 kg/m3 for X12max and BCPmax respectively). Higher density thresholds were optimal when the MLD was shallower than the SFM (optimal density value of 0.076 kg/m3 and 0.060 kg/m3 and for X12max and BCPmax respectively) or when a temperature minumum layer was detected (optimal density value of 0.21 kg/m3 and 0.24 kg/m3 for X12max and BCPmax respectively). Density thresholds across ocean regiems had little overlap. Success rates were highest when the MLD was shallower than the SFM (64.8 % and 63.19 % for X12max and BCPmax respectively), and lowest when the mixed layer depth was deeper than the SFM (81.3 % and 80.6 % for X12max and BCPmax respectively).

Light thresholds exhibited stable success rates over the range of 8 - 14 %. Success rates were higest when a temperature minimum layer existed (74.3 % and 81.3 %) and lowest when a fresh water layer existed (57.1 % and 61.9 %).

The optimal density threshold yielded higher success rate compared to the optimal light threshold, when the mixed layer depth was deeper than the SFM or a fresh water layer was present. Comparable success rates were observed between optimal light and density thresholds when a temperature minimum layer was present or when the X12max method was used and the mixed layer depth was shallower than the SFM. A higher success rate for the optimal light threshold, compared to the optimal density threshold was only found when the BCPmax method was used and the MLD was shallower than the SFM. Optimal temperature thresholds only outperformed the best of either optimal modelled light or optimal density thresholds, when the BCPmax method was used in the presenece of a fresh water layer. Reflecting these trends, two combinations of optimal thresholds were implemented to the X12max, S08max and BCPmax methods to assess performance.

#### Performance of NPQ corrections

[possible subheading about SCMs? Would it be possible to arrange this section in subsections about NPQ, SCM, HSC? It’s sort of like that but there is some mixing]

Compared to uncorrected profiles, all NPQ correction methods significantly reduced the surface bias and SCM bias in fluorescence profiles (??). All methods that only used fluorescence data (X12max, X12maxL, X12, B15 and P18) also significantly reduced the RSS compared to uncorrected profiles and on-average tended to underestimate the SCM. Only the BCPmax, BCPmaxL and B19 methods were not able to significanty reduce the RSS when informed by beam attenuation data. All correction methods, both using beam attenuation and not, improved the success of reproducing a SCM and HSC oserved in chlorophyll measurements. Only the X12, P18 and B19 methods were able to calculate a similar depth to NPQdobs.

The X12 and P18 methods performed the best when considering RSS (), surface bias (), tail bias () and SCM bias (). The X12 and P18 methods were also the only methods that calculated NPQ depths similar to NPQdobs (). No fluorescence methods that used information from backscatter and beam attenuation data outperformed the X12 and P18 methods. The B15, S15 and S15eu corrections performed the worst across RSS (), surface bias (), tail bias () and SCM bias (), tending to underestimate the SCM and overestimate surface chlorophyll.

Compared to chlorophyll informed corrected profiles using only fluorescence information, no method was able to reproduce significantly similar surface bias, only the X12 method reproduced significantly similar RSS, and the X12 and P18 methods displayed significantly similar SCM biases. In this case, the success of replicating a HSC feature was significantly similar across all methods, but the success of replicating a SCM feature was only significantly similar across X12 and P18 methods. Compared to chlorophyll informed corrected profiles supplemented by beam attenuation information, no method was able to reproduce significantly similar RSS, only the P18 method reproduced significantly similar surface bias, and the X12 and P18 methods displayed significantly similar SCM biases. In this case, the success of replicating a HSC feature was significantly similar across X12, B15, P18 and B19, and the success of replicating a SCM feature was significantly similar across X12, S08 and B19 methods.

The X12 and P18 methods were the best at replicating the chlorophyll informed correction method when considering RSS, surface bias, tail bias and SCM bias using only fluorescence information () and with beam attenuation information (). The B15, S15 and S15eu methods were the worst at replicating the chlorophyll informed correction method when considering RSS, surface bias, tail bias and SCM bias with application using only fluorescence information () and with beam attenuation information (). On average, all methods tended to calculate NPQ depths deeper that those calculated using chlorophyll informed correction methods.

[possible subheading about NPQ?]

When significant NPQ is observed, the X12 and P18 methods display higher RSS that is comparable to other methods (X12max, X12maxL, B15, BCPMax, BCPmaxL and S08) and on average tend to underestimate surface chlorophyll. The S08 method also tends to understimate surface chlorophyll when significant NPQ exists, while the X12max, X12maxL, B15 and BCPmaxL method on average overestimate surface chlorophyll, and the BCPmax method on average introduces litte bias. When significant NPQ is not observed, surface chlorophyll is on average overestimated [by which methods? all of them?]. There is little effect on the SCM bias when the effect of significant NPQ is considered.

The X12, P18 and B19 methods did not display a difference in the RSS when an SCM was observed, and the S08 method displayed very little difference. Other methods resulted in large increases in RSS when an SCM was observed. All methods showed increase surface error when an SCM was observed, but the effect was smaller on the X12 and P18 methods. When a HSC was observed, all methods displayed increased RSS and surface error. Again, this effect was smaller on the X12 and P18 methods.

[Subheading about differences across regimes?]

Increased RSS was observed when the MLD was shallower than the SFM or a FWL was present, compared to when the MLD was deeper than the SFM or a TML was observed. Higher variability in surface errors was also observed across all methods when the MLD was shallower than the SFM or a FWL was present, with X12max, X12maxL, B15, BCPmax, BCPmaxL, S15 and S15eu tending to overestimate surface chlorophyll, while X12, P18, S08 and B19 displayed little bias. The effects of mixing regiems on SCM bias was variable, and X12, P18, S08 and B19 displayed the least variance in performance across mixing regiems.

#### Application on Biogeochemical Argo

• Compare bbp methods using X12 and P18 as a baseline in BGC-Argo

#### The occurence of NPQ

Using the X12 correction method, we observed NPQ in 33% of profiles. Significant NPQ signals were constrained to the daytime and November/December with sun angles > -5 degrees. Only 44% of NPQ occurred above the 0.03 density MLD and only 46% of NPQ occurred above the 15% euphotic depth. These are common depth criteria used in NPQ correction methods. No significant correlation was observed between NPQ depth and sun angle, or NPQ depth and surface chlorophyll.

[Add in BGC Argo observation of NPQ]

[Add in community pigment analysis on NPQ]

## Discussion

In this work, we provide an in depth comparison of NPQ correction methods by comparing corrected profiles to co-incident chlorophyll observations from the Southern Ocean. This analysis supplements previous work by Thomalla et al. (2018) and Xing et al. (2018), who compared day and night profiles using glider and biogeochemical Argo datasets, respectively.

We find that the X12 and P18 methods perform best, even when considering methods supplemented by backscatter or beam attenuation profiles. These methods displayed the lowest biases and RSS when compared directly with chlorophyll measurements, and were the best at replicating chlorophyll-informed corrected profiles. These two methods on average chose NPQ depths that were very close to those calculated by chlorophyll-informed methods, and within 5 m of the observed NPQ depth.

The S15 and S15eu methods performed the worst, with RSS, SCM biases and surface biases that were double those observed when using the X12 and P18 methods. The B15 method also did not perform well, displaying SCM biases and surface biases that were double those observed when using the X12 and P18 methods. Compared to chlorophyll-informed methods, B15, S15 and S15eu calculated NPQ depths that were up to on average 18.92 m, 40.03 m and 35.58 m deeper, respectively. This resulted in a larger underestimation of the SCM, a larger overestimation of surface chlorophyll and a lower success of replicating an SCM. When an SCM existed, the deeper calculation of the NPQ depth led to the supression of the SCM and HSC magnitudes.

The S08 method introduced on average up to 3 % more RSS, surface bias and SCM bias compared to the X12 and P18 methods. Results were also more variable using the S08 method, displaying an increase in standard deviation over error measures of up to 2.5 %. This is due to a deeper calculation of the NPQ depth which leads to a larger effect on the underestimation of the SCM and overstimation of surface chlorophyll.

[discussion about the S08 method and BGC-Argo - Backscatter, unless extensive QC is applied, can often have large spikes.]

The performance of NPQ corrections was variable across the four mixing conditions considered. Cases where the MLD was shallower than the SFM or where a TML was observed, displayed significantly higher errors compared to where the MLD was deeper than the SFM and where a FWL was observed. Clear sensitivities were observed to the choice of the threshold used to calculate a maximum NPQ depth (NPQ\_dmax), but optimised methods (X12max, X12maxL, BCPmax and BCPmaxL) performed worse than the X12, P18 and S08 methods which used a fixed NPQdmax threshold across mixing regions. This suggests that the choice of NPQ\_dmax is not the largest driver in variability between mixing regions.

As suggested in previous studies, we find evidence that ocean conditions affect the performance of NPQ corrections. In cases where the MLD was shallower than the SFM or where a TML was found, there was a higher proportion of both significant NPQ and occurences of SCMs. No large difference was observed in the proportion of profiles that were quenched between when a SCM was present or absent when the MLD was shallower than the SFM (SCM: 40.5 % NPQ; no SCM: 46.3 % NPQ) or when a TML was present (SCM: 29.8 % NPQ; no SCM: 33.6 % NPQ). The presence of SCMs in these mixing regions is a known phenonemon (**Baldry2020?**), but this is the first evidence of these mixing regions being linked to higher rates of NPQ. Consequently, these mixing regions are more prone to the errors introduced by the effects of NPQ and the effects of a SCM being present.

The presence of diatom communities and nutrient limitation could be a driver for the higher rates of NPQ in cases where the MLD was shallower than the SFM and a TML exists. In the Southern Ocean, it is well known that diatoms dominate surface communities when nutrient limitation exists [reference for this statement?]. [there is a missing link in the logic here. Please elaborate on why nutrient limited diatoms, as opposed to nutrient limited anything else, would exacerbate NPQ] Further, deep, diatom dominated SCMs (dSCM) are observed under strong nutrient limitation and stratification (shallow MLD) (**Baldry2020?**). These dSCMs, exhibit different community compositions to surface communities, are very thin (< 10 m) and reside at or below the MLD.

[insert pigment/community composition insights]

Under nutrient limitation, the magnitude and occurance of significant NPQ is higher (8,**RyanKeogh2021?**,**RyanKeogh2021? Schallenberg et al., 2022**). So much so, that NPQ signals measured from gliders have been found to be a strong proxy for iron limitation in the Ross Sea (**RyanKeogh2021?**). As SCMs appear to occur mostly under nutrient limitation in the Southern Ocean (**Baldry2020?**), it follows that NPQ must be higher under conditions where SCMs may form. So in conditions where a shallow mixed layer exists and the SFM is deeper that the MLD (due either to quenching or the presence of a deep SCM) higher rates of significant quenching are observed. This leads to higher rates of error, as even with the best NPQ corrections we show that up to 20 % RSS error still remains after correction and performance is variable.

The consistently greater error observed when a SCM exists, is mostly due to differences in the SCM feacture compared to the SFM feature. This can be due to changes in chlorophyll fluorescence yield with depth, differences in sampling resolution, differences in the time of sampling and differences in the depth recorded between chlorophyll and fluorescence. Changes in chlorophyll fluorescence yield likely occurr when changes in community composition are observed with depth. Mis-alignment of the SFM and SCM depth is possible, as its depth may have changed between the downcast time (where fluorescence is often collected) and the time of upcast (where chlorophyll is observed). Particularly when SCMs are thin, depth mis-alignment or the inaccurate sampling of the true SCM can lead to large differences when comparing SCMs with calibrated fluorescence. Often firing of niskin bottles misses the true SCM.

This provides a good illustration into how the overall perfomance of a data-driven correction largely depends on the quality of the chlorophyll ship data, which was highly variable within the data compilation. This is due to variable observational errors and chlorophyll fluorescence yeild with depth. During quality assurance assessments we not only identified questionable data, but observed a number of profiles with variable chlorophyll measurements that showed no distinct profile shape.

There is evidence that NPQ is not the only driver of changes in chlorophyll fluorescence yield with depth. Significant, real changes in chlorophyll fluorescence yield with depth were observed in x profiles (X%) within the data compilation and a decoupling of chlorophyll fluorescence from chlorophyll values. This raises question about the magnitude of the uncertainty of chlorophyll fluorescence as a proxy for chlorophyll, even when calibrated to coincident chlorophyll measurments. Further study, using data compilations like BIOMATE, is needed to understand this variabillity in chlorophyll fluorescence and its implications for biogeochemical models that assimilate data from fluorometers.

The ship dataset used within this study is the largest and distributed ship-based data compliation used to-date. We have shown that data-driven correction methods, provide the best results for correcting chlorophyll fluorescence profiles. The methods are semi-automated and can be applied to other ship data compolations, or used over smal datasets to correct profiles in a uniform way.

## Tables and Figures

Table : Summary of BIOMATE data used in this study

A picture containing chart

Description automatically generated

Table : Spatio-temporal distribution of BIOMATE data used in this study

Chart, diagram

Description automatically generated

Table : Distribution of ocean regimes

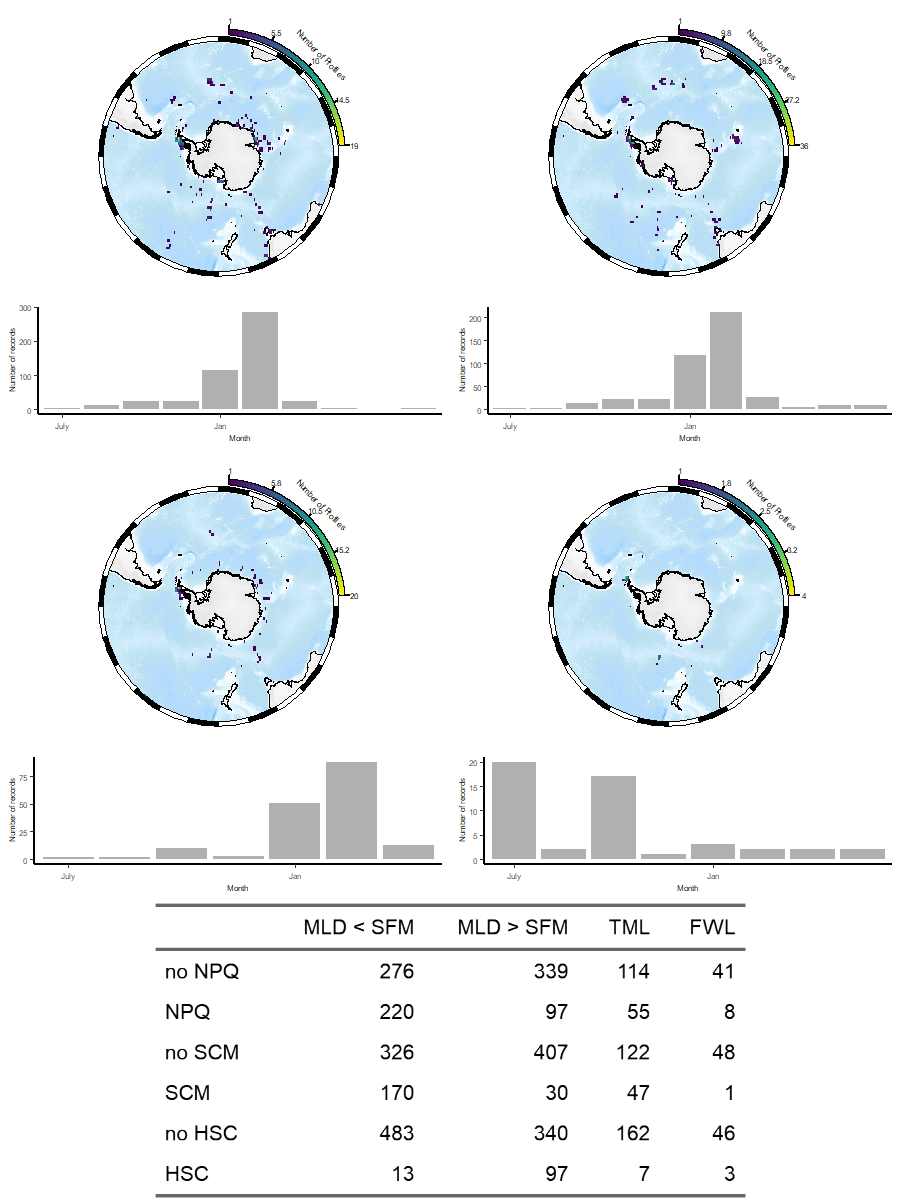


Table 2: Definition of Ocean properties

A picture containing rectangle

Description automatically generated

Table 3: NPQ correction methods considered in this study



Diagram

Description automatically generated

Figure 3: Method Figure

Table : Absolute values of error measures across uncorrected fluorescence and chlorophyll-informed corrected fluorescence compared to chlorophyll measurements.

A picture containing graphical user interface

Description automatically generated

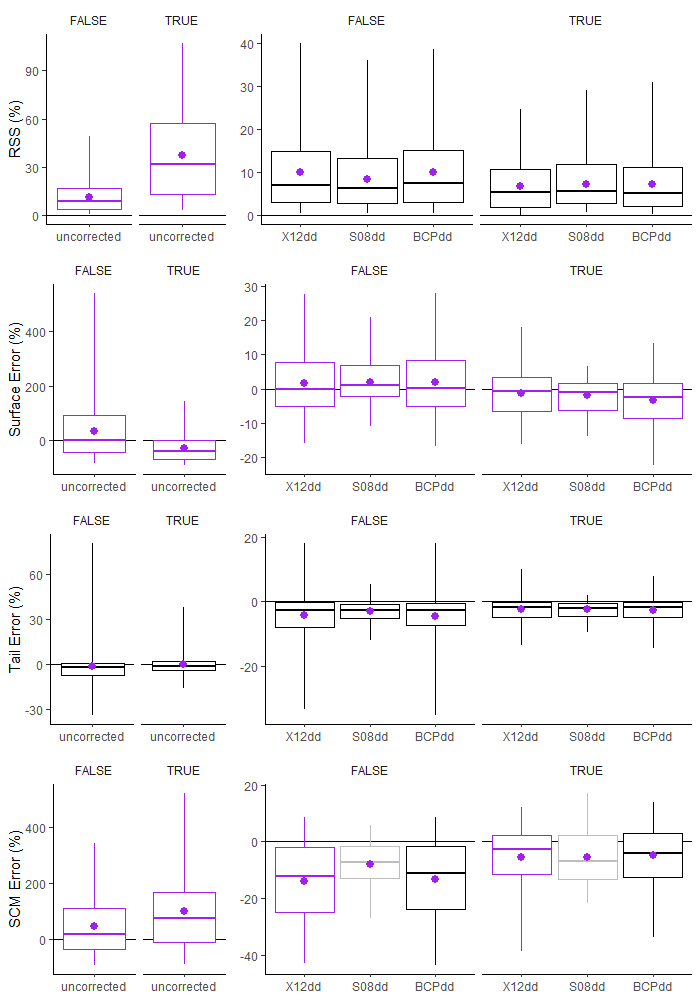


Figure 4: Absolute error measures, and the effect of the presence of significant NPQ across uncorrected and chlorophyll-informed data.

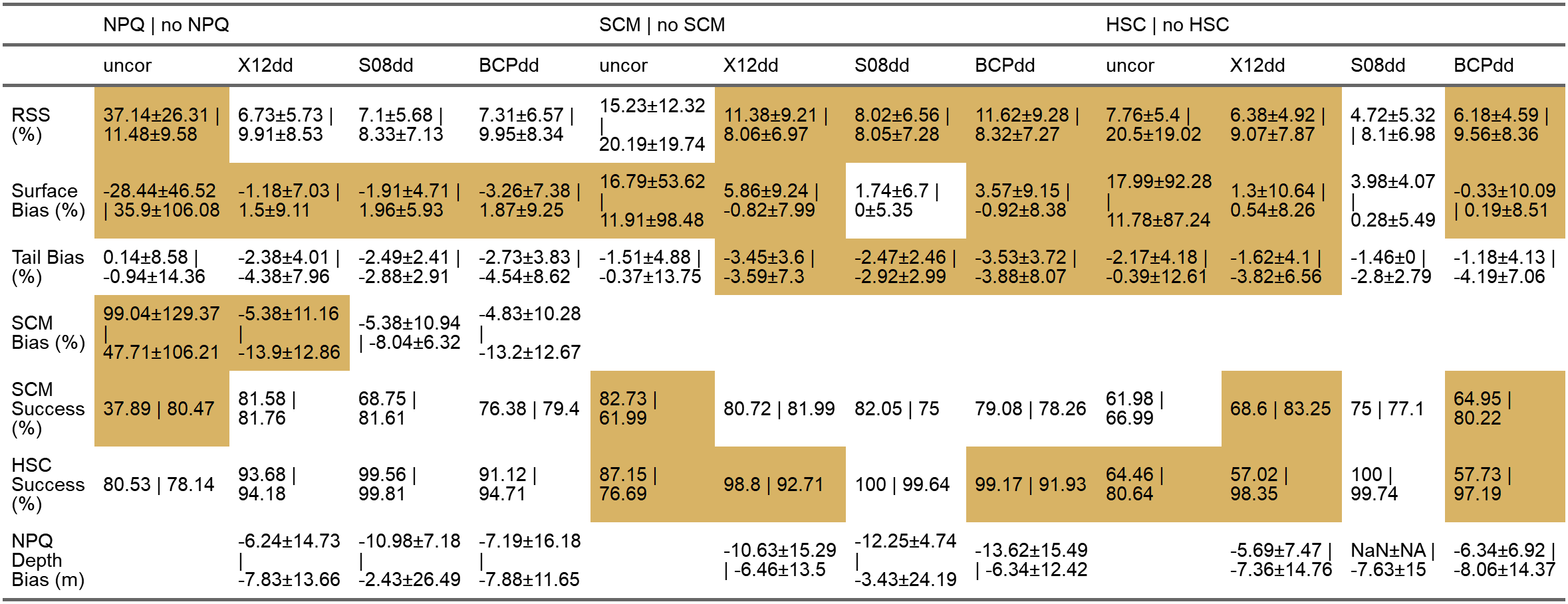


Table 5 (a and b) The effects of NPQ, SCM, HSC and ocean regime on error for uncorrected fluorescence and chlorophyll-informed methods. Shaded cells indicate a significant difference in distribution.

A picture containing timeline

Description automatically generated

Table : Optimal thresholds for NPQdmax, chosen according to the highest success rate



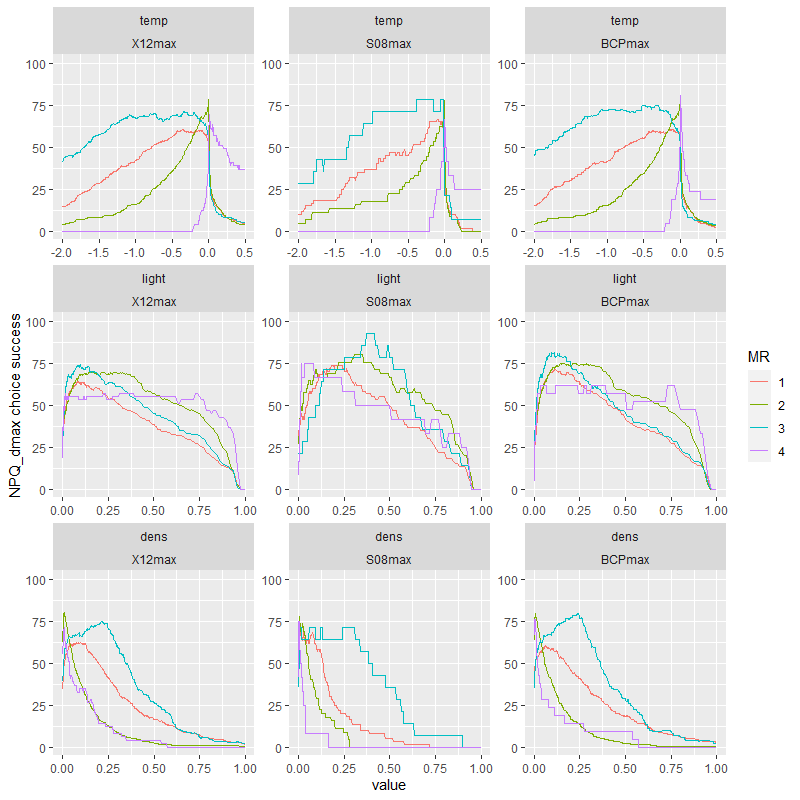


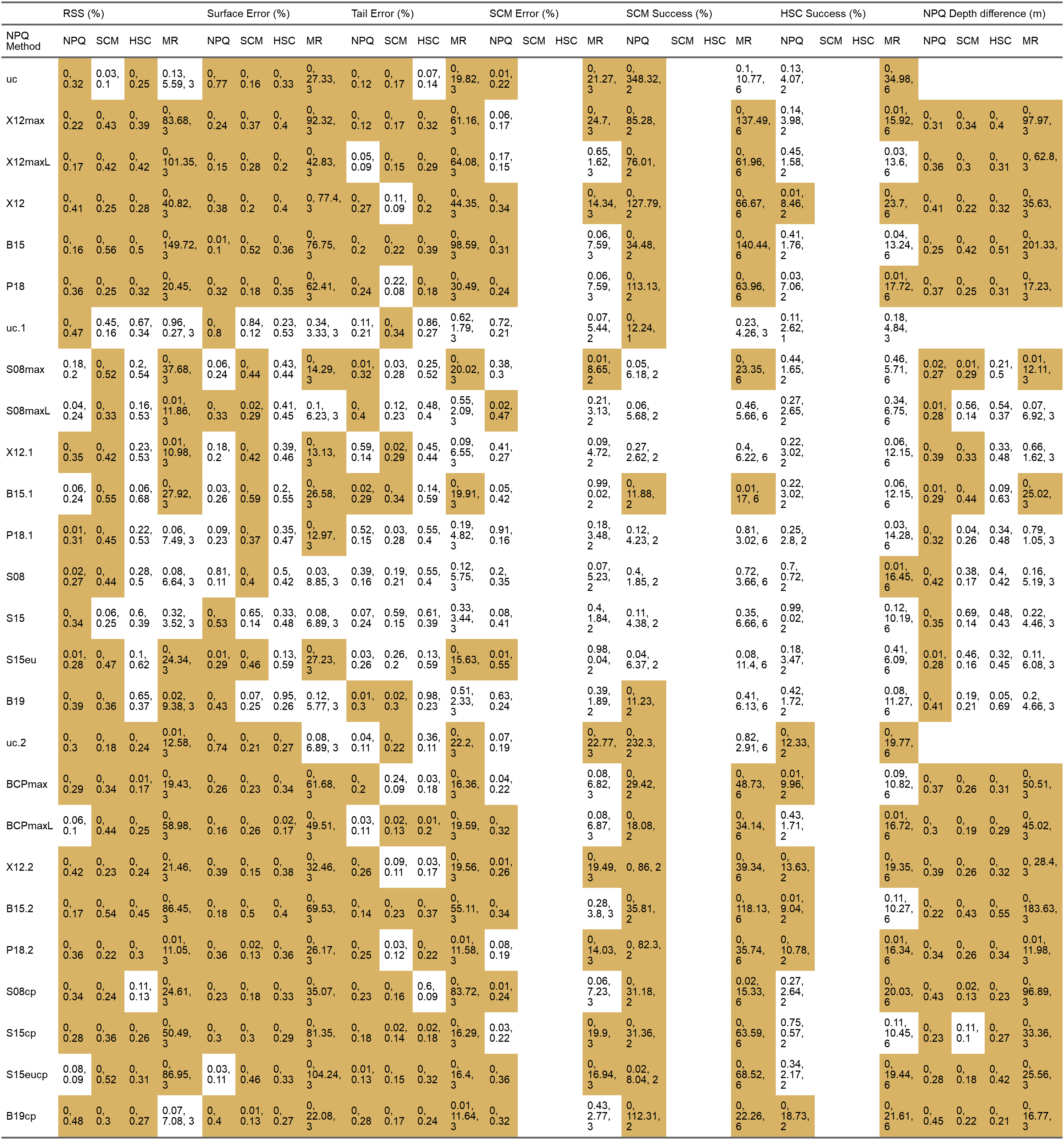
Figure 5: Threshold sensitivity test for NPQ\_dmax across light, temperature and density values

Table 6: The relative errors of NPQ correction methods, relative errors derived using chlorophyll-informed methods. Note first 6 methods are compared to X12dd, next 10 compared to S08dd and the last 10 compared to BCPdd. Shaded cells indicate significant differences using either a Wilcoxon test or a McNemars test.

Qr code

Description automatically generated

Table 7: Statistical results for effects of NPQ, SCM, HSC and ocean regime across error measures for all NPQ correction methods.



1. Thomalla, S. J., Moutier, W., Ryan-Keogh, T. J., Gregor, L. & Schütt, J. [An optimized method for correcting fluorescence quenching using optical backscattering on autonomous platforms](https://doi.org/10.1002/lom3.10234). *Limnology and Oceanography: Methods* **16**, 132–144 (2018).

2. Xing, X., Briggs, N., Boss, E. & Claustre, H. [Improved correction for non-photochemical quenching of in situ chlorophyll fluorescence based on a synchronous irradiance profile](https://doi.org/10.1364/OE.26.024734). *Opt. Express* **26**, 24734–24751 (2018).

3. Xing, X. *et al.* [Quenching correction for in vivo chlorophyll fluorescence acquired by autonomous platforms: A case study with instrumented elephant seals in the kerguelen region (southern ocean)](https://doi.org/10.4319/lom.2012.10.483). *Limnology and Oceanography: Methods* **10**, 483–495 (2012).

4. Biermann, L., Guinet, C., Bester, M., Brierley, A. & Boehme, L. [An alternative method for correcting fluorescence quenching](https://doi.org/10.5194/os-11-83-2015). *Ocean Science* **11**, 83–91 (2015).

5. Hemsley, V. S. *et al.* [Estimating oceanic primary production using vertical irradiance and chlorophyll profiles from ocean gliders in the north atlantic](https://doi.org/10.1021/acs.est.5b00608). *Environmental Science & Technology* **49**, 11612–11621 (2015).

6. Sackmann, B. S., Perry, M. J. & Eriksen, C. C. [Seaglider observations of variability in daytime fluorescence quenching of chlorophyll-*a* in northeastern pacific coastal waters](https://doi.org/10.5194/bgd-5-2839-2008). *Biogeosciences Discussions* **5**, 2839–2865 (2008).

7. Swart, S., Thomalla, S. J. & Monteiro, P. M. S. [The seasonal cycle of mixed layer dynamics and phytoplankton biomass in the sub-antarctic zone: A high-resolution glider experiment](https://doi.org/10.1016/j.jmarsys.2014.06.002). *Journal of Marine Systems* **147**, 103–115 (2015).

8. Falkowski, P. & Kolber, Z. [Variations in chlorophyll fluorescence yields in phytoplankton in the world oceans](https://doi.org/10.1071/PP9950341). *Functional Plant Biology* **22**, 341–355 (1995).

9. Schallenberg, C. *et al.* [Diel quenching of southern ocean phytoplankton fluorescence is related to iron limitation](https://doi.org/10.5194/bg-17-793-2020). *Biogeosciences* **17**, 793–812 (2020).

10. Ryan-Keogh, T. J. & Thomalla, S. J. [Deriving a proxy for iron limitation from chlorophyll fluorescence on buoyancy gliders](https://doi.org/10.3389/fmars.2020.00275). *Frontiers in Marine Science* **7**, (2020).

11. Ryan-Keogh, T. J. & Smith, W. O. [Temporal patterns of iron limitation in the ross sea as determined from chlorophyll fluorescence](https://doi.org/10.1016/j.jmarsys.2020.103500). *Journal of Marine Systems* **215**, 103500 (2021).

12. Claustre, H., Johnson, K. S. & Takeshita, Y. [Observing the global ocean with biogeochemical-argo](https://doi.org/10.1146/annurev-marine-010419-010956). *Annual Review of Marine Science* **12**, 23–48 (2020).

13. Ardyna, M. *et al.* [Parameterization of vertical chlorophyll a in the arctic ocean: Impact of the subsurface chlorophyll maximum on regional, seasonal, and annual primary production estimates](https://doi.org/10.5194/bg-10-4383-2013). *Biogeosciences* **10**, 4383–4404 (2013).

14. Carranza, M. M. *et al.* [When mixed layers are not mixed. Storm-driven mixing and bio-optical vertical gradients in mixed layers of the southern ocean](https://doi.org/10.1029/2018JC014416). *Journal of Geophysical Research: Oceans* **123**, 7264–7289 (2018).

15. Cornec, M. *et al.* [Deep chlorophyll maxima in the global ocean: Occurrences, drivers and characteristics](https://doi.org/10.1029/2020GB006759). *Global Biogeochemical Cycles* **35**, e2020GB006759 (2021).

16. Lavigne, H. *et al.* [On the vertical distribution of the chlorophyll a concentration in the mediterranean sea: A basin-scale and seasonal approach](https://doi.org/10.5194/bg-12-5021-2015). *Biogeosciences* **12**, 5021–5039 (2015).