General Response to reviewer comments:

Firstly we would like to thank all reviewers: Prof Boss, Dr Ciavatta and Dr Ford for their thorough review of our manuscript and their insightful comments. All reviewers commented on the number of formatting errors that were contained in the original submission. We'd like to apologize for this and point out that we remedied this in a version of the manuscript that is attached to the response to RC1.

We have addressed all of the reviewers comments below, but would like to take this opportunity to point out that the manuscript has undergone a major revision and now contains an extra assimilation experiment and numerous in-text changes. Not all in-text changes have been documented individually, excerpts to reviewer comments have been included where appropriate. A fully marked up version of the manuscript has been provided that shows the extent of the text modifications.

The most contentious issue that RC2 and RC3 raised, was our use of OC3M rather than the reflectances directly. There have been plenty of studies that have assimilated OC3M, and relatively few that use other bio-optical properties, but the most powerful advance that we have found was in directly coupling the BGC model with an in-water optics model to simulate a remote sensing reflectance. The key step towards assimilating remote sensing data, has been the ability to calculate a spectrally resolved simulated Remote-Sensing Reflectance (RSR), that accounts for the major bio-optically active constituents. Once a simulated remote sensing reflectance is available, there are a variety of assimilation options available.

In our original manuscript, we contest that whilst we are not directly assimilating the raw reflectances, we are using them in a consistent manner whereby there is a directly comparable quantity that is produced by the model. We are essentially assimilating like-for-like variables that are not subject to difference in kind errors. Both reviewers encouraged us to try assimilating reflectances directly. We have now done this and successfully demonstrate the assimilation of reflectance observations centered at 551 nm (R551). To this end, regardless of whether we use simulated R551 or OC3M, we are modelling a quantity that is predicted using a forward mechanistic model, rather than relying on an empirical-statistical relationship to relate reflectances with Chl-a and subsequently a model state variable.

Below we have address the reviewer comments and suggestions in "italics".

Response to Reviewer 1

Dear Prof. Boss,

We would like to thank you for reviewing our submission to Biogeosciences Discussion (BGD) entitled, "Use of remote-sensing reflectance to constrain a data assimilating marine biogeochemical model of the Great Barrier Reef." We would firstly like to apologize for allowing a series of Figure numbering and reference mistakes to propagate through into the online version of our discussion paper. Clearly these mistakes have made it difficult to appropriately assess the science contained in our paper.

Please find our response to your comments and suggestion below:

1. Chlorophyll fluorescence is strongly affected by ambient light (called nonphotochemical quenching). It seems (from your plots) that it has likely affected the glider data near the surface yet you do not mention it or a correction for it. This will result in a significant bias.

*The glider data used in the paper (Figures 7 and 8; latest version of the manuscript) shows very little evidence of NPQ, and thus we therefore did not mention it. We are certainly aware of NPQ effects and have seen NPQ in high resolution CTD casts and shallow Slocum glider missions and corrected for it in other work (**Baird, M. E.**, I. M. Suthers, D. A. Griffin, B. Hollings, C. Pattiaratchi, J. D. Everett, M. Roughan, K. Oubelkheir and M. Doblin (2011) Physical-biogeochemical dynamics of a surface flooded warm-core eddy off southeast Australia Deep Sea Res. II **58**, 592-605, which considers your work on NPQ.).*

The glider data presented in this manuscript was from an ocean glider that sampled to 1000 m. If important, we would expect NPQ to bias measurements taken in the upper 30 m, however, in almost all casts, the fluorescence-derived Chl-a concentration is uniform in the upper 60 m. It should also be noted that the glider deployments were during the Southern Hemisphere winter, which further reduces the impact of NPQ. In the revised manuscript, we have acknowledged that NPQ could bias fluorescence based Chl-a measurements, but there is no evidence of this.

2. It is obvious to practitioners that OC3M does not provide reasonable chlorophylls when CDOM or bottom reflectance contributes significantly to the signal (and in way not captured by the CDOM/chl relation in the open ocean). In addition, local algorithms (as built into your model's optics) are likely to work better than any global model (by design). There are, however, other products (the IOPs, Chlorophyll or OC3M are not IOPs) which are designed w/o assumptions of co-variability of IOPs, e.g. that reflectance spectra are only a function of chlorophyll (called, for example

GIOP). Have you tried to see if their product (when tuning the IOP shapes as you did in your model) provide you more useful outputs from Rrs to assimilate?

In this study we have not used a local semi-empirical algorithm to estimate satellite-derived Chl-a. We have used remote-sensing reflectances R_{rs} derived from a local atmospheric corrections (see Schroeder et al., 2007). The atmospherically-corrected R_{rs} are then used in the OC3M algorithm. As you rightly point out, we expect the satellite-derived OC3M results to be contaminated by CDOM, TSS and bottom reflectance. We acknowledge that this is a blue water algorithm that should not be used in coastal environments with case 2 waters to estimate in situ chlorophyll concentrations. But that is not what we are doing. We are using observed OC3M to compare a simulated OC3M (calculated from simulated R_{rs}) that also contains the simulated effects of CDOM, TSS and bottom reflectance. Both simulated and observed OC3M are equivalent and can be directly compared and assimilated.

If I interpreted your criticism correctly, other products such as kd490 would not be subject to the same errors as OC3M when CDOM, TSS and bottom reflectance is present. We have certainly considered including these products (and also R_{rs} 645 nm) in the assimilation state vector. However, for the purposes of conducting a series of experiments whereby we demonstrate the concept of simulating the observation (i.e. using the optics model of Baird et al., 2016), and exposing the problems of assuming a direct relationship between modelled surface Chl-a and observed OC3M, this precluded the inclusion of addition observation variables. Inclusion of R_{rs} 645 nm is certainly of interest and experiments such as these are underway, but beyond the scope of the present study. We have demonstrated the worth of assimilating R_{rs} at 551 nm as detailed in EXP5.

We would also like to point out that the field of BGC data assimilation is lagging perhaps a decade behind ocean hydrodynamics and two decades behind NWP, in that marine BGC DA, is still trying to assimilate observational products (such as remotely sensed chl-a) that try to convert an observed quantity, into a modelled quantity. This paradigm was dropped in the numerical weather prediction community in favor of assimilating temperature brightness, rather than using empirical algorithms to estimate temperature at individual levels of the model from temperature brightness. By applying a similar approach in marine BGC DA, it was anticipated, and subsequently found, that this is also the case when using ocean color.

3. Assimilating a single value out of 6bands of reflectance (the OC3M) seems too limiting, particularly wrt TSM who is usually inverted from magnitude of reflectance (rather than band ratio). This can give you another and independent information to constrain your model with (particularly near the coast where sediments become an important part of the model).

This is a very valid comment and we would agree that by including a direct measurement of R_{rs} at 645nm in the assimilation state vector, then we would very likely increase the information content derived from the observations. However, as stated above, our intention was to demonstrate the concept of simulating the observation in this study. In preliminary work (not shown), pairwise plots of individual R_{rs} bands plotted against each other, show very high degrees of correlation. Therefore adjacent bands do not contain substantial new amounts of information, for example by assimilating 6 ocean colour bands, we do not expect

to increase the information content by a factor of 6, compared to assimilating some functional form the band ratios.

There is also the additional theoretical complexity of requiring the assimilation system to include off diagonal elements within the observation error covariance matrix if correlated observations are to be assimilated. To date, it has been difficult to get estimates of uncertainty for individual reflectance bands on a pixel by pixel basis. To the best of our knowledge, there has been no studies that adequately provide estimates the spatial covariance structure between ocean color bands (on a pixel by pixel basis), and therefore we have chosen address this problem by using a band ratio approach in the first instance.

Response to comments on annotated PDF:

Page1: Authors seem unaware of problem using fluorometers to measure chlorophyll, e.g. Cullen 1982 and literature since. In particular the issue of non-photochemical quenching.

This has been addressed in our response above.

Page 1, Line 27: I think that what you mean here is that rather than using global semi-empirical algorithm you find locally tuned empirically tuned algorithm to work better.

We make the point here that by assimilating "like-for-like" we can use OC products in optically complex coastal water.

Page 3, line 3: I am afraid you are mixing between global and local algorithms, on top of attackiing the global algorithm for being used in regions for which they were not designed, e.g. shallow and coastal waters where their inherent assumptions fail.

We have acknowledged in multiple places that OC3M was not designed for use in complex coastal waters. It is not our intention to attack the global algorithms nor advocate for their use in areas they were not designed for. Nonetheless, we show there is value is using a global algorithm in the coastal zone for assimilation purposes so long as the simulated product contains the same difference in kind errors as the observation. We have added text in multiple locations throughout the manuscript to make this point. Furthermore, we have removed "error-prone" when discussing the global algorithms.

Page 3 line 26: as long as you used local data in the training, you de-facto constructed a local algorithm. Your exercise is designed to interpolate between conditions that have been observed. This is perfectly fine, but needs to be acknowledged.

There has been a misunderstanding, as we do not use local data to train the system, nor do we use a local algorithm. It is true that we use both observed OC3M and Rrs 551 to constrain our assimilation system, but we then undertaken an independent assessment against in-situ observation that have not been used in the assimilation system.

Page 5, Line 14: Note that computing Rrs in BGC models has been done before, e.g. Fujji et al., 2009 and other works co-authored by Fei Chai, Works by Dutkewitz, among others.

We have added extra references and a short discussion

Page 5, Line 33: FYI: you are assuming the attenuation up and down are equal. They are not. The mean cosine for Ku is often assumed to be 0.5.

We don't disagree that there could be an improved method, and our eventual goal is to use a more advanced physics based radiative transfer model. However, there is precedence for using exp(-2K) to depth weight the surface expression as mentioned in Baird et al., (2016):

"Note that the weighting of the surface expression of an IOP based on twice the vertical attenuation rate has been used in shallow-water semi-analytical reflectance models to consider the relative impact of water column and benthic constituents (Lee et al., 1998) and for considering the surface expression of depth-varying chlorophyll concentration (Moline and Prézelin, 2000)."

Further, u itself is based on backscatter which for pure water is the 0.5. So the error of our approach is in the depth-weighting of the surface expression, not the magnitude.

Nonetheless there are clearly approximations in our approach but these do not result, when compared to MODIS observations, in overly large errors in estimating Rrs (see Baird et al., 2016b).

Response to Reviewer 2

We would like to thank Dr Ciavatta for his thorough review, we have address his specific comments below in italics.

This work aims to demonstrate that assimilating satellite-derived remote sensing reflectance into biogeochemical-optical coastal models is better than assimilating empirical statistical products from satellite remote sensing, e.g. chlorophyll. To achieve this objective, the authors assimilate "super-observations" that were computed from remote sensing reflectance by using the OC3M algorithm and a set of empirically determined coefficients. The best model performance (i.e. OC3M forecasts) was obtained when assuming that observed OC3M represents a model OC3M diagnostic variable, rather than the sum of simulated chlorophyll. The assimilation system improved the simulation of independent in situ data (nitrate, ammonia, dissolved inorganic phosphorus and total suspended solids), as well as chlorophyll fluorescence from a glider, when compared to the model without assimilation (control).
Assimilation of remote sensing reflectance has the potential to be a crucial break- through in the area of marine ecosystem modelling. However, in my opinion, this work has some relevant issues and a major revision is needed before its publication. In addition, the editing of the manuscript definitively needs an improvement, though I am assuming that the numeration of the figures is missing because of the journal web system.
I feel that the methods and results are not supporting fully the objectives and conclusions of the present version of the work. The objective is to demonstrate that assimilating reflectance is better than assimilating empirical functions of reflectance (e.g. chlorophyll, Kd, PFTs). However, the authors assimilate OC3M, which is an empirical function of the reflectance ratios. I think they should have assimilated the reflectance directly. I acknowledge that the authors discuss this choice extensively, but I am not convinced by their arguments so far, for the following reasons.

A. Firstly, they assimilate OC3M, rather than reflectance, because "the relationship between individual state variables and remote-sensing reflectance is at time non-linear, thus violating the [sic] one of the underlying assumptions of the DEnKF." (page 45). If I am not missing something, the relationship between individual state variables and OC3M is also clearly non-linear, but still OC3M is assimilated. The authors add: "In contrast, the relationship between simulated and observed OC3M is linear": this is not "contrasting" or relevant, because the relationship between simulated and observed reflectance is also linear. Further doubts are casted by the conclusion (see point 8 below). I am not a DEnKF expert, but if it is based on the EnKF, the non-linear relationship between observed variable and other state variables is acceptable in practice, otherwise the authors could not have assimilated OC3M, and we could not even assimilate chlorophyll in models of normal complexity. Thus, direct assimilation of reflectance data should be possible.

Thank you for pointing this out, this was poor language on our behalf and needs clarification. We agree that there is likely to be non-linear relationships between the

observed and unobserved state variables, and the linear assumption is at best an approximation that is known introduce error. There have been various approaches to reduce errors associated with the linear Gaussian assumption (e.g. Gaussian anamophosis etc). The point we would like to make here is related to errors that arise due to what we are modelling (or including as an observed variable) is different to what we are observing. For instance if we make the assumption that surface total Chl-a is equivalent to OCM3 Chl-a, then we have two error sources that must be accounted for, 1.) model error and 2.) difference in kind error, and 3.) compounding this, the relationship between surface Chl-a and MODIS OC3M is not linear. We have highlighted this by including an additional figure (Figure 5). The sum of these can be large and is generally account for by inflating the diagonal elements of the observation error covariance matrix. If we can remove the difference in kind error, then we only have representation error and model error. We have included a brief discussion of these error types in the introduction. We also acknowledge that previous studies have used IOP's and AOP's in the assimilation system and have included additional references to these studies.

We have included an additional experiment where reflectance at 551 nm is assimilated (EXP5), and is shown to be comparable in skill to EXP4.

B. A second reason why the authors prefer assimilating OC3M rather than reflectance is the cross-correlation between observation errors of reflectance bands. However I think they should try to tackle this issue and assimilate reflectance bands. I fully acknowledge the importance and challenge of dealing with the above cross-correlations, which have not been addressed in biogeochemical assimilation so far. However, Numerical Weather Prediction systems (NWP, a point of reference for the authors) are tackling them, for example by inflating diagonal observational errors or by using more complex approaches (e.g. Weston et al., 2014 at UK Met Office). Tackling this issue is necessary in a work that advocates a "third novel approach to biogeochemical data assimilation, i.e. the assimilation of remote-sensed reflectance" (page 17).

*We agree that this point needs to be tackled, but feel that this is well beyond the scope of the current study. We have used univariate observations with spatial super-obing and increased of the diagonal of **R** to account for non-zero off diagonal elements. We recognize that this is sub-optimal, but the content of this manuscript should be seen as an example of assimilating reflectance, and we point to a number of issues that need to be resolved in future studies.*

C. Finally the authors motivate assimilation of O3CM rather than the 3 reflectance bands because of computational costs. They do not describe their computing facility and usage in the manuscript (not even which year(s) they have simulated), thus I cannot really comment on this point. In any case, I think spatial resolution or simulation window should be sacrificed if the objective is to propose a novel paradigm for biogeochemical data assimilation. Another option would be to use data assimilation approaches that are computationally cheaper, such as the SEEK the SEIK to stick on Kalman filtering (see e.g. Nerger et al., Tellus, 57a, 2005). Despite their

limitations, such approaches have been proved useful in several applications with marine models (see e.g. the papers by Triantafyllou and colleagues).

Additional material has been added to various points of the manuscript to detail the system we are running on, we have specifically added detail to section 2.5.2 detail the computational system. There is substantial activity associated with improving the efficiency of our DA system, and our ensemble approach has been recognized as an important step towards providing GBR managers with an improved management tool. We anticipate receiving a larger allocation in the next round on the super computer. However, the focus of this study is on developing a new DA approach and the eventual operationalization of the system, to this end we will have a much larger ensemble. We recognize the limitations of our current system and have explicitly acknowledged them in the manuscript text. Given that we are getting stable results and a substantial improvement over the control run, we believe that we have adequately demonstrated a "proof of concept". We deliberately chose not to adopt a cheaper approach as we know that given the dry/wet season cycle, using a static background error covariance would lead to poor results. Additionally, the operational system will be using an EnKF so we used this as an opportunity to develop some of the required tools to implement an operational system.

In essence, I think that the authors should do a further effort and assimilate reflectance directly, rather than OC3M. I do recognise that this implies a lot of work, but I feel it is necessary given the objective of the work. Crucially, this would make the work a milestone in biogeochemical data assimilation.

We would like to thank Dr Ciavatta for prompting us to make the additional step, and we have included the results of an additional experiment whereby remote-sensing reflectance at 551nm is assimilated. The results look very encouraging and in the future we will include additional bands as we increase our ensemble size.

Other issues:

1) The authors call OC3M in eq. 21 a super-observation, referring to the works of Cummings et al., 2005 and Oke et al., 2008. However, in those works, "super-observations" derive by super-obing, i.e. by spatially averaging observations to reduce their number, as wells as to deal with observation error correlations. In my opinion, the authors are doing something different. They are computing a nonlinear prognostic variable that is included in the augmented state vector and control variable space. This allows them to simplify the observation operator to a direct mapping of computed to observed OC3M (see Evensen2003, Ocean Dynamics 53, for such an approach with the EnKF). However, I was not any more sure after reading the conclusions (see my point 8 below). Can the authors clarify?

We have possible allowed for a broader definition of "super-obing" than we should have. We have therefore changed the description of our method, and clarified our pre-processing of observations. Reviewer 3 also raised similar concerns and we have responded more fully in the response to the RC3 comments.

2) The current set-up of the assimilation experiments 1-4 does not provide a fair assessment of OC3M versus chlorophyll assimilation, in my opinion. It is known that OC3M may overestimate chlorophyll in coastal waters because of TSS and terrestrial CDOM. In experiment 1, observed OC3M is compared with simulated chlorophyll, and just chlorophyll classes are updated in the analysis: experiment 1 failed. In experiment 4 observed OC3M is compared with simulated OC3M and used to update chlorophyll classes as well TSS and nutrients. I suspect that experiment 4 outperformed experiment 1 because it had the chance to correct TSS. What if also TSS and nutrients were updated in the analysis in experiment 1? The authors should present such additional experiment, using the same observational error than in experiment 4.

We agree with Dr Ciavatta that for a fair comparison between EXP1 and EXP4, additional variables should be included in the assimilation vector. It was not our intention to draw a detailed conclusion between EXP1 and EXP4. The experimental design was constructed to allow for a direct comparison between EXP1 and EXP2. We agree that we would need to include further experiments to differentiate between the effects you mention above. We have included additional comparison of EXP1 to glider data and found that it performs poorly compared with EXP4 and EXP5. Given that we ran additional experiments to include the assimilation of reflectance (EXP5) we were not in a position to rerun EXP1 with the modification suggest above. Further justification for the experimental design is given in section 2.5.2.

3) Related to the above point: it is not clear how the authors computed the total chlorophyll that they compared to OC3M in experiment 1. Does it include only Trichodesmium, small and large chlorophyll? Can the author assume that benthic microalgae, corals, macrophytes do not contribute to ocean colour? Furthermore, TSS does not appear as a model variable in Figure 15: is it the sum of more than one model variable?

That is correct, surface total Chl-a is a diagnostic variable that is the sum of Tricho, Small and Large Phytoplankton is the surface layer of the model.

We have clarified this by adding the following line to the experiment 1 description, " Surface Total Chl-a is the sum of Trichodesmium, Small and Large Phytonlonton. Other variables benthic microalgae, corals and macrophytes are sub-surface and are not included in the calculation of surface total Chl-a. In optically shallow waters, these variables will influence the surface reflectance, however, EXP1 has been set up ……"

4) An appendix summarizing the bio-optical optical model should be included in the manuscript. In my opinion, the current appendix describing the biogeochemical model could be deleted because the scheme in Figure 15 is sufficient (but please renumber the figures in the order you cite them in the text). On the other hand, one needs to read Baird et al., 2016b to find crucial information to appreciate the current manuscript, e.g. on the representation of CDOM in the model.

We have renumbered the figures.

We are open to suggestions from the editor at this point. Our preference is to remove a large portion of Appendix A, but we feel that some of the subtleties of this model are best reported here. On the other hand, Baird et al., (2016b), should almost be considered a companion paper, and to include a distilled version in this manuscript would increase the length considerably (we are already at ~14,000 word and 17 figures).

5) The authors should include a larger number of critical variables in the control variable space, i.e. in the state vector that is updated in the analysis. In particular, the most important optically-active components in figure 15 should be included. I do appreciate that CDOM absorption is computed from a regression with salinity (Baird et al., 2016b), and salinity should not be updated by OC3M assimilation. However CDOM absorption dominates the signal in figure 2, as stated at page 9. I think it would it be worth to use a passive tracer instead of salinity in the regression for CDOM and exploit reflectance assimilation to correct the tracer. Furthermore, also organic components of NAP should be added in the control variable space. The authors justify the small number of control variables with computational costs, but see my comment C). In any case, it seems more sensible to include optically active compounds rather than nutrients in exp-4, if a selection is needed. I am rising this point because I think the strength of assimilating reflectances (and bulk optical properties in general) is their stronger covariance with a larger number of model variables, if compared to chlorophyll. Such strength is lost if most of the optically-active compounds are not included in the control variable space.

This is a good point, and we have included most of the optically active constituents in the assimilation vector with the exception of bottom type(which we consider to be fixed in time and therefore not dynamic) and CDOM (see below). The focus of our next study is a coupled physics-BGC system, so that errors in the physical state can be included. This will therefore include the CDOM component via the CDOM/salinity relationship. This additional step is well beyond the scope of this study, but we do recognize, as your rightly point out, that excluding some optically active constituents may bias the system in some regions. Nonetheless, our results indicate that by including the subset of variables used in EXP4 and EXP5 the assimilation system substantially reduces errors when compared with the non-assimilating model.

6) 36 members in the dynamic ensemble sounds a quite small number compared to 100 often used with the EnKF and marine models. I understood that such choice is related to the small number of control variables, and in turn to computational costs (but see my point C above). Can the author show the sensitivity of the results to the number of members, at least for one analysis cycle?

We know that 36 members is a small ensemble size, however we are getting good results out of the configurations used in EXP4 and EXP5, with uni-variate observations, that beat the performance of the non-assimilating run. Our future operational system will use a much larger ensemble and a larger ensemble will improve the results and stability of the filter.

7) The dataset used for the assessment of the simulation skill is relatively small: e.g. 3 or maximum 4 data of nutrients at each station and depth (why so few if the IMOS sampling is

monthly?). With such a low number of data in time, there is the risk that improvements are not statistically significant. This doubt must be avoided in a work that presents a new paradigm. Can the simulation be extended, or performed for a different period that is covered better by data?

We agree with your comments and would like to also point out that these stations are all located close to the coast. The period simulated is considered one of the most data dense. Unfortunately in a region of high conservation, economic and social importance there is a very limited in-situ observing system when compared with other international coastal regions. This makes us almost entirely reliant on remotely sensed observations for assimilation and assessment purposes, which justifies our approach to improve the assimilation of remotely sensed observations.

8) In the conclusion, the authors wrote "The non-linear observation operator in the assimilation system subsequently converted remote-sensing reflectance into a simulated OC3M approximation of Chl-a". This contradicts what stated in the Methods and reported in the above point A), where OC3M is described as a prognostic variable computed in the model and included in the augmented state vector and control variable space. If a non-linear operator was used in the analysis instead, several issues would arise. Please clarify what has been done in this work.

This was a mistake, the RSR at 443, 488 and 551 has simply been converted into a univariate OC3M observation, and in light of the new results from EXP5 we have reworded the conclusion:

"In this study we have used a spectrally-resolved optical model coupled to a BGC model to simulate the remote-sensing reflectances centred at the MODIS ocean colour bands. A series of assimilation system configuration experiments were undertaken to test the assimilation system performance. When the simulated OC3M (EXP4) and remote sensing reflectance (EXP5) was assimilated into the model, the domain-wide forecast errors in Chl-a from fell from 100% to 55% when compared to the non-assimilating model. By using a functional derivation of the remote-sensing reflectances (OC3M), information from multiple bands are included in a univariate observation and the forecast error is halved compared to simply assuming the OC3M is directly related to the model prediction of surface total Chl-a. A comparison against in-situ observations of NO_3, NH_4, DIP and TSS shows the assimilating model (EXP4) reduces the MAPE from 90% to less than 20% at most stations. By using a forward model that includes a majority of error sources present in the observed OC3M, we have shown that the assimilation of remotely-sensed products in optically complex case 2 waters can be achieved, and adds substantial predictive skill when compared to the non-assimilating model. Furthermore, this approach can be generalized to non ocean–colour specific missions by assimilating the reflectances directly (e.g. EXP5), liberating a vast quantity of data that cannot be used using traditional BGC assimilation systems."

I have added further comments in the pdf of the manuscript, but I neglected the typos, letting their correction to the copy-editing office. I am now realizing that my review is extremely lengthy and I will be happy to elucidate the points where I was unclear. I stress

that I really appreciate the novelty of the work and I believe that its eventual publication will represent a crucial contribution to marine system modelling and prediction.
Kind regards, Stefano Ciavatta

Comments originating from marked up PDF:
There were a number of comments attached to an annotated PDF, we have responded to the major comments/suggestions below.

Page 7: line 40: 0.012,1 are the mean and variance of the natural logarithm of the parameter values, aren't they? Please justify these values.

Additional text has been added: "... and respective means of 0.012 and 0.007 are typical values as used in the control run (Baird et al., 2016b), and given relatively broad standard deviations of 1 for both parameters."

Page 8: Line 15:

OC3M is actually a dimensionless quantity, but if used literally it is meant to estimate Chl-a. We have tried to clarify our approach to reduce confusion at multiple places in the revised manuscript.

Page 8: lines 30-35
This has been rewritten in light of the new results from EXP5.

Page 8: line 45: please write explicitly the model components of such total chlorophyll

Done

Page 9: line 13: spell this notation: it's the first time it is used in the text

Done.

Page 9: line 37: Thus CDOM should be included in the control variable space

Discussed above.

Page 10: line 34 crashed? Why? Which variables?
Extra detail has been added. The bias in the observations was causing high PhyS and PhyL biomass, this drew down the Inorganic nutrients to very low values in the surface cells causing very rapid turn-over of the available nutrients. Hence the 4/5 integrators were taking progressively smaller steps, and the runs were becoming very slow and inefficient due the numerics.

Page 11: line 9 and 19:

Details of persistence is now given.

Page 13: line 4: I don't see bloom in the data: can you please clarify?

An area adjacent to Papua New Guinea is now mentioned in the manuscript.

Page 13: line 40: such analysis has been done in this work (please show the results) or in a previous paper (please cite)?

A reference has been added to support this statement.

Page 14 line 17: I can't understand very well. Why the data of two cells are represented in each plot? What is the meaning and how is it computed such unresolved variability?

Extra detail is given in the text. Each of the subplots do not represent adjacent cells, but rather the time window for which glider measures are aggregated (e.g. over a 2 hr period) means that the glider may sample "adjacent model cells".

Page 14: line 45

There are only 3-4 samples as the simulation period only spans 3-4 months. Extra detail is given in the updated text of the manuscript.

Page 15: lines 15-25

Additional references have been added

Page 17: lines 15-25:

We have substantially reworded the discussion to remove confusion and add in additional references here and in the intro.

Table 1: please justify these values of the errors

Additional justification added to section 2.5.2.

Figure 5: Plots in figures 5-7 could be put in a single figure to make easier the comparison.

We would prefer the keep figures 5-7 separate and in landscape orientation (see comments above).

Figure 6: Here the innovations are large, but the increment of OC3M is small. That's because the increments of Phy_S and Phy_L are opposite is sign. Can the author comment this in the manuscript?

We don't necessarily agree with Reviewer 2 in this case. There are areas where the innovations are large and increments small, and vice-versa. There are also many cases where the increments for large and small phytoplankton are both positive or negative. When we look into the ratio between the small:large phytoplankton between the control and

assimilating run, there is not a strong indication that that the assimilation system is changing the balance. We agree though that this would be an interesting line of inquiry, but beyond what we would wish to include in the current version of the manuscript.

Final Comments:

Again, we'd like to thank Dr Ciavatta for his comments as they have substantially improved the manuscript and prompted us to extend the work to the direct assimilation of reflectance data.

Response to Reviewer 3

Reviewer: D. A. Ford (Referee) david.ford@metoffice.gov.uk

Received and published: 6 June 2016

This study assimilates ocean colour data into a coastal biogeochemical model, and aims to show that using remote-sensing reflectance in the assimilation provides better results than using chlorophyll. This is done by applying the OC3M algorithm to both observed and modelled remote-sensing reflectance, rather than simply using the chlorophyll concentration from the surface level of the model. The study concludes that the former approach gives better results than the latter, and that the resulting assimilation improves model results compared with independent in situ observations, and gives skilful forecasts. The paper provides useful and novel results, but I'm not convinced that the authors have achieved what they claim to have done. The paper claims to have assimilated remote-sensing reflectance, and shown this to be superior to the assimilation of chlorophyll. However, unless I've misunderstood, what's actually been assimilated is still chlorophyll, just with a more intelligent construction of the observation operator than has been used in other schemes. Whilst still novel, important, and of interest to the BG community, the work does not seem to be quite as ground-breaking as the paper claims. It might be that I'm not doing the paper justice, but in that case the methods need to be much more clearly explained. I will go through the paper providing my comments, before concluding with my recommendations. I'll note here that the manuscript could have done with proofreading before submission. There are typos throughout, some of the explanations are unclear, and figure, table and section references are often missing or incorrect. Rather than going through all of them in this review, I'll instead recommend that the authors give the manuscript a thorough proofreading before resubmission.

We'd like to thank Reviewer 3 for his insightful comments and thorough review. We agree that in the first instance the title and results could be considered confusing if some of the subtleties of our approach were not immediately appreciated. We have addressed this in our initial response to RC2 via AC2. Furthermore, as detailed below we have also undertaken additional experiments and included an additional experiment in the MS (EXP5) that does assimilate the raw reflectance data (551 nm). We'd also like to take this opportunity to apologize for introducing editorial mistakes in our initial submission. Once it became clear that there were problems with the initial manuscript, we submitted a revised version as an Author Comment (AC1) in response to RC1. Below we have addressed all the specific suggestions raised by Reviewer 3. In most instances we have included excerpts of text that have been added to the MS, but we would like to point out that the manuscript has undergone a major revision and text has been altered in many places to reflect the suggestions made by all reviewers. To this end, we have included a manuscript that has been "marked up" to make such changes obvious.

Specific comments: Section 1 - The introduction would benefit from a clearer description of the ocean colour processing chain, where different products (e.g. Rrs, Kd, Chl-a, etc) fit into that, and what their errors are. The introduction should also include a brief discussion of previous ocean colour assimilation studies, especially those that have assimilated products other than chlorophyll, such as Shulman et al. (2013) and Ciavatta et al. (2014).

We have added additional text, references and a figure to the introduction; see marked up manuscript:

"Satellites record the top of atmosphere irradiance, not chlorophyll-a (Chl-a; or other modelled variables) directly. To attain estimates of Chl-a concentration or other variables, the spectrally resolved top of atmosphere irradiance is converted into atmospherically corrected surface reflectances, which are then related to Chl-a concentration via empirical statistical relationships derived from in-situ observations. Figure 1 graphically demonstrates the key steps in the OC processing chain."

Section 2.2 – state here that you convert to OC3M yourself, even if the details are still given in a later section. What's the time resolution of the data?

Thankyou, done. Additional information has been added to the methods section. We use daily images and have indicated this now.

Section 2.3 – a lot of information is given in the appendix, but you don't seem to refer to it anywhere.

Thankyou for pointing this out, we have now included a reference to this.

Section 2.4 – state explicitly that a is absorption and bb is backscattering, and what z0 and z1 represent.

Thankyou for pointing this out, we have clarified the symbols used.

Section 2.5.1 – please ensure that notation, italicisation, etc. is consistent, describe the localisation operator, comment on whether 36 ensemble members is sufficient, and whether log-transformation is sufficient for normalisation. Please also clarify what time window you use – 5-day window, 1-day window every 5 days, are 5-day composites or daily data assimilated? Model formulation, model physics, fluxes, etc. also introduce sources of error.

We have made the following comment in section 2.5.1:

"*We acknowledge that a 36 member ensemble is small, and there is further comment on the ensemble size included in the discussion."*

"The application of the log-transform to biological variables is discussed in Parslow et al., (2013), and is commonly used in other BGC assimilation schemes (e.g. Ciavatta et al., 2014)."

Further detail has been added to the manuscript to clarify the time windows used:

"*The assimilation system iterates through time using a five-day forecast length. The assimilation system is cycled by calculating the analysis fields at time t, using the forecast from the previous cycle, $\boldsymbol{X}^f(t)$, and observations $\boldsymbol{Y}(t)$ at time t, using all observations that fall within a window of t +/- 3 hrs. The numerical mode is initialized using the analysis fields $\boldsymbol{X}^a(t)$ and the next five day forecast is made. This forecast at t+5 days, $\boldsymbol{X}^f(t+5)$, is then used in the next assimilation cycle."*

We have added a reference and brief description of the localisation operator:

"Where l is the localization operator, applied in the form of covariance localization (Sakov and Bertino, 2011), with an isotropic localization radius of 60km. "

Section 2.5.2 – I can see why the term's been used, but "super-observations" typically refer to spatially aggregated observations, rather than a variable transformation, so this is potentially confusing.

To avoid confusion, we have altered the text such that we only use "super-obing" to refer to spatial aggregation.

By applying the OC3M algorithm to the observed remote-sensing reflectance, you're essentially just creating a standard ocean colour chlorophyll product, is this right? I don't understand how this is any different to just assimilating a standard off-the-shelf OC3M chlorophyll product from (e.g.) NASA, which is what you're trying to get away from?

There is a subtle difference in our approach for Experiments 1-4. In Experiment 1, we assume that there is an equivalent relationship between simulated surface Chl-a and observed OC3M (calculated using a locally tuned atmospheric correction, which has a lower error than the NASA reflectances). In experiments 2-4, we use our simulated reflectances (which account for the Chl-a distributed within an optical depth of the surface, and also includes other optically active constituents and bottom reflectances, to derive a "simulated OC3M", which is directly equivalent to the observed OC3M. We are trying to get away from the assumptions used in EXP1, and use an optical model to compare like with like. We have also included an additional experiment that directly assimilates reflectance (EXP5). We have made substantial changes to the first 2 paragraphs of section 2.5.2 to clarify this. The objective is to reduce the "difference in kind error". A short discussion of the error types has been included in section 2.5.2.

Furthermore, whilst this has been discussed, I'm not convinced about the use of OC3M rather than a regional algorithm that's more appropriate for the region. This should either be justified more strongly, or a regional algorithm compared as well. Both here and later on, the discussion of what is and isn't linear, and the extent to which this matters, needs to be clearer.

An additional figure has been detailing the non-linear relationship between simulated OC3M and simulated Chl-a and the likely effect in the forecast innovations. We have changed the text to read:

"the relationship between individual state variables (e.g. Simulated Surface Chl-a) and observed OC3M is at times non-linear, thus violating the one of the underlying assumptions of the DEnKF"

As pointed out below, whilst there is data from a regional algorithm available, we do not consider it to be a long term objective to assimilate it. Due to requiring a very expensive field campaign to develop such regional algorithms, the eventual objective is to use the multispectral reflectance data, thus circumventing the need to develop regional algorithms for assimilation purposes.

You state that using super-observations "eliminates the possibility of cross-correlation". It will do so between the bands, but there will still be cross-correlations due to nearby observations having similar error characteristics.

We have included additional text to clarify our approach:

"Using simulated and observed OC3M eliminates the possibility of cross-correlation and contains information derived from multiple bands. The OC3M algorithm can be considered a band-ratio function, f(RSR), that transforms a multi-variate observations into univariate observations. It is likely that there are other function forms that could combine information from multiple bands into a single non-correlated observations."

The methodology of the different experiments should be explained in this section rather than in the results section.

In section 2.5.2 (Assimilation system experiments and configuration, we have included a description and rationale behind the choice of assimilation system configurations:
"Experiment 1 (EXP1) was designed to test the assimilation system under the assumption that modelled surface Chl-a concentration was equivalent to the Chl-a concentration thought to be represented by MODIS OCM3. This experiment in analogous to those of Natvik and Evenson (2003), Gregg (2008) and Ford et al., (2012). This is an entirely reasonable assumption in offshore waters, however, OC3M is known to unreliable in coastal waters where sediments (E.g. TSS), bottom reflectance and CDOM cause artificially high OC3M values. In EXP2-4 we assume that the simulated OC3M is equivalent to the observed OC3M and it is used as input into the observation operator. The simulated OC3M (Eq 21) contains the signature from simulated TSS, CDOM and bottom reflectance as per section 2.3, 2.4 and Appendix A. In EXP5, we assimilate the reflectance at 551nm using the simulated reflectance at the equivalent wave length."

Please include the dates that have been run for.

We have included an introductory paragraph to the methods section that introduces the time period for which the experiments are conducted, and justifies its choice:
"The assimilation system was tested between the 25th of May 2013 to the 22nd of September. This period was chosen as it coincides with a field program and glider deployments. Additionally, this period also has the largest number of cloud free days. Details of the observations, model and assimilation system are given below. "

Either here or elsewhere in the paper, please discuss more fully how your approach compares with the typical NWP approach of using a radiative transfer model to convert the model variables, and assimilating the radiances directly.

A short paragraph is now included in the discussion and reference is made to this in the abstract and introduction.

Section 3.1 – please provide a reason why 14/7/2013 was chosen for Fig. 2 – I assume this date is representative?

Additional text has been added to justify this choice:
"This date is representative of dry season conditions with little cloud contamination along the inshore region."

Section 3.2 – define "adaptive ODE integrator".

We use a Dormand-Prince 4/5 integrator (now detailed in the text).

Whilst definitely an indication, a line plot of only nine cycles' worth of mean increments isn't really enough to convince me that using model OC3M rather than model surface level chlorophyll is giving more realistic model chlorophyll. I would like to see some comparison against in situ chlorophyll observations for all the model runs.

We have added an additional panel to Figure 7 and extra results to Figure 9 to present a comparison of EXP1 against a subset of the glider data. We have added additional text in the results section. Unfortunately due to the short period that we ran EXP1 we don't have enough in-situ observations from the Reef Rescue or IMOS stations to adequately comment on the EXP1 performance against in-situ nutrients.

I would also like to see some more discussion, either here or later in the paper, about why EXP1 performed so poorly it could only be run for nine assimilation cycles. This is the standard procedure that other groups use successfully, so for the community to be convinced that using an alternative method will be beneficial generally, and it's not just due to how your system is set up, some more information would be helpful.

We have altered our text to reflect that given the limited in-situ observation we cannot make broad statements on how this system would perform elsewhere, however, given the additional information as presented in Figure 7 and 9, EXP1 performance is worse than the control run and therefore is of no value. This is due to the OC3M observations being positively biased and therefore overestimating Chl-a concentration (even in deep water), and this is apparent in Figure 4. We have added the following text to the discussion:

"EXP1 performed poorly for a number of reasons. It is likely that even an 80% observation error was insufficient to adequately account for positive biases in the OC3M observations. These biases, present even in offshore waters, lead to positive innovations that result in adding Phytoplankton biomass in the increments. These large increments lead to persistently high biomass, that draw available nutrients down to very low levels. The only way to account for this form of observation error, is to inflate the "difference in kind error" to a large value. Given that the non-assimilating control run of the model, had forecast errors that range between 70%-100% (region dependent), running an assimilation system with an observation error larger than the error present in the non-assimilating model, does not make sense."

Section 3.3 – whilst observations from t+5 days haven't been assimilated, they will share a lot of characteristics with observations which have, since they come from the same source. I would therefore only class them as "semi-independent", rather than fully independent.

Wording in the text has been changed to semi-independent as your rightly point out there may be a temporal correlation in the remote sensing observations.

If OC3M follows a logarithmic distribution, would it be more appropriate to calculate statistics on log-transformed OC3M? Especially since you state that the distribution of errors is positively skewed.

This is true, however as the absolute values only vary by an order of magnitude, we wanted to present a figure not using log-scaling to ease in interpretation. We are happy to change this if there is a strong suggestion by the editor and/or other reviewers.

Be clear, both here and in the caption for Fig. 4, that your depth ranges refer to the bottom depth of the water column, rather than intervals within the water column.

Thankyou for this suggestion, we have clarified this in the caption text.

"Figure 12: Box and whisker plots of RMSD (top row) and MAPE (bottom row) of the mis-match between simulated OC3M and ANN-observed OC3M. Each panel contains the control run (C) and EXP4 showing forecast (F), persistence (P) and analysis (A). Presented are statistics for the whole domain (left column), and regions as defined by bottom depth depth (three rightmost columns)."

Section 3.3.1 – you pick a single cycle (the final one?) and make clear that the features you describe are just for "this particular analysis cycle". This is fine, but please make some mention of how representative it is of other cycles. The overlaid observations are too small to be seen without zooming in to about 400% on the pdf – on a printed copy they're just black blobs. Please make sure these can be seen in the final version. Perhaps because they're on different pages, or because the colour scale is saturated, but Fig. 5a and 7a look very similar to me.

We have tried a variety of plotting options to improve the visibility of the observations, but due to the spatial density of the observations, if they were any larger, then they obscure the background solution. Pending the outcome of this revision we will suggest that figures 13-15 are made landscape.

It is true that the forecast and analysis fields are similar, however as shown in the increment field there are subtle changes. This highlights one of the advantageous properties of this approach in that the increments being applied are small in most cases (due to the relatively low P biomass in the surface). Additional text has been added to the manuscript providing context for the cycle chosen. There has also been considerable effort invested into model development for this region. The eReefs BGC model has been shown to be skillful, therefore the assimilation system does not have to work hard to reduce model-data misfits.

"Cycle 22 was chosen to demonstrate the spatial impact of the system as it is representative of the last 10 cycles and was relatively cloud free."

Section 4 – whilst your results have "been achieved by explicitly assimilating like-forlike variables", you haven't actually shown any assimilation results against observations from not doing so. You state that "Our approach of simulating the observation is the opposite to the conversion of observed remote-sensing reflectances into modelled variables, e.g. the assimilation of ... Chl-a." But as far as I can tell, this is exactly what you've done – you've used the OC3M algorithm to convert the remote-sensing reflectance to Chl-a. Please tell me if I've misunderstood! It's true that you've processed the model and observations in the same

way, which is a good thing to have done and novel in biogeochemical assimilation, but it's not the same as directly assimilating the remote-sensing reflectance.

This is true, in the first version of this manuscript we assimilated f(RSR), and chose OC3M as our function. We still consider the assimilation of RSR, as the model explicitly predicts RSR, and ***we don't rely on the empirical statistical relationship between RSR and Chl-a****. Reviewer 2 also suggested we undertake an additional experiment whereby reflectance is directly assimilated. This has required substantial work and an additional 3 experiments were undertaken to accomplish this, however, we have only reported on the best performing configuration in the manuscript and denoted these results as EXP5. There are numerous changes scattered throughout the manuscript that deal with the additional results.*

The final paragraph of this section needs re-writing to be clearer (currently it advocates "remote-sensed reflectance" as an alternative approach to "remote-sensing re- flectances"), and to place it in the context of differing model complexities and regions of interest.

We have carefully address the language used throughout and have reworded numerous sections to avoid any confusion.

Section 5 – you state you "halve the forecast error compared to simply assuming the OC3M is directly related to the model prediction of surface total Chl-a" – has this been shown? Is this just based on Fig. 3, or something else? If non-ocean colour-specific sensors can be reliably assimilated then that would indeed be a great breakthrough. But given that you still appear to transform the remotesensing reflectance to chlorophyll, will the current approach still require the use of ocean colour-specific sensors? Please explain this more clearly.

With the additional experiment included, we have demonstrated that we can directly assimilate RSR. And the comparison against withheld ins-situ observations (i.e. the glider data shows a reduction in RMSD from 0.3 to 0.1 in the upper 80m (Figure 9). The forecast error statistics contained in Figure 12 also support this statement.

Finally, please state the variables and units in your Table descriptions.

Apologies, we have added units where needed.

Recommendations: Assuming that I haven't misunderstood what's already been done, then ideally I would like to see the direct assimilation of the remote-sensing reflectance, as that would make for a very important step forward in biogeochemical assimilation. However, I acknowledge that this would involve a considerable amount of extra effort, and would in effect constitute a separate study. Therefore, I'm happy that the current work is suitable for publication in BG, provided that the authors are clear about what has been done (the like-for-like processing of model and observations, rather than the direct assimilation of remote-sensing reflectance), the comments above are addressed, and the manuscript is tidied up. In particular, I would like to see greater comparison between the different experiments, with all runs compared to in situ observations. Furthermore, either a regional algorithm should be included in the comparison, or a much stronger justification should be provided for why this hasn't been done. Whilst in this case the direct assimilation of remote-sensing reflectance doesn't seem to have been performed, it is an important step towards it.

References:

Ciavatta, S., Torres, R., Martinez-Vicente, V., Smyth, T., Dall'Olmo, G., Polimene, L., Allen, J. I. (2014). Assimilation of remotely-sensed optical properties to improve marine biogeochemistry modelling. Progress in Oceanography, 127, 74-95.

Shulman, I., Frolov, S., Anderson, S., Penta, B., Gould, R., Sakalaukus, P., Ladner, S. (2013). Impact of bio-optical data assimilation on short-term coupled physical, biooptical model predictions. Journal of Geophysical Research: Oceans, 118(4), 2215- 2230.

As mentioned previously, we have included an additional experiment to address your concerns and also those of Reviewer 2. Additional analysis against in-situ observations has been included when appropriate. Given that we have only run the EXP1 configuration for a short period, we have altered some of our wording such that we don't make "global" statements, however, for the period that has be tested there is a clear indication that the configurations in EXP4 and EXP5 are superior. We have also included a larger number of references to acknowledge previous work in assimilating IOP's and AOPS.
The inclusion of a regional algorithm could be done, but would require further DA experiments to be run. In the long term we do not see the advantage of regional algorithms as they require a substantial and expensive field campaign to appropriately tune an empirical statistical algorithm, that can be circumvented through the use of RSR.

Use of remote-sensing reflectance to constrain a data assimilating marine biogeochemical model of the Great Barrier Reef.

Emlyn M. Jones[1], Mark E. Baird[1], Mathieu Mongin[1], John Parslow[1], Jenny Skerratt[1], Nugzar Margvelashvili[1], Richard J. Matear[1], Karen Wild-Allen[1], Barbara Robson[2], Farhan Rizwi[1], Peter Oke[1], Edward King[1], Thomas Schroeder[3], Andy Steven[3] and John Taylor[4]

[1] CSIRO Oceans and Atmosphere, Hobart, Australia, 7000

[2] CSIRO Land and Water, Canberra, Australia, 2601

[3] CSIRO Oceans and Atmosphere, Brisbane, Australia, 4102

[4] CSIRO Data61, Canberra, Australia, 2601

Correspondence to: Emlyn M. Jones (emlyn.jones@csiro.au)

Key words: Great Barrier Reef, Data Assimilation, Remote Sensing, observed OC3M, Reflectances, Biogeochemical (BGC) Model, Chlorophyll

Abstract: Skillful marine biogeochemical (BGC) models are required to understand a range of coastal and global phenomena such as changes in nitrogen and carbon cycles. The refinement of BGC models through the assimilation of variables calculated from observed in-water inherent optical properties (IOPs), such as phytoplankton absorption, is problematic. Empirically-derived relationships between IOPs and variables such as Chlorophyll-a concentration (Chl-a), Total Suspended Solids (TSS) and Color Dissolved Organic Matter (CDOM) have been shown to have errors that can exceed 100% of the observed quantity. These errors are greatest in shallow coastal regions, such as the Great Barrier Reef (GBR), due the additional signal from bottom reflectance. Rather than assimilate quantities calculated using IOP algorithms, this study demonstrates the advantages of assimilating quantities calculated directly from the less error-prone satellite remote-sensing reflectance (RSR). To assimilate the observed RSR, we use an in-water optical model to produce an equivalent simulated RSR, and calculate the mis-match between the observed and simulated quantities to constrain the BGC model with a Deterministic Ensemble Kalman Filter (DEnKF). Following the traditional assumption that simulated surface Chl-a is equivalent to the remotely-sensed OC3M estimate of Chl-a resulted in a forecast error of approximately 75%. We show this error can be halved by instead using simulated RSR to constrain the model via the assimilation system. When the analysis and forecast fields from the RSR-based assimilation system are compared with the non-assimilating model, a comparison against independent *in situ* observations of Chl-a, TSS, and dissolved inorganic nutrients (NO_3, NH_4 and DIP) show that errors are reduced by up to 90%. In all cases, the assimilation system improves the simulation compared to the non-assimilating model. Our approach allows for the incorporation of vast quantities of

remote-sensing observations that have in the past been discarded due to shallow water and/or artefacts introduced by terrestrially-derived TSS and CDOM, or the lack of a calibrated regional IOP algorithm.

1 Introduction:

Aquatic biogeochemical (BGC) models have been used to understand a range of coastal and global phenomena such ocean acidification (Mongin et al., 2016), nutrient pollution (Skerratt et al., 2013) and carbon cycles, and are central to our predictions of global climate (Sarmiento and Gruber, 2006). At the coastal / regional scale, non-linear biogeochemical processes driven by planktonic interactions, as well as non-linear circulation features such as mesoscale eddies, limit the timescale over which biogeochemical properties are deterministically predictable (Baird, 2010). For the purposes of prediction, it is therefore necessary to assimilate observations to correct for model errors and non-linear processes.

In situ observations of phytoplankton pigments and macro-nutrients are sparse in space and time due to the prohibitive expense of collecting them. Optical sensors on gliders and floats provide high resolution *in situ* observations that are used to estimate pigment and nutrient concentrations. Nonetheless, these observations in large parts of the ocean remain sparse. The most spatially comprehensive dataset available for BGC assimilation is from Ocean Color (OC) remote sensing. The assimilation of remotely-sensed data into marine biogeochemical models has been problematic due to differences between the variables represented in models and the variables that are routinely observed (Baird et al., 2016b), typically referred to as "difference in kind" errors. Satellites measure the top of atmosphere radiance, not chlorophyll-a concentration (Chl-a; or other modelled variables) directly. To attain estimates of Chl-a or other variables, the spectrally-resolved top of atmosphere radiance is converted into atmospherically-corrected remote-sensing reflectances (RSRs), which are then related to Chl-a via empirical statistical relationships derived from *in situ* observations. Figure 1 graphically demonstrates the key steps in the OC processing chain, and points at which models and remotely-sensed and *in situ* observations can be compared.

Early studies investigating the benefits of assimilating OC products, predominantly SeaWiFS-derived Chl-a, into BGC models include those of Carmillet et al., (2001) and Natvik and Evenson (2003), with a comprehensive review of algorithms used and observations assimilated detailed in Gregg (2008). Considerable effort has been invested in DA algorithm development, with ensemble and variation approaches being the most common. A thorough review of these approaches in a statistical sense is presented in Dowd et al., (2014). There are now examples of operational and pre-operational global systems that routinely assimilate Chl-a products (Ford et al., 2012). Additionally, there has been further experimentation with assimilating alternative remotely-sensed Apparent Optical Properties (AOPs) such as the vertical attenuation coefficient at 443 nm, Kd_{443} (Ciavatta et al., 2014) and Inherent Optical Properties (IOPs) such as phytoplankton absorption (a_{ph}), as described in Shulman et al., (2013).

It is well known that OC algorithms, such as observed OC3M (Moderate Resolution Imaging Spectroradiometer, MODIS, three-band Chl-a algorithm) that are optimized for global

applications suffer from errors due to a variety of optically-active constituents in coastal and shelf waters (Odermatt et al., 2012). Furthermore, it has also been noted that even globally, there is a non-uniform distribution of error, and substantial bias, in the OC3M-derived Chl-a. Satellite-derived ocean color products such as OC3M are a function of observed RSRs. A substantial effort is invested in empirical studies that convert RSRs to biogeochemical quantities such as Chl-a, Total Suspended Solid concentration (TSS), phytoplankton functional types (PFTs), and Colored Dissolved Organic Matter (CDOM) (Odermatt et al., 2012). Each of these empirical relationships have differing error magnitudes stemming not only from a difference in kind, but also representation errors. In optically deep regions (e.g. offshore waters) not influenced by sediment resuspension and terrestrial runoff, typical errors for OC derived Chl-a (e.g. OC3M) are less than 40%, and as low as 5-20% for IOPs and AOPs. However, in optically-complex coastal areas where there is river discharge, sediment resuspension and optically shallow water, errors can exceed 300% (e.g. Schroeder et al., 2016; Qin et al., 2007).

Numerical weather prediction (NWP) avoids difference in kind errors through using the model to simulate directly-observed quantities, such as temperature brightness, in preference to deriving other quantities such as temperature and humidity profiles from the brightness measurements. Thus NWP systems commonly assimilate brightness temperature in preference to satellite-derived temperature and humidity (Derber and Wu, 1998; Dee et al., 2011). The goal of this study is to apply this approach to marine biogeochemical modelling, assimilating RSRs rather than quantities calculated using empirical-statistical relationships. We assess this approach through comparison against withheld *in situ* observations.

The Great Barrier Reef (GBR) region, located along the northeast coastline of Australia, is used to demonstrate the assimilations of RSRs in optically-complex and shallow waters (Blondeau-Patissier et al., 2009). The GBR is characterized by fringing reefs along the continental slope that create a semi-connected inshore lagoon that spans over 3000 km of coastline (Figure 2). The GBR ecosystem, described as one of the seven natural wonders of the world, is under increasing pressure from local and global anthropogenic stressors (De'ath et al., 2012). Decreasing water clarity due to nutrient and sediment pollution is considered a serious threat to the GBR ecosystem (Thompson et al., 2014), with major concerns being the impact of lower benthic light levels on coral and seagrass communities (Collier et al., 2012; Baird et al., 2016a) and the impacts of invasive species (e.g. Morello et al., 2014).

The paper is structured in the following manner, in Section 2 (Methods) the data for assimilation and skill assessment is presented, along with a description of the model and assimilation methods used. Section 3 contains the results from the control (non-assimilating) run of the model, and subsequent data assimilation experiments. In Section 4 we discuss the approach used and implications of the findings more generally. And we conclude with major findings in Section 5.

2 Methods:

The assimilation system was tested between the 25th May 2013 to 22nd September 2013. This period was chosen as it coincides with a field program by the Great Barrier Reef Marine Park Authority (GBRMPA) inshore Marine Monitoring Program (MMP) and an autonomous glider deployment (Integrated Marine Observing System, IMOS). Additionally, this period also has a large number of cloud free days. Details of the observations, model and assimilation system are given below.

2.1 In-situ observations

All *in situ* observations have been withheld from the assimilation system for validation purposes, and are primarily obtained from two different programs.

The Integrated Marine Observing System (IMOS) has deployed fluorometers on moorings at Yongala and North Stradbroke National Reference Stations (NRS) (Figure 2). This study uses the monthly observations of dissolved inorganic nutrients (NO_3, NH_4 and DIP) at the NRS sites (Lynch et al., 2014). Additionally, glider data obtained from the IMOS-operated Australian National Facility for Ocean Gliders (ANFOG) provide cross-shelf profiles of water column properties, including temperature, salinity, and chlorophyll fluorescence. These locations are shown in Figure 2.

The GMRMPA MMP (Figure 2, yellow circles) contains 13 sites in inshore regions and in the GBR lagoon, and sample nutrients and Chl-a extractions 3 times a year (Thompson et al., 2011; Rolfe and Greg, 2015). The moorings were deployed at the GBRMPA MMP sites from 2009 to 2014 and included a Sea-Bird water quality monitors (WQM) that measure chlorophyll fluorescence and turbidity (NTU). A comparison of the 2011-2014 control run simulation against the GBRMPA MMP observations, and other observations, is available in a 180 page skill assessment report at: https://research.csiro.au/ereefs/models.

2.2 MODIS observations

The daily observations of RSR are obtained from MODIS-Aqua and an atmospheric correction developed for the region is applied (Schroeder et al., 2007). The atmospheric correction applied an Artificial Neural Network (ANN) approach trained by a radiative transfer model to invert the top of atmosphere (TOA) signal measured by MODIS-Aqua. The ANN algorithm was adapted to an approach previously-developed for the MEdium Resolution Imaging Spectrometer (MERIS) sensor but on the basis of a different learning algorithm (Schroeder et al., 2007). Algorithm performance is described in detail in Goyens et al. (2013) and King et al. (2014).

SeaDAS-provided Level-2 flags were used to quality control the observed RSR and to exclude erroneous and out-of-range pixels. We filtered the data for land and severe sun glint affected pixels, cloud contamination including cloud shadows, and rejected pixels with observing and solar zenith angles above 52°and 70°, respectively. Super-observations (Cummings et al, 2005; Oke et al., 2008b), of RSR and OC3M (see Section 2.5.2, eq. 21) are generated for the assimilation system by taking the mean and variance of all the 1 km resolution observations that fall within a 4 km model grid cell. The mean value of the super-

observation is then assimilated, while the variance is used as an estimate of the representation error in the observation error covariance matrix. More detail on this process in given in Section 2.5.2.

2.3 The eReefs modelling system

We used the eReefs coupled hydrodynamic, sediment and biogeochemical modelling system (Schiller et al., 2014). The hydrodynamic model is a fully three-dimensional finite-difference baroclinic model based on the three dimensional equations of momentum, continuity and conservation of heat and salt, employing the hydrostatic and Boussinesq assumptions (Herzfeld 2006; Herzfeld and Gillibrand 2015). The sediment transport model adds a multilayer sediment bed to the hydrodynamic model grid and simulates sinking, deposition and resuspension of multiple size-classes of suspended sediment (Margvelashvili 2009). The complex biogeochemical model simulates optical, nutrient, plankton, benthic organisms (seagrass macroalgae and coral), detritus, chemical and sediment dynamics across the whole GBR region, spanning estuarine systems to oligotrophic offshore reefs (Figure 3, Baird et al., 2016b), an expanded description of the BGC model is given in Appendix A. Briefly, the biogeochemical model considers four groups of microalgae (small and large phytoplankton, *Trichodesmium* and microphytobenthos), two zooplankton groups, three macrophytes types (seagrass types corresponding to *Zostera* and *Halophila*, macroalgae) and coral communities. Photosynthetic growth is determined by concentrations of dissolved nutrients (nitrogen and phosphorous) and photosynthetically active radiation. Microalgae contain two pigments (chlorophyll-a and an accessory pigment), and have variable carbon:pigment ratios determined using a photo-adaptation model (described in Baird et al., 2013). Overall, the model contains 23 optically-active constituents (Baird et al., 2016b; and Appendix A).

The model is forced with freshwater inputs at 21 rivers along the GBR and the Fly River in southwest Papua New Guinea. River flows input into the model are obtained from the DERM (Department of Environment and Resource Management) gauging network. Statistical flow/load relationships are used to account for nutrient and sediment inputs from rivers into the model (statistical relationships between river flow and nutrient concentrations (Furnas 2003). Nutrient concentrations flowing in from the ocean boundaries were obtained from the CSIRO Atlas of Regional Seas (CARS) 2009 climatology (Ridgway et al., 2002).

2.4 Calculation of remote-sensing reflectance from biogeochemical state

The model contains 23 optically-active constituents (Baird et al., 2016b; and Appendix A). To calculate the RSR at the surface, we need to consider the light returning from multiple depths, and from the bottom. Rather than using a computationally-expensive radiative transfer model, we approximate RSR based on an optical-depth weighted scheme (Baird et al., 2016b), alternative methods are given in Fujii et al., (2007) and Dutkiewicz et al., (2015). The ratio of the backscattering coefficient to the sum of backscattering ($\mathbf{b}_b$) and absorption ($\mathbf{a}$) coefficients for the whole water column at wavelength, λ, is:

$$\mathbf{u_\lambda} = \sum \frac{\mathbf{w_{\lambda,z'}\, b_{b,\lambda,z'}}}{\mathbf{a_{\lambda,z'} + b_{b,\lambda,z'}}} \qquad \text{Eq 1}$$

where $\mathbf{w_{\lambda,z'}}$ is a weighting for each layer representing the component of the RSR due to the absorption and scattering at depth **z'.**

The weighting fraction is given by:

$$\mathbf{w_{\lambda,z}} = \frac{1}{\mathbf{z_1 - z_0}} \int_{\mathbf{z_0}}^{\mathbf{z_1}} \mathbf{exp}\left(-\mathbf{2K_{\lambda,z'}}\right)\mathbf{dz'} \qquad \text{Eq 2}$$

where $\mathbf{K_\lambda}$ is the vertical attenuation coefficient at wavelength $\boldsymbol{\lambda}$, $\mathbf{z_0}$ and $\mathbf{z_1}$ are the top and bottom depths of the layer and the factor of 2 accounts for the path length of both downwelling and upwelling light. The vertical attenuation coefficient is calculated from the sum of the absorption and scattering properties of each of the optically-active constituents, and the zenith angle (for each of these relationships, and more information, see Baird et al., 2016b).

The integral of $\mathbf{w_{\lambda,z'}}$ to infinite depth is 1. In areas where light reaches the bottom, the integral of $\mathbf{w_{\lambda,z'}}$ to the bottom is less than one, and benthic reflectance is included in the sum as an extra term with a weighting of 1 - $\sum w_{\lambda,z'}$.

The sub-surface RSR, $\mathbf{r_{rs}}$, is given by:

$$\mathbf{r_{rs}} = \mathbf{g_0 u_\lambda} + \mathbf{g_1 u_\lambda^2} \qquad \text{Eq 3}$$

where $\mathbf{g_0}$ = 0.0895 and $\mathbf{g_1}$ = 0.1247 are coefficients for the nadir-view in oceanic waters that vary with wavelength and other optical properties (Morel et al., 2002), but can be approximated as constants (Lee et al., 2002). The constants result in a change of units from the unitless **u** to a per unit of solid angle, sr^{-1}, quantity, $\mathbf{r_{rs}}$.

The above-surface RSR is given by (Lee et al., 2002):

$$\mathbf{R_{rs,\lambda}} = \frac{\mathbf{0.52\, r_{rs,\lambda}}}{\mathbf{1 - 1.7 r_{rs}}} \qquad \text{Eq 4}$$

Thus, the above-surface RSR is calculated from the inherent optical properties of the optically-active constituents in the biogeochemical model.

2.5 Data Assimilation System

The Data Assimilation (DA) algorithm used in this study is the Deterministic Ensemble Kalman Filter (DEnKF; Sakov and Oke, 2008). The full biogeochemical state variable list contains over 130 2D and 3D variables. Including all of these variables in the assimilation system is impractical due to memory constraints, but we also acknowledge that for many variables the observations will be uninformative, and therefore not good candidates to include in the assimilation state vector. We therefore limit the variables that are updated within the system to a select subset that are detailed in Section 2.5.2.

2.5.1 Data Assimilation Algorithm

The DEnKF is based on the Kalman filter analysis equation, of which various flavors have had general success in state estimation in other marine BGC data assimilation problems (e.g. Hu et al., 2012, Ciavatta et al., 2016). The derivation of the DEnKF is given in Sakov and Oke (2008) and is a modification to the traditional Kalman filter equation:

$$\mathbf{x^a} = \mathbf{x^f} + \mathbf{K}(\mathbf{Y} - \mathbf{H}\mathbf{x^f}) \qquad \text{Eq 5}$$

Where **x** is the model state, **Y** is the vector of observations, **H** is the observation operator, and the superscripts of ^{a} and ^{f} denote the analysis and forecast fields respectively. In this study, we only use a subset of the state variables in the DEnKF update, those variables not included in the state vectors denoted in Table 1 are not altered by the state update. The Forecast Innovations (**FI**) are defined by:

$$\mathbf{FI} = (\mathbf{Y} - \mathbf{H}\mathbf{x^f}) \qquad \text{Eq 6}$$

The Kalman gain matrix, ***K***, is given by:

$$\mathbf{K} = \mathbf{l}\mathbf{P^f}\mathbf{H^T}\left(\mathbf{H}\mathbf{P^f}\mathbf{H^T} + \mathbf{R}\right)^{-1} \qquad \text{Eq 7}$$

Where $\mathbf{l}$ is the localization operator, applied in the form of covariance localization (Sakov and Bertino, 2011), with an isotropic localization radius of 60 km, and **R** is the observation error covariance matrix. Off diagonal elements of **R** are set to 0, while the diagonal elements of **R** are the sum of the representation error, difference in kind error and analytical measurement error, termed $\mathbf{E_{tot}}$. These error terms are discussed in details in Section 2.5.2 (additionally see Schaeffer et al., 2016 for a further discussion of these error sources relating to BGC variables and methods to estimate them from glider data).

The background error covariance matrix, ***P***, is given by:

$$\mathbf{P^f} = \frac{1}{\mathbf{m}-1}\sum_{\mathbf{i}=1}^{\mathbf{m}}(\mathbf{X_i^f} - \mathbf{x^f})(\mathbf{X_i^f} - \mathbf{x^f})^{\mathbf{T}} = \frac{1}{\mathbf{m}-1}\mathbf{A}\mathbf{A^T} \qquad \text{Eq 8}$$

where **m** is the ensemble size, and **i** denotes the **i**th member of the ensemble. Given that we are using a flavor of the ensemble Kalman filter, the background error covariance is approximated by a 36 member dynamic ensemble whereby $\mathbf{X_i^f}$ is the ith ensemble member, and **x** is the ensemble mean. We acknowledge that a 36 member ensemble is small, however as demonstrated in the results section this small ensemble performs adequately for univariate assimilation. The ensemble size will need to be increased to assimilate multivariate observations. To avoid negative values and normalize the state, we log-transform the state before forming the state vector. The application of the log-transform to biological variables is discussed in Parslow et al., (2013), and is commonly used in other BGC assimilation schemes (e.g. Ciavatta et al., 2014). The background error covariance matrix is never computed; rather a series of anomaly fields are constructed and denoted by **A**. We then construct the Kalman gain matrix in the observation subspace as per Sakov and Oke (2008):

$$\mathbf{A} = [\mathbf{A_1} \ldots \mathbf{A_m}] \qquad \text{Eq 9}$$

where the **i**th anomaly field is given by:

$$\mathbf{A_i} = \mathbf{X_i^f} - \mathbf{x^f} \qquad \text{Eq 10}$$

where **x** is the the ensemble mean and is updated via Eq 5, each anomaly field is updated by:

$$\mathbf{A^a} = \mathbf{A^f} - \frac{\mathbf{1}}{\mathbf{2}}\mathbf{KHA^f} \qquad \text{Eq 11}$$

The full analyzed ensemble is then given by:

$$\mathbf{X^a} = \mathbf{A^a} + [\mathbf{x^a}, \ldots, \mathbf{x^a}] \qquad \text{Eq 12}$$

The assimilation system iterates through time using a five-day forecast duration. The assimilation system is cycled by calculating the analysis fields at time t, using the forecast from the previous cycle, $\mathbf{X}^f(t)$, and observations **Y**(t) at time t, using all observations that fall within a window of t +/- 3 hrs. The numerical model is initialized using the analysis fields $\mathbf{X}^a(t)$ and the next five day forecast is made. This forecast at t+5 days, $\mathbf{X}^f(t+5)$, is then used in the next assimilation cycle.

The DEnKF requires the ensemble to be perturbed in such a way that it captures the main source of error. These perturbations are introduced in a way that captures our prior understanding of the dominant errors. In this system we expect that errors will stem from uncertainty in the Initial Conditions (ICs) as per most assimilation system. Additional sources of error can stem from uncertainty associated with BGC process parameters, which has been discussed at length in Parslow et al. (2013), and river boundary conditions (BCs). In this study we have ignored the effects of uncertainty that propagate from the model physics (hydrodynamics), short wave radiation forcing and open ocean boundary fluxes of BGC tracers, but do acknowledge that they contribute to the overall uncertainty in the predicted system state.

In the context of this study, we have introduced perturbations to the ensemble by sampling initial conditions randomly from a four year run (where **T** = 4 years) of the BGC model:

$$\mathbf{X(t=0)_i} \sim \mathbf{Uniform(X(t=0)\ ,X(t=T)\ )\ \ for\ i = 2 \ldots m} \qquad \text{Eq 13}$$

Where $\mathbf{X}(t=0)_i$ is the initial condition for the model state for the **i**th member sampled for a uniform distribution with no replacement. Where $\mathbf{X}(t=0)_{i=1}$ is the ensemble mean. Sensitivity experiments have shown the model is sensitive to perturbation in the quadratic zooplankton mortality rate for large ($\mathbf{m_{Q,ZS}}$) and small zooplankton ($\mathbf{m_{Q,ZL}}$) with units of d^{-1} $(mg\ N\ m^{-3})^{-1}$. These are considered system parameters, and are as such uncertain. To this end we have perturbed the ensemble, by sampling space and time invariant parameters from:

$$\mathbf{m_{q,ZL,i} \sim LN(0.012, 1) \quad for\ i = 2 \ldots m} \qquad \text{Eq 14}$$
$$\mathbf{m_{q,ZS,i} \sim LN(0.007, 1) \quad for\ i = 2 \ldots m} \qquad \text{Eq 15}$$

Where **LN** is a log-normal distribution, and respective means of 0.012 and 0.007 are typical values as used in the control run (Baird et al., 2016b), and given relatively broad standard deviations of 1 for both parameters. The river nutrient and sediment loads were altered by a time invariant scaling factor ($\boldsymbol{\theta}$) to all rivers:

$$\mathbf{\theta_{NO3,i} \sim N(1, 0.3) \quad for\ i = 2 \ldots m} \qquad \text{Eq 16}$$
$$\mathbf{\theta_{NH4,i} \sim N(1, 0.3) \quad for\ i = 2 \ldots m} \qquad \text{Eq 17}$$
$$\mathbf{\theta_{DIP,i} \sim N(1, 0.3) \quad for\ i = 2 \ldots m} \qquad \text{Eq 18}$$
$$\mathbf{\theta_{FineSed,i} \sim N(1, 0.3) \quad for\ i = 2 \ldots m} \qquad \text{Eq 19}$$

Where **N** is a normal distribution truncated at 0. Each ensemble member has their load ($\mathbf{Q_i}$) scaled according to:

$$\mathbf{Q_i = \theta_i Q_{control}} \qquad \text{Eq 20}$$

where $\mathbf{Q_{control}}$ is the load entering the control run.

2.5.2 Assimilation system experiments and configuration

The assimilation system was assessed using five assimilation system configurations (Table 1) using a subset of the full model state in the assimilation state vector, and corresponding diagonal elements of the observation error covariance matrix (**R**). The assimilation experiments were conducted on Raijin, a super-computer host at the National Computational Infrastructure (NCI; http://nci.org.au/). The assimilation cycle progressed via a two-step process. The first step generated the forecast fields by integrating the ensemble forwards on 36 nodes of the system. Each node has a dual 8 core processor, with 16 Gb of RAM. This step takes approximately 90 minutes to simulate 5 model days. The second step generates the analysis fields and updates model scripts to allow for the next forecast cycle. The analysis step is undertaken on a single high-memory node and requires 800 Gb of RAM and takes 50 minutes using 16 cores.

We only allow the observations to update the variables contained in the assimilation state vector. The analyzed assimilation state vector is then inserted into the full model state vector.

There are three sources of observational error ($\mathbf{E_{tot}}$) that must be accounted for when relating remote-sensed observations to the modelled state variable:

1. Representation error ($\mathbf{E_r}$) – errors that arise due to the approximation that the modelled tracer quantities are an average over a whole model cell. This can be thought of as unresolved spatial variability.
2. Difference in kind error ($\mathbf{E_D}$) – these errors arise when the variables that are being modelled and included in the assimilation state vector differ from the observations. For example, many studies have included surface Chl-a (or some optical depth weighted average) in the assimilation state vector, and assume there is a direct

relationship with OC3M (or other quantities). OC3M is known to have typical errors of between 30-70% in blue water domains, and errors that exceed 300% in optically complex (or optically shallow) waters (Qin et al., 2007).

3. Analytical/sensor/processing error (E_A) – Depending on the observational platform in use, these errors can be small (e.g. ARGO floats), or in the case of remote-sensing products, even with the ANN atmospheric correction, as large as 15-20% (see Schroeder et al., 2007).

The total observation error used on the diagonal of **R** is then given by the sum of error types: $E_{tot} = E_R + E_D + E_A$

The sum of these can be large and forms the diagonal element of the observation error covariance matrix. The larger E_{tot}, the lower the impact of the observations. If we can remove the difference in kind error (E_D), then we only have representation error and analytical/sensor/processing error. Given that Level 3 ocean color remote sensing products rely on empirical/statistical relationships, E_D dominates, therefore if we can minimize E_D (or remove it entirely), then the information content of the observations increases. Conversely, if we don't have a large enough observation error, we run the risk of overfitting the observations and generating unrealistically large increments for the unobserved model state variables.

Experiment 1 (EXP1) was designed to test the assimilation system under the assumption that modelled surface total Chl-a (the sum of small and large Phytoplankton Chl-a and *Trichodesmium* Chl-a) was equivalent to the Chl-a from MODIS OC3M. This experiment is analogous to those of Natvik and Evenson (2003), Gregg (2008) and Ford et al., (2012). This is a reasonable assumption in offshore waters, however, OC3M is known to be unreliable in coastal waters where sediments (e.g. TSS), bottom reflectance and CDOM cause artificially high OC3M values. We expect that EXP1 will contain difference in kind errors that stem from the assumption that surface Chl-a is equivalent to observed OC3M. The observation error prescribed for this experiment is considered to be the lower bound of OC3M errors as reported in Qin et al., (2007).

In Experiments 2-4 (EXP2-4) we assume that the simulated OC3M is equivalent to the observed OC3M and it is used as an input into the observation operator. The simulated OC3M (Eq 21) contains the signature from simulated TSS, CDOM and bottom reflectance as per Section 2.3, 2.4 and Appendix A. Thus for EXP2–4, the observed and simulated OC3M contains the same error characteristics. In other words, the configuration used for EXP2 is the same as EXP1, except the effect of difference in kind error has been removed by using simulated OC3M, in place of total surface Chl-a. In EXP3, we reduce the observation error to account for the reduction in E_D, and in EXP4 we add additional variables to the assimilation state vector.

In EXP5, we assimilate observed RSR at 551 nm (R551) using the simulated RSR at the equivalent wavelength. The progressively lowering of the E_{tot}, inserted on the diagonal element of **R**, through EXP1-5 shows the progressive reduction in difference in kind error associated with each experimental configuration (Table 1). Typically errors in RSR as

reported in Schroeder et al., 2007 are in the order of 10-20%, and we have used a figure of 20% for this experiment.

In all experiments we have generated a set of super-observations (Cummings et al, 2005; Oke et al., 2008) by spatially averaging the 1 km observations, onto the 4 km model grid. Prior to super-obing in Experiments 1-4 (EXP1-4), the observed atmospherically-corrected RSRs (Schroeder et al., 2007) at 443, 488 and 551 nm were transformed into a single observation using the OC3M algorithm:

$$\mathbf{OC3M} = \mathbf{10}^{(\mathbf{a0}+\mathbf{a1}.\mathbf{B}+\mathbf{a2}.\mathbf{B}^2+\mathbf{a3}.\mathbf{B}^3+\mathbf{a4}.\mathbf{B}^4)} \qquad \text{Eq 21}$$

Where **a0**, **a1**, … **a4** are a set of empirically determined coefficients (e.g. NOMAD version 2, http://seabass.gsfc.nasa.gov/wiki/article.cgi?article=NOMAD) and **B** is:

$$\mathbf{B} = \mathbf{log}_{10}(\frac{\mathbf{R}_{rs,\lambda 1}}{\mathbf{R}_{rs,\lambda 2}}) \qquad \text{Eq 22}$$

$\mathbf{R}_{rs,\lambda 1}$ and $\mathbf{R}_{rs,\lambda 2}$ are determined by the absolute magnitude of the remote-sensing reflectance, and $\mathbf{R}_{rs,\lambda 1}$ is either the band centered on 443 nm or 488 nm and $\mathbf{R}_{rs,\lambda 2}$ is the band centered on 551 nm. We apply the OC3M algorithm (Eq 21) to both the observed and simulated RSRs.

The assimilation system preserves the stoichiometry of the small and large phytoplankton (referred to as PhyS_Chl-a and PhyL_Chl-a respectively) as follows. In the biogeochemical model, each phytoplankton cell (small, large, benthic or *Trichodesmium*) is represented by a quantity of structural material, B, and reserves of nitrogen, R_N, reserves of phosphorus, R_P, reserves of energy, R_I, and an intracellular chlorophyll-a concentration, c_i. Our intention in the assimilation is to change the number of cells, as quantified by B, not the physiological status of the cell, as represented by R_N, R_P, R_I, and c_i. Since the reserves are quantified as the total of these reserves across the entire population, each of the reserves is changed by the same proportion as the biomass. Thus, for example, the nitrogen reserve of an individual cell, R_N / B, is unchanged. Once the analyzed quantity of c_i is determined (e.g. PhyL_Chl-a and PhyS_Chl-a), the quantities of R_N, R_P, R_I and B are updated such that the respective ratios prior to assimilation are preserved.

3 Results:

3.1 Control Run

The modelling system has been designed to represent the spatially-resolved water quality dynamics (phytoplankton, nutrients, turbidity and oxygen) of the GBR World Heritage Area for informed management. A number of indicators have been used to assess the skill of the model, including RMS errors, Pearson's correlation coefficients, Wilmott's skill indicators (Wilmott et al., 1985; https://research.csiro.au/ereefs/models).

The simulated state variable concentrations resemble both the regional climatology for offshore-reef, lagoon-reef and near-shore zones and water quality observations under

contrasting seasons/loads and flood events [not shown here, but detailed at https://research.csiro.au/ereefs/models, with optical (Baird et al., 2016b) and carbon chemistry (Mongin et al., 2016) skill assessment published elsewhere]. As mentioned above, the model simulates RSR. It is therefore possible to incorporate these RSRs into standard, well-recognized remote-sensing products. Figure 4 presents a snapshot of simulated surface Chl-a and the simulated OC3M, as well as remotely-sensed products (regional ANN-observed OC3M and NASA-observed OC3M). The images are during the dry season conditions with little cloud contamination along the inshore region.

The two panels on the right side of Figure 4 represent the simulated (top) and NASA remotely-sensed (bottom) OC3M estimate of Chl-a. Both combine individual RSRs into proxies for Chl-a using the OC3M algorithm (Eq 21). OC3M poorly represents surface Chl-a close to the coast where sediment resuspension and CDOM absorption dominate. By comparing the two panels, we can conclude that the model represents accurately the general distribution of Chl-a throughout the region, with high values along the coast and above each reef systems, and low concentrations offshore. The simulated surface Chl-a inside the coastal band is lower than in the remotely-sensed OC3M observation. The two panels on the left of Figure 4 represent simulated surface Chl-a (top) and OC3M based on the regionally-optimized RSR (bottom; ANN-observed OC3M). Regardless of the remote-sensing products used, there are clear differences between the simulated Chl-a and the simulated OC3M, and these exceed the differences between those of the ANN-observed and NASA-observed Chl-a.

Surface Chl-a and OC3M are typically assumed to be equivalent. However, many studies have shown that there are regional biases and substantial overestimation in optically-complex coastal waters (Qin et al., 2007). This is demonstrated in Figure 5, where we plot simulated surface Chl-a against simulated OC3M for deep water (Figure 5, left) and the whole domain (Figure 5, right) for the 26th May 2013. It is immediately obvious that these variables are not equivalent and the error structure is non-linear, especially when coastal regions are included. Figure 5 shows that for the GBR region, there is a substantial risk of OC3M over estimating the *in situ* Chl-a even in optically simple deep water regions of the domain.

3.2 Assimilation system configuration experiments

To choose the best configuration for the assimilation of RSR into a coastal biogeochemical model, five experiments were undertaken using a variety of state variables in the state vector (**X**, Table 1), and by altering the diagonal elements of the observation error covariance matrix (**R**, Eq 7).

3.2.1 Forecast Innovations

The forecast innovations (Eq 6) for the five assimilation system configuration experiments are shown in Figure 6:

- EXP 1 (green lines, Figure 6): Assumes that total surface Chl-a is equivalent to observed OC3M and is used to calculate the forecast innovations, which are then used to update small and large phytoplankton Chl-a. An 80% error in the ANN-observed OC3M observation is prescribed on the diagonal elements of **R**.
- EXP 2 (red line, Figure 6): Simulated OC3M (calculated from simulated RSR at 443, 488 and 551 nm, as described by Baird et al., 2016b) is used to calculate the forecast innovations. The state variable and observation errors are the same as EXP1.
- EXP 3 (black line, Figure 6): The same configuration as EXP 2, with a reduced observation error imposed on the diagonal of the observation error covariance matrix.
- EXP 4 (blue line, Figure 6): Additional variables are included in the assimilation state vector, which now comprises of small and large phytoplankton, ammonia, nitrate and total suspended solids concentrations. The observation error is the same as EXP3.
- EXP5 (dashed magenta line, Figure 6): The assimilation vector as EXP4, but using R551 observations, with a lower observation error of 0.2. Note that these innovations relate to the difference in simulated and observed R551, not OC3M as in EXP1-4.

The forecast innovation (Eq 6) statistics for Experiments 1-5 (Figure 6) provide insight into the assimilation system performance. An optimal assimilation system should result in mean forecast innovations (mismatches between observations and the model) of close to 0, and low mean absolute innovations. The assimilation system where the observation operator assumed that there was a direct relationship between simulated surface Chl-a and ANN-observed OC3M (EXP1, Figure 6 green line), performed very poorly and was discontinued after 9 cycles; the model at times became numerically stiff, requiring the adaptive $4^{th} - 5^{th}$ order Ordinary Differential Equation (ODE) integrator to take progressively smaller steps. The innovation statistics for EXP1 suggested the model was constantly over-predicting Chl-a with the mean absolute innovation exceeding 0.7 more than 50% of the time. Calculating the forecast innovations with simulated OC3M and ANN-observed OC3M, rather than simulated surface Chl-a and observed ANN-observed OC3M, improved innovation statistics dramatically (EXP2-4, Figure 6). The configuration used for EXP4 (Figure 6, blue line) gave the best performance when assimilating OC3M. While EXP5 configuration gave best performance in terms of a forecast statistic, a further examination against independent (non-assimilated) observations is given in Sections 3.2.2 – 3.2.5.

3.2.2 Independent assimilation system assessment: Glider

The control and assimilating runs were compared with with-held ocean glider data that was deployed on 26th May 2013, and recovered on the 4th of August 2013. The glider track largely followed the shelf break and headed in a south-easterly direction. To make a comparison between glider observations and the model, we take a sub sample of glider observations centered at the time of model output, with a time window of two hours. For each glider observation that falls within this time period, we find a corresponding 3-D cell from the model and extract the equivalent model solution. No 3-D interpolation is performed as the non-interpolated solution will yield insight into the unresolved sub grid scale variability.

A persistent feature of the observed Chl-a (Figure 7, top panel) is the relatively low values in the upper 100 m of water column. For much of the record, there is a persistent weak deep chlorophyll maximum (DCM) that is centered at between 80 and 120 m depth. Rarely do concentrations in the upper 80 m exceed 0.5 mg/m^3. When the control run is examined, Chl-a in the top 80 m regularly exceeds 1 mg/m^3, and DCMs when they exist, are located between 30 and 60 m deep. A detailed analysis of the control run demonstrates the model is able to reliably produce DCMs (contained in the 180 page skill assessment report at: https://research.csiro.au/ereefs/models), however, in this particular location in time and space, the model does not generate one consistent with the observations. Using the EXP1 configuration, the assimilation system does not improve the solution when compared with the control run. However, using the EXP4 and EXP5 configurations, the assimilation system improves on the control when compared to the glider observations. The assimilation cannot place the DCM in the correct location because the remote sensing observations provide no information about such a deep feature. The remote sensing observations do remove the bias in the upper 80 m with concentrations in the assimilating run ranging between 0.1 - 0.3 mg/m^3 (Figure 7). The configuration used in EXP5 captures the DCM that is predicted on the 14/7/13 and also maintains the lowest mixed layer Chl-a for the feature captured on from the 7/7/13 – 13/7/13 (Figure 7). Assimilating OC3M still over-predicts the mixed layer Chl-a.

A comparison between individual profiles from the glider and equivalent sampling of the model shows substantial unresolved sub-grid scale variability (or instrument noise). Figure 8 (top left) shows that in the upper 70 m the observed Chl-a as measured by the fluorometer ranges between 0.14 and 0.3 mg/m^3, with a mean value of 0.18 mg/m^3. There is no quenching evident in this or other profiles, as the observed fluorescence is constant in the upper 50 m. Within each of the subplots, the glider profiles collected over the 2 h window may sample multiple model grid cells. However, due the relatively slow horizontal glider speed, these glider profiles fall within two adjacent model cells. In most cases, the control and assimilating runs gave indistinguishable solutions between adjacent with the exception of Figure 8, top left. In all cases the assimilation of observations of both OC3M and R551 have reduced the error in the simulated Chl-a when assessed against independent glider data. The noise in the fluorometer observations appears to range between 0.04 and 0.08 mg m^{-3}.

The Chl-a RMSD for each layer of the model, and aggregated in time for all glider observations, is shown in Figure 9. Above 80 m, there is a substantial reduction in RMSD between the control and assimilating runs (EXP4 and EXP5). The RMSD from EXP1 is marginally higher than the RMSD in the control run. The assimilating runs (EXP4 and EXP5) have an RMSD of between 0.10 and 0.17 mg m^{-3} compared with 0.30 to 0.41 mg m^{-3} in the control run. The assimilation of R551 (EXP5) results in a minor improvement in RMSD when compared with the assimilation of OC3M (EXP4). Below 80 m, the RMSD profile is similar between the control and assimilating runs. However, the assimilation of R551 results in the lowest RMSD below 80 m. The largest impact of assimilating remote sensing observations is in the top 80 m of the water column.

3.2.3 Independent assimilation system assessment: GBRMPA MMP Chl-a and TSS

The GBRMPA MMP moorings were deployed at 13 sites in the shallow inshore regions of the Great Barrier Reef Lagoon (Figure 1). The control run typically had RMSDs of between 0.4 and 0.6 mg/m^3 for *in situ* Chl-a (Figure 10). In most cases the assimilation of OC3M (EXP4) reduced the RMSD of *in situ* Chl-a by 0 – 10 %. The assimilation of R551 (EXP5) was more variable. Some sites improved, whereas other were degraded by the assimilation of this observation type. In only one case did the assimilation of R551 beat OC3M for Chl-a at the GBRMPA MMP sites.

The TSS RMSD varies widely across all the GBRMPA MMP sites (Figure 11), driven by the strong variation in magnitude of the spring – neap tidal forcing. The combination of perturbed forcing, and the inclusion of the TSS constituents in the state vector in the assimilating model, has generated realistic time varying correlations between the observed OC3M/R551 and inshore TSS. These cross-correlations allow for the correction of simulated TSS from OC3M/R551 observations. The TSS RMSD at all sites for the assimilating run is less than 5 to 20% that of the control run. The performance of EXP4 and EXP5 were indistinguishable.

3.2.4 Independent assimilation system assessment: Nutrients

Within the GBR region, there are two IMOS NRS sites (Yongala and North Stradbroke, NS, Figure 1). The dissolved inorganic nutrients of NO_3, NH_4 and DIP are taken monthly. At Yongala, water samples are taken at the surface (0 m), 10 m, 20 m and bottom (26 m). At NS, samples are taken at the surface (0 m) and 10 m. It should be noted that there are only 3 to 4 samples per depth at each site during the 4 month simulation period. At Yongala, typical control run RMSDs for NO_3 range from 5 to 12 mg/m^3. Assimilating OC3M/R551 halved these errors. The improvement at NS is evident with improvements found in the upper 10 m of the water column, with EXP4 and EXP5 giving indistinguishable results.

With the exception of the surface samples at Yongala, the assimilation system improved the prediction of NH_4 at all depths for each site. Most notably was the 70 to 90% reduction in RMSD at the deeper locations at Yongala. There were marginal improvements to DIP, which displayed a 0 to 30% reduction in RMSD across all sites. EXP4 resulted in lower DIP errors at NS, but EXP5 resulted in lower DIP errors at Yongala.

3.2.5 Summary of assimilation system experiments

The direct assimilation of RSR (EXP5: R551), or a function of RSR, *f(RSR)*, (EXP4: OC3M) improved the solution when compared with the control run against three independent non-assimilated in-situ observational datasets. The subtle differences between the two approaches will be discussed in Section 4. Based on these findings, the preferred assimilation system for state estimation on the GBR is that used in EXP4. A further examination of the EXP4 forecast error statistics and examples forecast, analysis and increment field is given Section 3.3.

3.3 EXP4: Assimilation system forecast errors.

The assimilation system was run with a 5-day forecast cycle. Using the forecast at *t+5* days and comparing the temporal mean (across all cycles) of the Root Mean Square Difference

(RMSD) and Percentage Error against observations provides insight into the value of the assimilation system, when compared with a non-assimilating system. By comparing the forecast fields against yet to be assimilated observations, we are providing a semi-independent estimate of forecast skill. Additionally, by comparing the forecast against the persisted analysis field from the previous analysis cycle (e.g. the analysis field from t-5 days), it can be determined if the dynamic model is adding skill to the forecast.

A comparison of simulated OC3M and observed OC3M for the non-assimilating control run gives a domain wide median error (range) of 0.32 (0.27 – 0.48) mg m^{-3} (Figure 12). This is approximately equivalent to a domain-wide median percentage error (range) of 100% (80% - 130%) (Figure 12). The data assimilation system reduces the forecast errors and percentage errors to a median value (range) of 0.23 (0.20 – 0.30) mg m^{-3} and 55% (43% - 63%) respectively. The analysis errors are again reduced when observations are assimilated, with median and percentage errors (range) of 0.19 (0.14 – 0.23) mg m^{-3} and 39% (37% - 42%) respectively. When the analysis field from the previous assimilation cycle is persisted forward, the errors (and percentage errors) slightly exceed that of the forecast field with values of 0.26 (0.21 – 0.29) mg m^{-3} and 52% (44% - 65%) respectively. However, it is not expected that these error statistics are spatially uniform given the large percentage of area that is dominated by deep oceanic waters. To understand the spatial variability of the forecast error statistics, the whole domain is divided into three regions representing shallow coastal waters (depth < 30 m), lagoon and shelf waters (30 m < depth < 500 m), and deep oceanic waters (depth > 500 m).

In shallow coastal areas (Figure 12, 2nd column), the non-assimilating control run has a median error (range) of 1.35 (1.1 – 2.45) mg m^{-3}, which corresponds to a percentage error (range) of 130 (105 – 180) %. The distribution of control run errors in the coastal zone is positively skewed, with the mean value of the distribution sitting some way from the median. The assimilation system marginally reduces the median forecast error when compared with the control run, though most notably, it reduces the median percentage error and associated variability. The forecast also beats persistence in this region. There is a marked improvement for lagoon and shelf waters with the assimilation system reducing the median error from 0.34 to 0.25 mg m^{-3}, which corresponds to a reduction in percentage error from 96% to 48%. In the oceanic regions of the domain, the assimilation system reduces the error from 0.16 to 0.10 mg m^{-3}, corresponding to a percentage error reduction from 91% to 45%. In all cases the forecast fields beat persistence. A summary of the results can be found in Table 2 and Table 3.

3.3.1 EXP4: Forecast, Increment and Analysis Fields

The sum of the surface *Trichodesmium* Chl-a, Small Phytoplankton (PhyS) Chl-a and Large Phytoplankton (PhyL) Chl-a biomass differs substantially from the simulated OC3M as shown in Figure 5. The assimilation system updates all of the state variables included in the assimilation state vector. Simulated OC3M is a diagnostic variable that is a function of all the optically-active dynamic state variables as described in Section 2. To demonstrate the impact on the dynamic variables of PhyS and PhyL, results from the forecast step, and the assimilation update, are presented in Figure 13, Figure 14 and Figure 15. Cycle 22 was chosen to demonstrate the spatial impact of the system as it is representative of the last 10 cycles and was relatively cloud free.

The simulated OC3M forecast field for cycle 22 (12th September 2013) displays elevated OC3M in the shallow near shore environment throughout the whole of GBR region and southern shelf of Papua New Guinea (PNG) (Figure 13). Additional features are elevated OC3M in the vicinity of the central and southern fringing reefs, and a plume originating from the eastern region of PNG. Offshore oceanic waters generally have low OC3M of 0.2 mg m^{-3} or less. There is some evidence of mesoscale blooms in the northern and southern sections of the domain. Observed OC3M is overlaid on Figure 13 (left). Where there is a difference in colour, the simulated OC3M differs from the observed OC3M.

The forecast surface layer fields for PhyS and PhyL appear substantially different to the simulated OC3M field (Figure 13). The differences near the coast are where TSS and CDOM are known to cause artefacts in OC3M. While there are patchy blooms of small phytoplankton at various locations within the domain, rarely does the PhyS Chl-a exceed 0.5 mg m^{-3}. The exception to this in the inner central coastal region of the GBR and in the vicinity of the Fly River plume on the south coast of PNG. Similarly, the PhyL Chl-a remains very low for large areas of the domain, however in regions with additional nutrient supply (e.g. in upwelling regions, mesoscale eddies and some river mouths) blooms do occur.

When the observed OC3M is assimilated, increment fields are calculated using Eq 5 and are presented in Figure 14 for simulated OC3M, PhyS Chl-a and PhyL Chl-a. The innovations are overlaid on the increment field for OC3M to give an indication of how well we are fitting the observations. In areas where the model is over-predicting OC3M, the increments will be negative. In areas where the model is under-predicting OC3M, the increments will be positive. The increments and innovations here are presented as a fractional change with respect to the background (forecast) field.

For this particular analysis cycle, it appears that the model is under-estimating inshore OC3M by up to 10-30% and over-estimating OC3M by upwards of 50% offshore (Figure 14, left). By using the background ensemble correlation structure, the increments applied to PhyS biomass increase its concentration in the inner lagoon by up to 20%, and substantially increase the PhyS biomass offshore of the central outer reefs by more than 50% (Figure 14, centre). It should be noted that the increments being applied to the background fields contain meso and sub-meso scale information. Significantly, features such as upwelling filaments, eddies and plumes are maintained through the assimilation procedure, demonstrating that they are allowed to dynamically evolve in the assimilation system. The increments applied to PhyL biomass (Figure 14, right) differ substantially to those of PhyS biomass (Figure 14, centre). For large areas of the domain, the assimilation system decreases the PhyL biomass by up to 50%, whereas there are some areas that it increases. These areas correspond to regions where a bloom may be occurring (there is a small westward shift in the major bloom off the Papua New Guinea coast). The increment applied to the central region of the domain, offshore of the outer reefs, is linear and coherent and likely a result of shifting a dynamic feature such as an upwelling-induced bloom to better match observations.

When the increments contained in Figure 14 are applied to the forecast fields, the resulting analysis field for simulated OC3M better fits the observed OC3M, with a substantially

reduced error inshore and in the vicinity of the outer reefs (Figure 15). The difference between simulated and observed OC3M is small in the deeper offshore regions and shallow sections of the lagoon. The greatest error in OC3M occurs in the central lagoon and the outer reefs where spatial variability is highest. The corresponding analysis fields for small phytoplankton and large phytoplankton are contained in Figure 15. There are elevated concentrations of small phytoplankton biomass in the near shore region near river mouths and the outer fringing reefs. The large phytoplankton biomass is concentrated in the region of Broad Sound, the Fly River plume, and the Papua New Guinea upwelling. Each of these features was predicted by the forecast, as little biomass is added or subtracted by the assimilation update. However, there is substantial removal of large phytoplankton biomass from the northern and central offshore regions. This leaves very little large phytoplankton biomass present in substantial areas of the domain during this particular analysis cycle.

4 Discussion

In the optically-complex waters of the GBR, the use of an optics model to calculate a simulated RSR and subsequent simulated OC3M to constrain the BGC model substantially reduces the errors in biogeochemical state. This has been achieved by explicitly assimilating like-for-like variables of simulated R551 or simulated OC3M. The data assimilation is constrained by the mis-match between simulated and observed R551 and OC3M. A summary of the root mean square differences (RMSDs) is contained in Table 4. Our approach of simulating the observation is the opposite to the conversion of observed RSRs into modelled variables, e.g. the assimilation of phytoplankton functional types, TSS, CDOM and Chl-a. We therefore advocate for the use of a "forward" model to predict optical properties, rather than rely on the "inversion" to back calculate the model state from reflectances. The conversion of RSRs into derived variables (e.g. Chl-a, PFTs) have associated errors that are as large as 300% in some locations, and can be biased by up to 70% (Qin et al., 2007), whereas the forwards approach has errors of approximately 20% (Baird et al 2016b). The errors stemming from the "inversion" technique are at times difficult to characterise in data assimilation systems. Using a forwards model avoids these errors and therefore the dominant source of error stems from model error, rather than observation error (via difference in kind errors). The approach of Ciavatta et al., (2014) contained similar results where they found the assimilation of Kd_{443}, to be superior to remote-sensed Chl-a.

A significant source of error in algorithms such as OC3M is that they produce a single value for each horizontal pixel, generally considered to be representative of the first optical depth of the water column. If this is to be compared to a single value in biogeochemical model, then it must be assumed that the water column is well-mixed to the optical depth, and that there is an equal optical depth of each of the wavebands used in the algorithm. Both of these conditions are rarely met in coastal waters. Matching RSR requires no assumptions about the structure of the water column, or of the vertical distribution of the optically-active constituents, because both observed and modelled quantities are two dimensional fields.

EXP1 performed poorly for a number of reasons. It is likely that even an 80% observation error was insufficient to adequately account for positive biases in the OC3M observations.

These biases, present even in offshore waters, lead to positive innovations that result in adding phytoplankton biomass in the increments. These large increments lead to persistently high biomass, that draw available nutrients down to very low levels. The only way to account for this form of observation error is to inflate the "difference in kind error" to a large value. Given that the non-assimilating control run of the model had forecast errors that range between 70%-100% (region dependent), running an assimilation system with an observation error larger than the error present in the non-assimilating model, does not make sense.

The similarity in RMSDs between EXP4 and EXP5 (Table 4) suggests there is little difference between using simulated R551 and simulated OC3M. Univariate observations were assimilated in each case, and the information content of the observations is likely going to be similar in deep water, hence the similar results for the Ocean Glider. However, due to the simulated OC3M containing information from 443 and 488 nm, which are important in shallow regions due the absorption of these wavelengths from sediments and CDOM, there is additional information in the OC3M observation that is not present in the 551 nm RSR. Ultimately the greatest information content will come from the simultaneous assimilation of multivariate observations obtained from the multi-spectral OC sensors and including the longer wavelengths.

4.1 True color visualization of EXP4

In order to visualize the impact of the assimilation system on the prediction of water clarity, we compare the observed true color (Figure 16) with the simulated true color of the control run (Figure 17, top left) and the EXP4 assimilating run (Figure 17, top right). Simulated true colour images are generated from RSR at the red, green and blue wavelengths calculated using the optical model and the three dimensional fields of the model-predicted 23 optically-active constituents.

The observed true colour image on the 12 Sep 2013 shows brown / yellow features associated with high suspended sediment concentrations. As these concentrations become more diluted, and mixed with phytoplankton, the water appears more greenish blue. Offshore reefs, with clear water above white substrates, appear as light blue features, with the intensity depending on the reef depth. Qualitatively, the control run (Figure 17, top left) does a reasonably good job of reproducing the observed true colour. The quantification of this mismatch can be done on individual color bands (not shown, Baird et al., 2016b). Qualitatively, the control run does not have enough suspended solids in the surface water in the mouth of Broad Sound (22.2°S, 149.5°E), and has too high phytoplankton concentrations offshore, especially in a feature centred at 23°S, 151.5°E. The assimilated run, while not that different to the control run, corrects some of these errors.

To approximately quantify impacts of the assimilation of water clarity, it is possible consider the colour of the added (and subtracted) constituents in the assimilation procedure. To avoid confusion with the phrases 'falsely-coloured' or 'negative', which have distinct meanings in visualisation science, but to still provide a phrase for true colour error, we use the term "off-colour", and distinguish between off-colour that requires correction through addition (Figure 14, bottom left) and subtraction (Figure 14, bottom right). The assimilation procedure added yellow colours (suspended sediment) within Broad Sound and green

colours (phytoplankton) in the mouth. Offshore the assimilation removed green, particularly, as noted above at 23°S, 151.5°E. By removing green it made the water more blue (Figure 14, top right).

4.2 Like-for-like assimilation

The general methodology presented in this study is similar to that used in numerical weather prediction (NWP). The NWP community moved away from assimilating satellite-derived temperature and humidity profiles more than two decades ago, in favour of assimilating radiances (or temperature brightness), this approach is detailed in Derber and Wu (1998). By assimilating satellite-derived radiance data, the NWP community avoided any reliance on empirical statistical relationships used to predict the temperature and humidity profiles. Similarly, the approach taken here is to avoid the use of an empirical/statistical inverse model, and use a physics-based forward model to predict RSR centred at the MODIS bandwidths. We then post-process these simulated RSRs into a simulated OC3M. The simulated OC3M is directly comparable to the observed OC3M with both containing quantitatively similar sources of error derived from bottom reflectance, and turbid coastal waters. By avoiding the use of an inverse empirical statistical model, we are presenting a BGC DA approach that had been adopted by the NWP community decades ago (Derber and Wu, 1998; Dee et al., 2015).

An additional benefit of avoiding the use of IOP / AOP based empirical/statistical products, assimilation of RSRs can take advantage non-ocean-color specific missions such as Himawari 8. The spectral resolution can be altered to simulate reflectances at the Himawari 8 true colour bands. If the Himawari 8 data can be assimilated it will provide a step change in the data available for areas such as the GBR, due to the high spatial resolution (nominally 500 m) and temporal resolution (every 30 minutes). This data density far exceeds that available from orbiting satellites, and will provide coverage similar to the products being assimilated in NWP systems.

If multiple reflectance bands are to be simultaneously assimilated, it will be necessary to account for cross-correlation between observation errors. For example, the RSR at 443 nm is strongly correlated with the RSR at 488 nm. Therefore, the observations of adjacent bands are no longer independent and it is likely that the assumption that the off diagonal elements of the observation error covariance matrix (**R**) needs to be reconsidered. Using simulated and observed OC3M eliminates the possibility of cross-correlation and contains information derived from multiple bands. The OC3M algorithm can be considered a band-ratio function, f(RSR), that transforms multi-variate observations into univariate observations. It is likely that there are other function forms that could combine information from multiple bands into a single non-correlated observation.

The underlying configuration of the data assimilation presented in this study requires the dominant error sub-space to be spanned by the ensemble. Pragmatic choices have been made to allow the system to run on the available compute resources. To this end we have perturbed 2 sensitive model parameters, and river loads of nutrients and sediments. The distributions that have been sampled to perturb the zooplankton mortality rates and θ_i along with their respective shape parameters could be considered a subjective choice. There is substantial scope to recast the problem with a Bayesian Hierarchical Modelling

(BHM) framework (as in Parslow et al., 2013; and Dowd et al., 2014), whereby the prior distribution are assigned to uncertain parameters, and a thorough meta-analysis of the literature could be used to construct informative distributions. The observations could then be used to construct not only a posterior over the state, but a full joint posterior over the state and parameters. Furthermore, we have not allowed uncertainty in the physics to propagate into the BGC solution. We recognise this is a shortcoming of the study, however, given the computational constraints, we are not in a position to expand the ensemble to include physics perturbations (which would require an ensemble that is up to an order of magnitude larger). As more computing power becomes available, ensemble sizes could be increased, stochastic parameterisations introduced (Garnier et al., 2016), and DA methods with less parametric assumptions (e.g. Parslow et al., 2013), could be adopted.

There have been two recent discussion papers released that detail the pathway towards operationalising BGC forecasting systems (Gehlen et al., 2015; and Ford and Barciela, 2015), analogous to the current NWP and hydrodynamic predictions system that routinely run at numerous operational centres. It has been acknowledged that satellite remote sensing will play a key role in such systems, there appears to be two divergent pathways to achieve this vision. Ford et al., (2016) advocate for the assimilation of empirical statistical products such as Chl-a and Phytoplankton Functional Types (PFTs), with the alternative being the assimilation of diffuse attenuation coefficient(s) (e.g. Ciavatta et al., 2014). For complex coastal regions that are dominated by case 2 waters, the assimilation of RSRs avoids the costly requirement of calibrating an empirical/statistical algorithm that is regionally-specific. Whilst the results from this study have shown to be valuable in the GBR region, further work needs to be undertaken to demonstrate the broad applicability of this approach. Nonetheless, we would advocate a third approach should be considered – the assimilation of RSR.

5 Conclusion

In this study we have used a spectrally-resolved optical model coupled to a BGC model to simulate the remote-sensing reflectances (RSR) centred at the MODIS ocean colour bands. A series of assimilation system configuration experiments were undertaken to test the assimilation system performance. When the simulated OC3M (EXP4) and remote-sensing reflectances (EXP5) were assimilated into the model, the forecast errors in Chl-a fell from 100% to 55% when compared to the non-assimilating model. By using a function of the remote-sensing reflectances (OC3M), information from multiple bands are included in a univariate observation and the forecast error is halved compared to simply assuming the OC3M is directly related to the model prediction of surface total Chl-a. A comparison against in-situ observations of NO_3, NH_4, DIP and TSS shows the assimilating model (EXP4) reduces the MAPE from 90% to less than 20% at most stations. By using a forward model that includes a majority of error sources present in the observed OC3M, we have shown that the assimilation of remotely-sensed products in optically complex case 2 waters can be achieved, and adds substantial predictive skill when compared to the non-assimilating model. Furthermore, this approach can be generalized to non ocean–colour specific missions by assimilating the remote-sensing reflectances directly (e.g. EXP5), liberating a vast quantity of data that cannot be used in traditional BGC assimilation systems.

Acknowledgements

We would firstly like to acknowledge the thorough and insightful comments provided by the three reviewers: Prof Emmanuel Boss (Reviewer 1), Dr Stefano Ciavatta (Reviewer 2) and Dr David Ford (Reviewer 3). The scope and clarity of this manuscript have improved greatly due to their input. The model simulations were developed as part of the eReefs project, a public-private collaboration between Australia's leading operational and scientific research agencies, government, and corporate Australia (research.csiro.au/ereefs). Atmospherically-corrected MODIS products were sourced from NASA and the Integrated Marine Observing System (IMOS) - IMOS is supported by the Australian Government through the National Collaborative Research Infrastructure Strategy (NCRIS) and the Super Science Initiative. Simulations and the processing of the remote sensing data was undertaken with the assistance of resources from the National Computational Infrastructure (NCI), which is supported by the Australian Government via NCRIS. We thank the many colleagues involved in developing the eReefs model, particularly Mike Herzfeld, John Andrewartha, Philip Gillibrand, Cedric Robillot and Richard Brinkman. Mark Baird was additionally funded by the CSIRO Wealth from Oceans Flagship, the Gas Industry Social & Environmental Research Alliance (GISERA), CSS FSP and the CSIRO Coastal Carbon Cluster. River flow data and information used to specify river nutrient and sediment loads were provided by the Queensland Government.

References

Ariathurai, R. & Krone, R. B. 1976. Finite element model for cohesive sediment transport. Journal of the Hydraulics Division, ASCE, 104, HY2, 323–328.

Baird, M. E. (2010) Limits to prediction in a size-resolved pelagic ecosystem model. J. Plankton Res. 32, 1131-1146.

Baird, M. E., M. P. Adams, R. C. Babcock, K. Oubelkheir, M. Mongin, K. A. Wild-Allen, J. Skerratt, B. J. Robson, K. Petrou, P. J. Ralph, K. R. O'Brien, A. B. Carter, J. C. Jarvis, M. A. Rasheed (2016a) A biophysical representation of seagrass growth for application in a complex shallow-water biogeochemical model Ecol. Mod. 325: 13-27.

Baird, M. E., N. Cherukuru, E. Jones, N. Margvelashvili, M. Mongin, K. Oubelkheir, P. J. Ralph, F. Rizwi, B. J. Robson, T. Schroeder, J. Skerratt, A. D. L. Steven and K. A. Wild-Allen (2016b) Remote-sensing reflectance and true colour produced by a coupled hydrodynamic, optical, sediment, biogeochemical model of the Great Barrier Reef, Australia: comparison with satellite data. Env. Model. Software 78: 79-96.

Baird, M. E., P. J. Ralph, K. Wild-Allen, F. Rizwi and A. D. L. Steven (2013). A dynamic model of the cellular carbon to chlorophyll ratio applied to a batch culture and a continental shelf ecosystem. Limnol. Oceanogr. 58, 1215-1226.

Blondeau-Patissier, D., Brando, V.E., Oubelkheir, K., Dekker, A.G., Clementson, L.A., Daniel, P., 2009. Bio-optical variability of the absorption and scattering properties of the Queensland inshore and reef waters, Australia. J. Geophys. Res. (Oceans) 114, C05003.

Carmillet, V., Brankart, J.M., Brasseur, P., Drange, H., Evensen, G. and Verron, J., 2001. A singular evolutive extended Kalman filter to assimilate ocean color data in a coupled physical–biochemical model of the North Atlantic Ocean. *Ocean Modelling*, *3*(3), 167-192.

Ciavatta, S., Torres, R., Martinez-Vicente, V., Smyth, T., Dall'Olmo, G., Polimene, L. and Allen, J.I., 2014. Assimilation of remotely-sensed optical properties to improve marine biogeochemistry modelling. *Progress in Oceanography*, *127*, 74-95.

Ciavatta, S., Kay, S., Saux-Picart, S., Butenschön, M. and Allen, J.I., 2016. Decadal reanalysis of biogeochemical indicators and fluxes in the North West European shelf-sea ecosystem. *Journal of Geophysical Research: Oceans*. 121, 1824-1845

Collier, C.J.,Waycott, M., McKenzie, L.J., 2012. Light thresholds derived from seagrass loss in the coastal zone of the northern Great Barrier Reef, Australia. Ecol. Indic. 23, 211-219.

Cummings, J.A., 2005. Operational multivariate ocean data assimilation. *Quarterly Journal of the Royal Meteorological Society*, *131*(613), 3583-3604.

De'ath, G. K. E. Fabricius, H. Sweatman and M. Puotinen (2012) The 27-year decline of coral cover on the Great Barrier Reef and its causes. PNAS 109, 17995-17999.

Dee, D. P. et al. (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q. J. R. Meteorol. Soc. 137: 553 – 597.

Derber, J.C. and Wu, W.S., 1998. The use of TOVS cloud-cleared radiances in the NCEP SSI analysis system. *Monthly Weather Review*, *126*(8), 2287-2299.

Dowd, M., Jones, E. and Parslow, J., 2014. A statistical overview and perspectives on data assimilation for marine biogeochemical models. *Environmetrics*, *25*(4), 203-213.

Dutkiewicz, S., Hickman, A. E., Jahn, O., Gregg, W. W., Mouw, C. B., & Follows, M. J. (2015). Capturing optically important constituents and properties in a marine biogeochemical and ecosystem model.

Egbert, G.D. and Erofeeva, S.Y., 2002. Efficient inverse modeling of barotropic ocean tides. Journal of Atmospheric and Oceanic Technology, 19(2), 183-204.

Ford, D.A., Edwards, K.P., Lea, D., Barciela, R.M., Martin, M.J. and Demaria, J., 2012. Assimilating GlobColour ocean colour data into a pre-operational physical-biogeochemical model. *Ocean Science*, *8*(5), 751-771.

Ford, D., & Barciela, R., 2015. Forecasting Research Technical Report No: 609 - Marine biogeochemical data assimilation–literature review and scoping report. 46 .

Fujii, M., Boss, E., & Chai, F. (2007). The value of adding optics to ecosystem models: a case study. *Biogeosciences Discussions*, *4*(3), 1585-1631.

Furnas, M. 2003. Catchments and corals: terrestrial runoff to the Great Barrier Reef. Australian Institute of Marine Science, Queensland., 334 p.

Garnier, F., Brankart, J.M., Brasseur, P. and Cosme, E., 2016. Stochastic parameterizations of biogeochemical uncertainties in a 1/4° NEMO/PISCES model for probabilistic comparisons with ocean color data. *Journal of Marine Systems*, *155*, 59-72.

Gaspari, G. and Cohn, S.E., 1999. Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, *125*(554), 723-757.

Gehlen, M., Barciela, R., Bertino, L., Brasseur, P., Butenschön, M., Chai, F., Crise, A., Drillet, Y., Ford, D., Lavoie, D. and Lehodey, P., 2015. Building the capacity for forecasting marine biogeochemistry and ecosystems: recent advances and future developments. *Journal of Operational Oceanography*,*8*(sup1), s168-s187.

Gillibrand, P.A. and Herzfeld, M., 2016. A mass-conserving advection scheme for offline simulation of scalar transport in coastal ocean models. *Ocean Modelling*, *101*, 1-16.

Goyens, C., Jamet, C., Schroeder, T., 2013. Evaluation of four atmospheric correction algorithms for MODIS-Aqua images over contrasted coastal waters. Rem. Sens. Env. 131, 63-75.

Gregg, W.W., 2008. Assimilation of SeaWiFS ocean chlorophyll data into a three-dimensional global ocean model. *Journal of Marine Systems*, *69*(3), 205-225.

Herzfeld, M. 2006, An alternative coordinate system for solving finite difference ocean models., Ocean Modelling, 14, 174-196.

Herzfeld, M., and P. A. Gillibrand. "Active open boundary forcing using dual relaxation time-scales in downscaled ocean models." *Ocean Modelling* 89 (2015): 71-83.

Hu, J., Fennel, K., Mattern, J.P. and Wilkin, J., 2012. Data assimilation with a local Ensemble Kalman Filter applied to a three-dimensional biological model of the Middle Atlantic Bight. *Journal of Marine Systems*, *94*, 145-156.

King, E. A., Schroeder, T., Brando, V. E., Suber, K., 2014. A pre-operational System for Satellite Monitoring of the Great Barrier Reef Marine Water Quality. Tech. rep., CSIRO Report, pp56.

Lee, Z., Carder, K. L., & Arnone, R. A. (2002). Deriving inherent optical properties from water color: a multiband quasi-analytical algorithm for optically deep waters. *Applied optics*, *41*(27), 5755-5772.

Lynch TP, Morello EB, Evans K, Richardson AJ, Rochester W, et al. (2014) IMOS National Reference Stations: A Continental-Wide Physical, Chemical and Biological Coastal Observing System. PLoS ONE 9(12): e113652. doi: 10.1371/journal.pone.0113652.

Madsen, O.S., 1994. Spectral wave-current bottom boundary layer flows, in Coastal Engineering 1994 Proceedings, 24th International Conference Coastal Engineering Research Council/ASCE, 384-398.

Margvelashvili, N., 2009. Stretched Eulerian coordinate model of coastal sediment transport. Comput. Geosci. 35, 1167-1176.

Margvelashvili, N., Saint-Cast, F., Condie, S., 2008. Numerical modelling of the suspended sediment transport in Torres Strait. Continental Shelf Research 28, 2241-2256.

Mongin, M., M. E. Baird, B. Tilbrook, R. J. Matear, A. Lenton, M. Herzfeld, K. A. Wild-Allen, J. Skerratt, N. Margvelashvili, B. J. Robson, C. M. Duarte, M. S. M. Gustafsson, P. J. Ralph, A. D. L. Steven (2016). The exposure of the Great Barrier Reef to ocean acidification. Nature Communications 7, 10732.

Morel, A., Antoine, D., and Gentili, B., 2002. Bidirectional reflectance of oceanic waters: accounting for Raman emission and varying particle scattering phase function. *Applied Optics* 41(30), 6289-6306.

Morello EB, Plagányi ÉE, Babcock RC, Sweatman H, Hillary R, Punt AE (2014) Model to manage and reduce crown-of-thorns starfish outbreaks. Mar. Ecol. Prog. Ser. 512:167-183

Natvik, L.J. and Evensen, G., 2003. Assimilation of ocean colour data into a biochemical model of the North Atlantic: Part 1. Data assimilation experiments. *Journal of Marine Systems*, *40*, 127-153.

Odermatt, D., Gitelson, A., Brando, V.E. and Schaepman, M., 2012. Review of constituent retrieval in optically deep and complex waters from satellite imagery. Remote sensing of environment, 118, 116-126.

Oke, P.R., Brassington, G.B., Griffin, D.A. and Schiller, A., 2008a. The Bluelink ocean data assimilation system (BODAS). Ocean Modelling, 21(1), 46-70.

Oke, P.R. and Sakov, P., 2008b. Representation error of oceanic observations for data assimilation. *Journal of Atmospheric and Oceanic Technology*, *25*(6), 1004-1017.

Parslow, J. S., Cressie, N., Campbell, E., Jones. E. M., and Murray, L. 2013. Bayesian learning and predictability in a stochastic nonlinear dynamical model. Ecological Applications 23:4, 679-698

Qin, Y., Brando, V.E., Dekker, A.G. and Blondeau-Patissier, D., 2007. Validity of SeaDAS water constituents retrieval algorithms in Australian tropical coastal waters. *Geophysical Research Letters*, *34*(21).

Ridgway K.R., J.R. Dunn, and J.L. Wilkin 2002 Ocean interpolation by four-dimensional least squares -Application to the waters around Australia. J. Atmos. Ocean. Tech., 19, 1357-1375.

Robson, B.J., Baird, M. and Wild-Allen, K., 2013. A physiological model for the marine cyanobacteria, *Trichodesmium*. In MODSIM2013, 20th International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, ISBN (978-0).

Rolfe, J. and Gregg, D., 2015. Factors affecting adoption of improved management practices in the pastoral industry in Great Barrier Reef catchments. Journal of environmental management, 157, 182-193.

Sakov, P. and Oke, P.R., 2008. A deterministic formulation of the ensemble Kalman filter: an alternative to ensemble square root filters. *Tellus A*, *60*(2), 361-371.

Sakov, P. and Bertino, L., 2011. Relation between two common localisation methods for the EnKF. *Computational Geosciences*, *15*(2), 225-237.

Sarmiento, J. L and N. Gruber (2006) Ocean Biogeochemical Dynamics. Princeton University Press. ISBN: 9780691017075. 528

Schiller, A., Herzfeld, M., Brinkman, R., Stuart, G., Jan. 2014. Monitoring, predicting and managing one of the seven natural wonders of the world. Bull. Am. Meteor. Soc. 23-30.

Schroeder, T., Behnert, I., Schaale, M., Fischer, J., Doerer, R., 2007. Atmospheric correction algorithm for MERIS above Case-2 water. J. Int. Remote Sens. 28, 1469-1486.

Schroeder, T., Lovell J., King E., Clementson L, and Scott R, (2016), *IMOS Ocean Colour Validation Report 2015-16*.

Schaeffer, A., Roughan, M., Jones, E. M. and White, D. (2016) Physical and biogeochemical spatial scales of variability in the East Australian Current separation from shelf glider measurements Biogeosciences 13(6): 1967-1975.

Shulman, I., Frolov, S., Anderson, S., Penta, B., Gould, R., Sakalaukus, P. and Ladner, S., 2013. Impact of bio-optical data assimilation on short-term coupled physical, bio-optical model predictions. *Journal of Geophysical Research: Oceans*, *118*(4), 2215-2230.

Skerratt, J., Wild-Allen, K., Rizwi, F., Whitehead, J. and Coughanowr, C., 2013. Use of a high resolution 3D fully coupled hydrodynamic, sediment and biogeochemical model to understand estuarine nutrient dynamics under various water quality scenarios. Ocean & coastal management, 83, 52-66.

Thompson, A., Costello, P., Davidson, J., Logan, M., Schaffelke, B., Uthicke, S. and Takahashi, M., 2011. Reef rescue marine monitoring program. Report of AIMS Activities–Inshore coral reef monitoring, p.128.

Thompson, A., Schroeder, T., Brando, V.E., Schaffelke, B., 2014. Coral community responses to declining water quality: Whitsunday Islands, Great Barrier Reef, Australia. Coral Reefs 33, 923-938.

Willmott, C.J., Ackleson SG, Davis RE, Feddema JJ, Klink KM,Legates DR, ODonnell J, Rowe CM. 1985. Statistics for the evaluation of model performance. Journal of Geophysical Research 90: 8995-9005.

Appendix A: Detailed description of the eReefs modelling system

The eReefs modelling system is a suite of coupled hydrodynamic, sediment, optical and biogeochemical models specifically tailored to the Great Barrier Reef. The hydrodynamic model is a three-dimensional, finite-difference, baroclinic model based on the three-dimensional equations of momentum, continuity and conservation of heat and salt, employing the hydrostatic and Boussinesq assumptions (Herzfeld et al., 2006, Schiller et al., 2015). The equations of motion are discretized on a finite-difference stencil corresponding to the Arakawa C grid. In the vertical z-coordinate scheme, there are 47 fixed z-levels. The atmospheric forcing products (wind, pressure, rain and heat fluxes) are supplied by Bureau of Meteorology (BOM) reanalysis products. A tidal signal was superimposed on the low-frequency sea level oscillation provided by BRAN2.3 (Oke et al., 2008a) on the regional grid open boundary. This tidal signal was introduced via a local flux adjustment. The OTIS tidal model (Egbert and Erofeeva, 2002) was used to generate the tidal signal from amplitude and phase information for 8 constituents. The local grid open boundary was forced with temperature, salinity and velocity (with local flux adjustment) derived from the regional grid. A mass conserving flux-based advection scheme is used to transport sediment and biogeochemical tracers.

The sediment transport model adds a multilayer sediment bed to the hydrodynamic model grid and simulates sinking, deposition and resuspension of multiply size-classes of suspended sediment (Margvelashvili et al., 2008). The model solves advection-diffusion equations of the mass conservation of suspended and bottom sediments and is particularly suitable for representing fine sediment dynamics, including resuspension and transport of biogeochemical particles. The model is initialised with the observed distribution of gravel, sand and mud in the seabed of the shelf region. Sediment particles settle on the seabed due to gravity and resuspend into the water column whenever the bottom shear stress, exerted by waves and currents, exceeds the critical shear stress of erosion. The resuspension and deposition fluxes are parameterised with the Ariathurai and Krone (1976) formula. The bottom friction under combined waves and currents is estimated through the nonlinear bottom boundary layer model (Madsen, 1994).

Sediments in benthic layers undergo vertical mixing due to bioturbation, represented by local diffusion. The corresponding diffusion coefficient scales with the sediment depth so that the bioturbation ceases to operate beneath the biologically active layer. The resistance of sediments to resuspension also varies with the sediment depth to reflect the consolidated nature of deep sediments. The numerical grid for sediment variables in the water column coincides with the numerical grid for the hydrodynamic model. Within the bottom sediments, the model utilises a time-varying sediment-thickness-adapted grid, where the thickness of sediment layers varies with time to accommodate the deposited sediment. Horizontal resolution within sediments follows the resolution of the water column grid.

The biogeochemical model is organised into 3 zones: pelagic, epibenthic and sediment. The epibenthic zone overlaps with the lowest pelagic layer and the top sediment layer, sharing the same dissolved and suspended particulate material fields. Dissolved and particulate

biogeochemical tracers are advected and diffused throughout the model domain. Additionally, biogeochemical particulate substances sink and are resuspended in the same way as sediment particles. Biogeochemical processes are organized into pelagic processes of phytoplankton and zooplankton growth and mortality, remineralisation of particulate and organic material, and fluxes of dissolved oxygen, nitrogen, phosphorus and carbon (including nitrogen fixation, phosphorus adsorption and desorption, surface gas exchanges, respiration and phorosythensis, and fluxes to and from biotic pools); epibenthic processes of growth and mortality of macroalgae, seagrass and corals, and sediment based processes of phytoplankton mortality, microphytobenthos growth, detrital remineralisation and fluxes of dissolved substances (Fig.2).

The biogeochemical model includes four groups of microalgae (small and large phytoplankton, *Trichodesmium* and microphytobenthos) and three macrophytes types (seagrass types corresponding to *Zostera* and *Halophila*, macroalgae and coral communities). Photosynthetic growth is determined by concentrations of dissolved nutrients (nitrogen and phosphate) and photosynthetically active radiation. Autotrophs take up dissolved ammonium, nitrate, phosphate and inorganic carbon, and in the case of *Trichodesmium*, fix atmospheric nitrogen (Robson et al., 2014). Microalgae incorporate carbon (C), nitrogen (N) and phosphorus (P) at the Redfield ratio (106C:16N:1P, Redfield 1963) while macrophytes do so at the Atkinson ratio (550C:30N:1P, Atkinson1983). Microalgae contain two pigments (chlorophyll-*a* and an accessory pigment), and have variable carbon:pigment ratios determined using a photoadaptation model (described in Baird et al., 2013).

Micro-zooplankton graze on small phytoplankton and meso-zooplankton graze on large phytoplankton and microzooplankton, at rates determined by particle encounter rates and maximum ingestion rates. Of the grazed material that is not incorporated into zooplankton biomass, half is released as dissolved and particulate carbon, nitrogen and phosphate, with the remainder forming detritus. Additional detritus accumulates by mortality. Detritus and dissolved organic substances are remineralised into inorganic carbon, nitrogen and phosphate with labile detritus transformed most rapidly (days), refractory detritus slower (months) and dissolved organic material transformed over the longest timescales (years). The production (by photosynthesis) and consumption (by respiration and remineralisation) of dissolved oxygen is also included in the model and depending on prevailing concentrations, facilitates or inhibits the oxidation of ammonia to nitrate and its subsequent denitrification (in the sediment) to di-nitrogen gas which is then lost from the system. Full details of equations used in the biogeochemical model are given by Baird et al. (2016b) and details of parameter values and implementation for the Great Barrier Reef are given by Herzfeld et al. 2016

The model is forced using flow and concentrations of dissolved and particulate constituents from 21 rivers along the Queensland coast (north to south: Normanby, Daintree, Barron, combined Mulgrave+Russell, Johnstone, Tully, Herbert, Haughton, Burdekin, Don, O'Connell, Pioneer, Fitzroy, Burnett, Mary, Calliope, Boyne, Caboolture, Pine, combined Brisbane+Bremer, and combined Logan+Albert) and the Fly River in Papua New Guinea (Herzfeld et al., 2015). To determine river concentrations, sediment and nutrient observations were statistically evaluated over 10 years (Furnas 2003). Separate analysis was

undertaken for wet- (the Fly, and the northern most 6 rivers in Queensland) and dry- (remainder) catchment rivers. Volume-averaged wet season export coefficients based on this observed dataset were derived for wet- and dry-catchment river types, and mean flow-weighted concentrations determined. These constant concentrations are multiplied by higher frequency (daily) observed discharge data to calculate the flux of constituents at the river mouths.

The eReefs BGC and sediment model has 3 open ocean boundaries. Nutrient concentrations flowing in from the boundaries were obtained from the CSIRO Atlas of Regional Seas (CARS) 2009 climatology (Ridgway et al., 2003) and empirical nutrient-temperature relationships. The initial conditions are specified by a generalised empirical relationship and scaled nutrient profiles on the model density profile specifying top and bottom water column values from CARS ocean atlas. Surface NO_3 is usually low (< 3 mg m^{-3}). In deeper waters nutrient concentrations increase from 0 to 1500 m depth and then remain constant down to the ocean floor (4000 m depth, 500 mg m^{-3}). The initial conditions for most other tracers were not spatially resolved, since observations for the outer reef and Coral Sea are limited temporally and spatially.

Tables:

Table 1: The subset of state variables included in the state vector and corresponding observation error (E_{tot}) standard deviations used for the five assimilation system configurations, and prescribed on the diagonal elements of R. The bold variables in the state vector are used in the input to the observation operator. It should be noted than the state variables are transformed by taking the natural logarithm of the variables. The observation error is then applied to the log-transformed state vector.

	Assimilation State Vector (**X**)	Observation Error (E_{tot})
EXP1	Log(**Surface Total Chl-a**, PhyS Chl-a, PhyL Chl-a)	0.8
EXP2	Log(**Simulated OC3M**, PhyS Chl-a, PhyL Chl-a)	0.8
EXP3	Log(**Simulated OC3M**, PhyS Chl-a, PhyL Chl-a)	0.4
EXP4	Log(**Simulated OC3M**, PhyS Chl-a, PhyL Chl-a, NO3, NH4, TSS)	0.4
EXP5	Log(**Simulated R551**, PhyS Chl-a, PhyL Chl-a, NO3, NH4, TSS	0.2

Table 2: Forecast error statistics for OC3M (mg Chl-a m^{-3}) in EXP4 by region (inshore, lagoon, offshore) for the Control (C), Forecast (F), Analysis (A) and Persistence (P) fields.

Region	Whole of Domain				Coastal				Lagoon and Shelf				Oceanic			
Field	C	F	A	P	C	F	A	P	C	F	A	P	C	F	A	P
Median	0.32	0.23	0.19	0.26	1.35	1.29	1.12	1.46	0.34	0.25	0.19	0.25	0.16	0.1	0.06	0.11
Mean	0.37	0.24	0.2	0.27	1.92	1.37	1.25	1.46	0.38	0.24	0.2	0.25	0.16	0.1	0.06	0.11
25% Quartile	0.27	0.2	0.14	0.21	1.1	0.95	0.92	0.94	0.29	0.16	0.14	0.21	0.14	0.08	0.04	0.7
75% Quartile	0.48	0.3	0.23	0.29	2.45	1.87	1.64	1.08	0.43	0.29	0.22	0.29	0.19	0.13	0.08	0.12

Table 3: Forecast Percentage (%) errors in EXP4 by region (inshore, lagoon, offshore) for the Control (C), Forecast (F), Analysis (A) and Persistence (P) fields.

Region	Whole of Domain				Coastal				Lagoon and Shelf				Oceanic			
Field	C	F	A	P	C	F	A	P	C	F	A	P	C	F	A	P
Median	100	53	39	53	130	95	90	105	95	47	38	52	93	48	31	48
Mean	107	54	38	55	180	97	90	107	102	51	37	56	96	50	31	51
25% Quartile	81	41	35	42	105	105	95	74	75	43	36	41	71	44	35	39
75% Quartile	131	61	41	62	181	85	81	145	118	62	42	62	126	55	30	62

Table 4: Root Means Square Differences (RMSDs) for the withheld observations and the control run (control), assimilation experiment 4 (EXP4) and assimilation experiment 5 (EXP5). The lowest RMSD is given in bold.

Observed Variable	Control	EXP4	EXP5
Chl-a (mg Chl-a m^{-3}) -GBRMPA MMP	0.5327	**0.4524**	0.4957
Chl-a (mg Chl-a m^{-3}) – IMOS Glider	0.2064	0.1344	**0.1232**
TSS (mg solids m^{-3}) -GBRMPA MMP	0.0081	**0.0017**	0.0018
NO3 (mg N m^{-3}) - IMOS NRS	6.5126	**1.1361**	1.6508
NH4 (mg N m^{-3}) - IMOS NRS	4.2795	1.6384	**1.6180**
DIP (mg P m^{-3}) - IMOS NRS	2.1579	**1.3785**	1.5387

Figures:

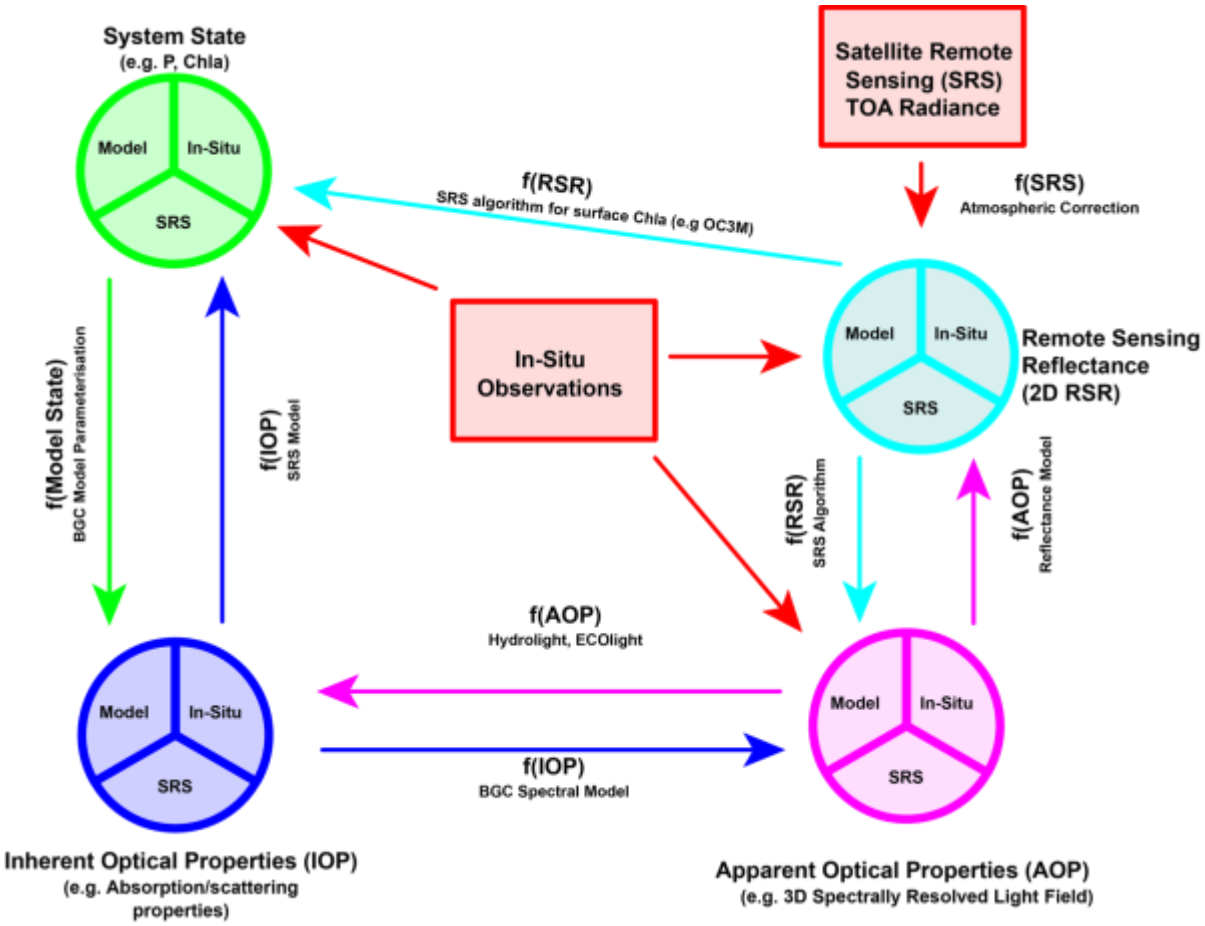

Figure 1: A diagram denoting the steps used to relate and transform various optical properties with the system state and observations. The green circle represents the in-water system state as predicted by a marine BGC model. The blue circle represents the Inherent Optical Properties (IOPs) of the system state. The magenta circle represents the depth resolved Apparent Optical Properties (AOPs). The cyan circle represents the 2D remote-sensing reflectance (RSR). The two red boxes represent either in-situ or satellite remote-sensing observations. Each circle is partitioned into three segments where each segments represents the possibility to compare like for like variables.

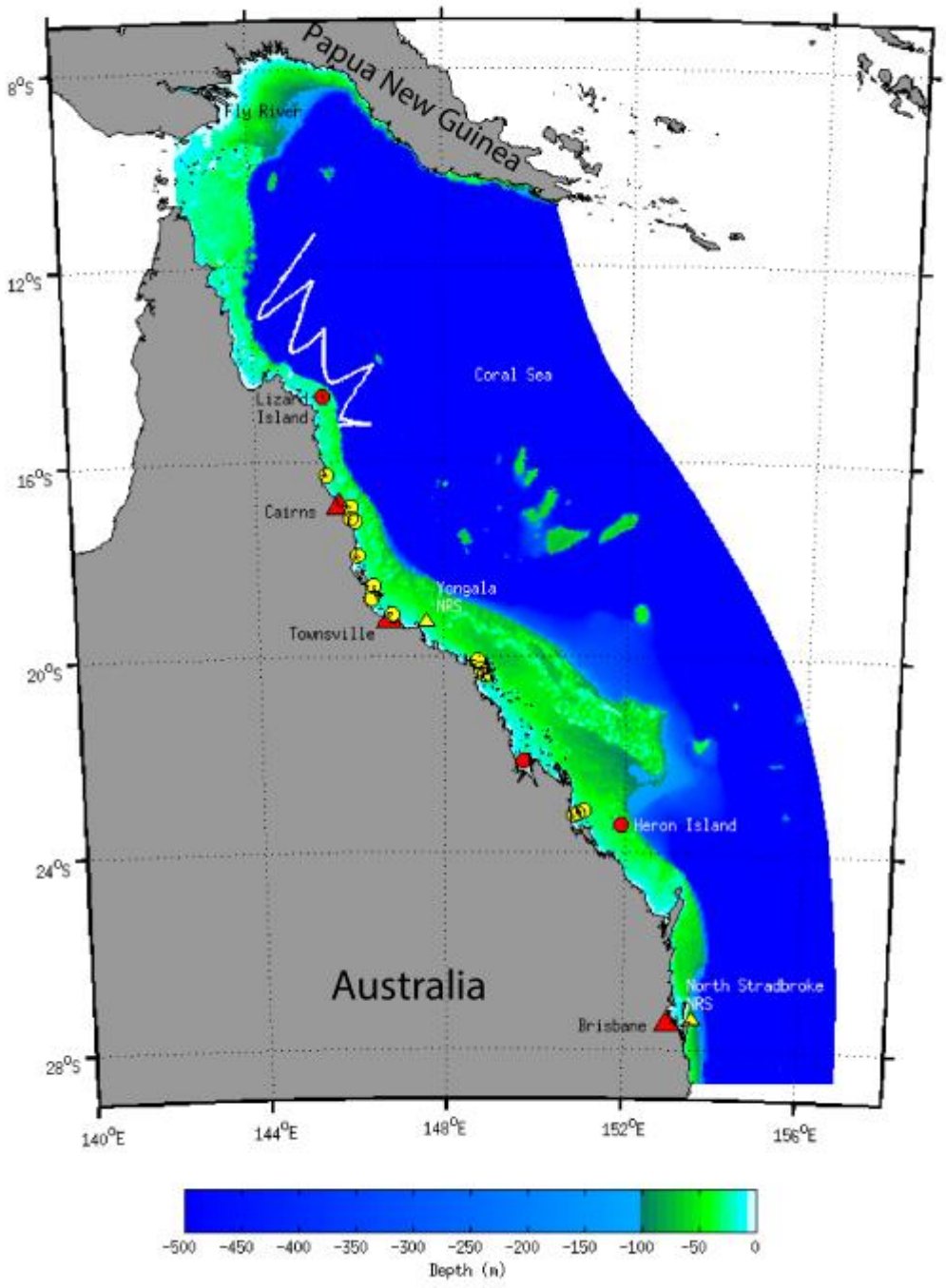

Figure 2: A map of the Great Barrier Reef region, with the color bar denoting the water depth, markers denote the population centres (red triangles), IMOS NRS sites (yellow triangles), GBRMPA MMP Water Quality Meters (WQMs; yellow circles) and points of interest referred to in the text (red circles), with the glider track (white line adjacent to Lizard Island). The *in situ* sampling locations and glider observations are used to assess the data assimilation system performance.

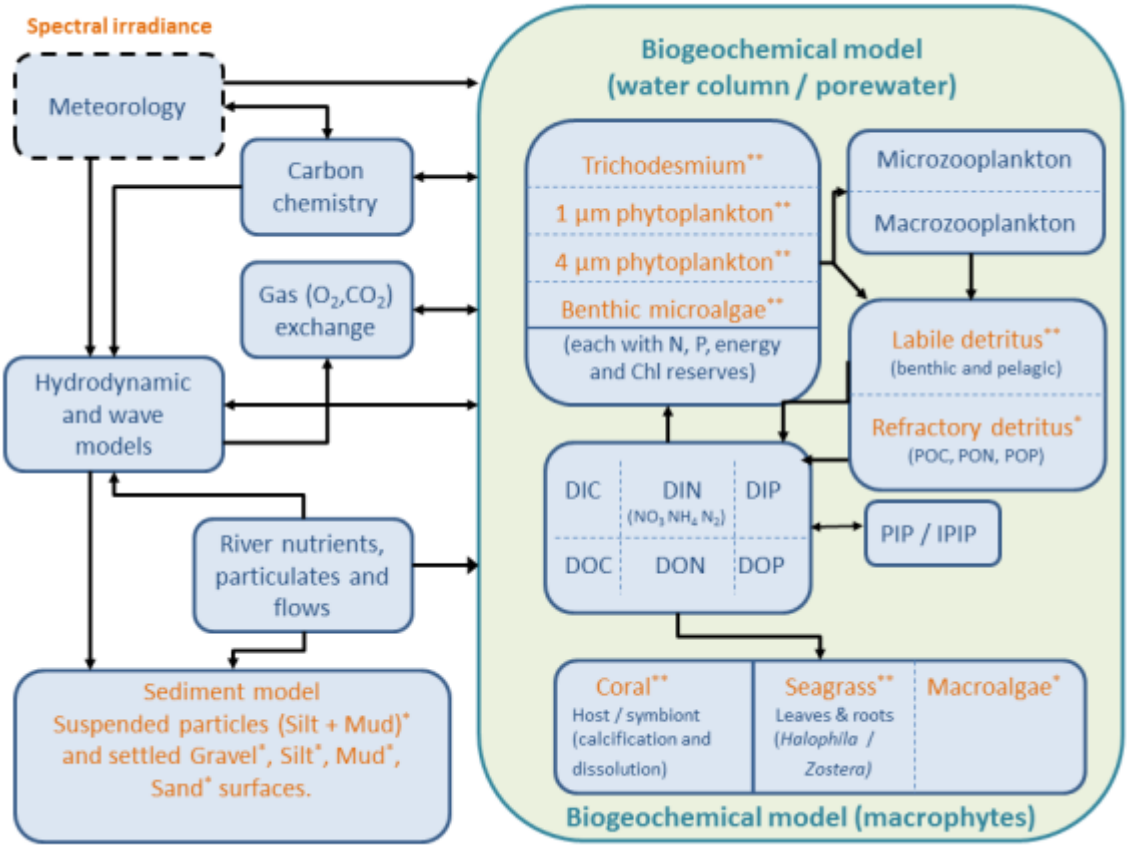

Figure 3: The eReefs modelling system with optically-active components identified with beige colouring and an asterisk, with the number of asterisks denoting the number of different optically-active elements with this component. Thus, each of the four microalgae have two pigment types, one absorbing like divinyl chlorophyll-a, and the other like photosynthetic carotenoids; there are two seagrass types, corals have both skeletons and zooxanthellae; three types of detritus absorb and scatter, and the sediment model contains a suspended fraction and four (mixed) sediment compositions. Additionally, pure seawater both absorbs and scatters light.

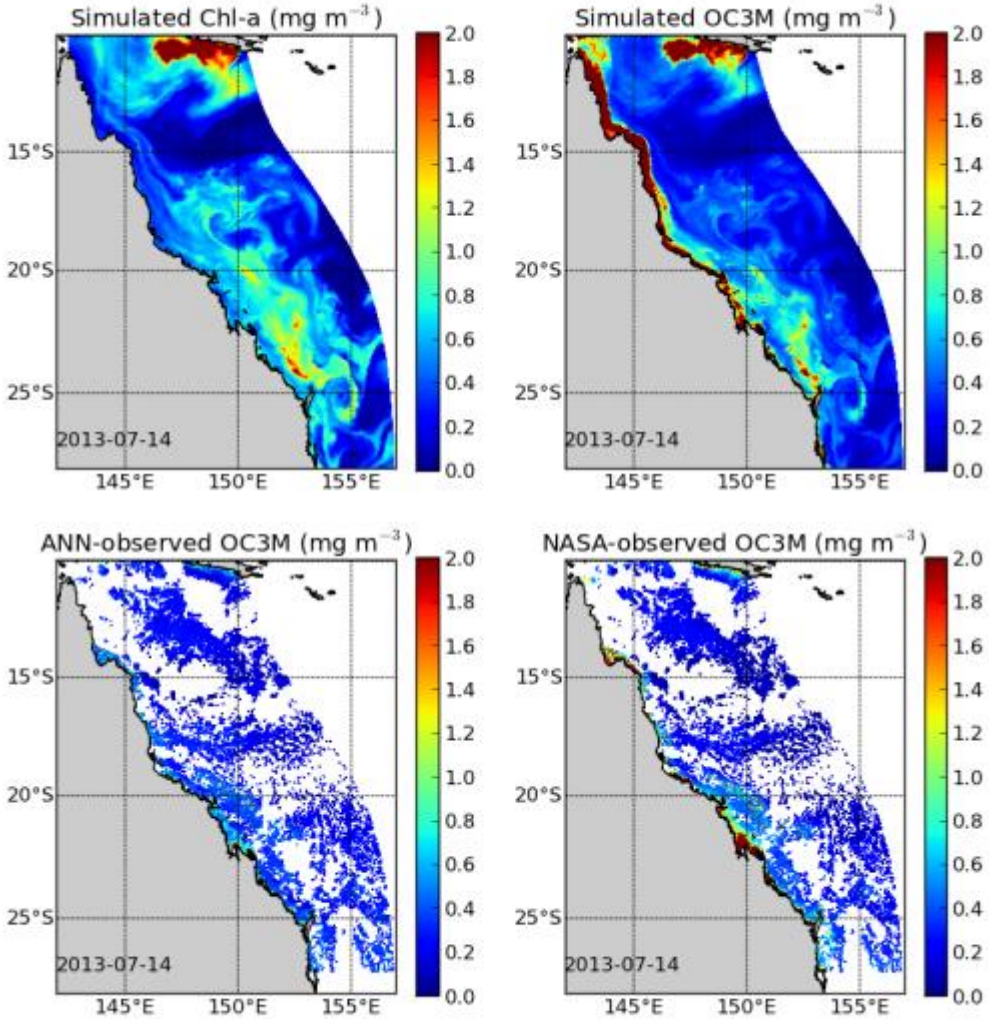

Figure 4: Simulated surface Chl-a (mg m^{-3}) of the non-assimilating control run (top left) and the simulated OC3M (top right) derived from the simulated remote-sensing reflectance for the 14/7/2013. The observed OC3M with ANN-derived observed remote-sensing reflectance (bottom left) and NASA-derived observed remote-sensing reflectance (bottom right).

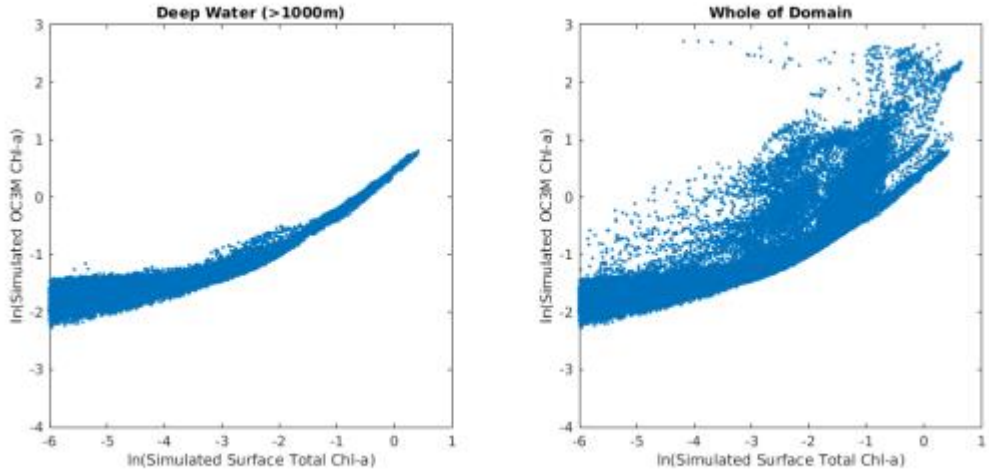

Figure 5: A log-scaled scatter plot of simulated surface total Chl-a and simulated OC3M for deep water regions (left) and whole of domain (right).

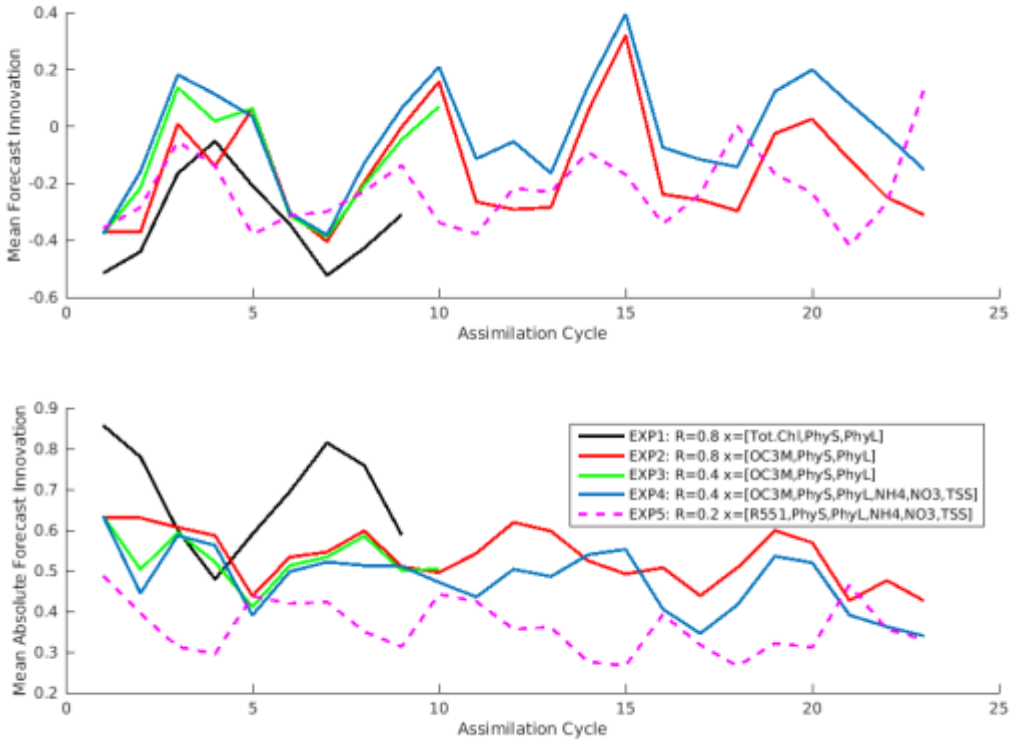

Figure 6: Comparison of mean innovation statistics for EXP 1-5. EXP 5 should be analysed with caution as these innovations relate to R551, not OC3M. An innovation of 0 indicates perfect agreement between model and observations. The top panel plots the mean innovation for each assimilation cycle. The lower panel plots the mean absolute innovation against assimilation cycle. The colours correspond to: black (EXP1), red (EXP2), green (EXP3), blue (EXP4) and magenta (EXP5). The variables in the legend correspond to the observation error (R) and the assimilation state vector (X).

.

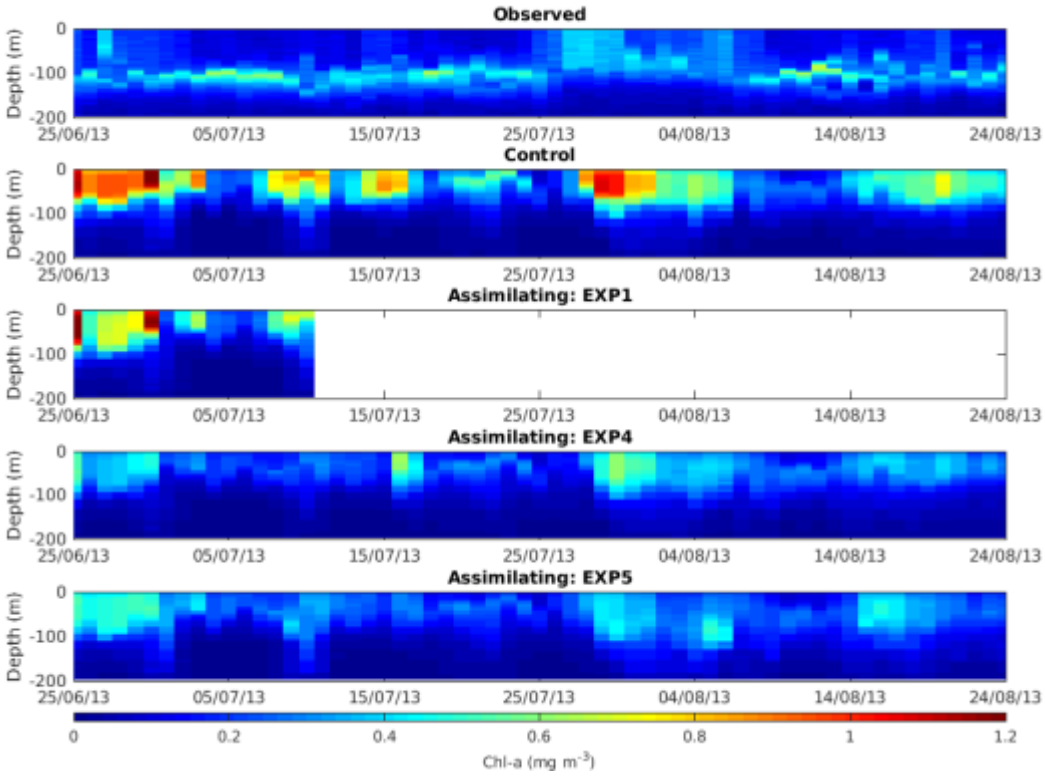

Figure 7: Comparison against independent IMOS glider observations, with the observed section of Chl-a derived from the onboard fluorescence sensor. The comparison is undertaken in model space, whereby all glider data that fall within a 1 hour window either side of when there is a 3D model output available is extracted, and equivalent simulated Chl-a are extracted from the model. The glider and model data is then spatially aggregated and interpolated onto the vertical model grid. The resulting section for the glider (top), and the simulated glider sections from the non-assimilating model (control) and assimilating model (EXP1, EXP4 and EXP5).

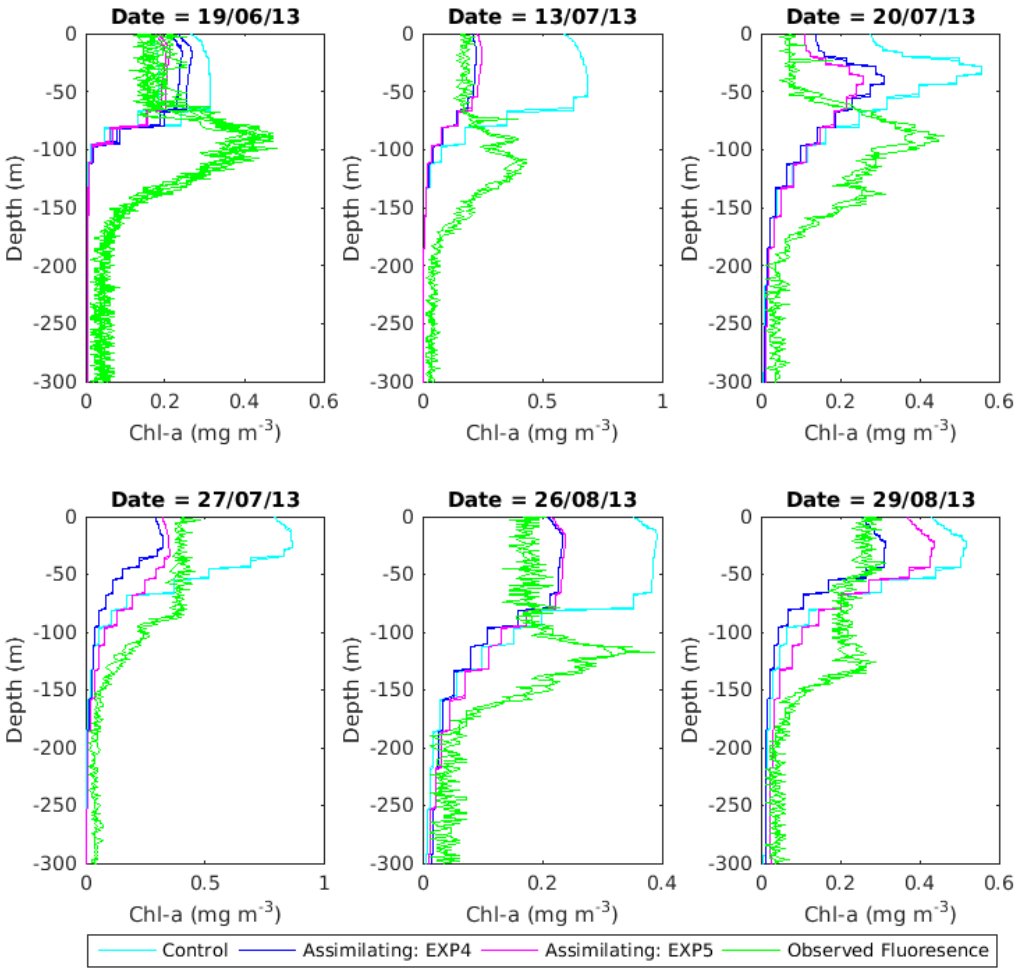

Figure 8: Comparison of depth-resolved Chl-a against six individual glider profiles using the method described in the caption of Figure 7. Glider observations are green, the non-assimilating control run profiles are cyan, and the assimilating run profiles for EXP4 are blue and EXP5 are magenta.

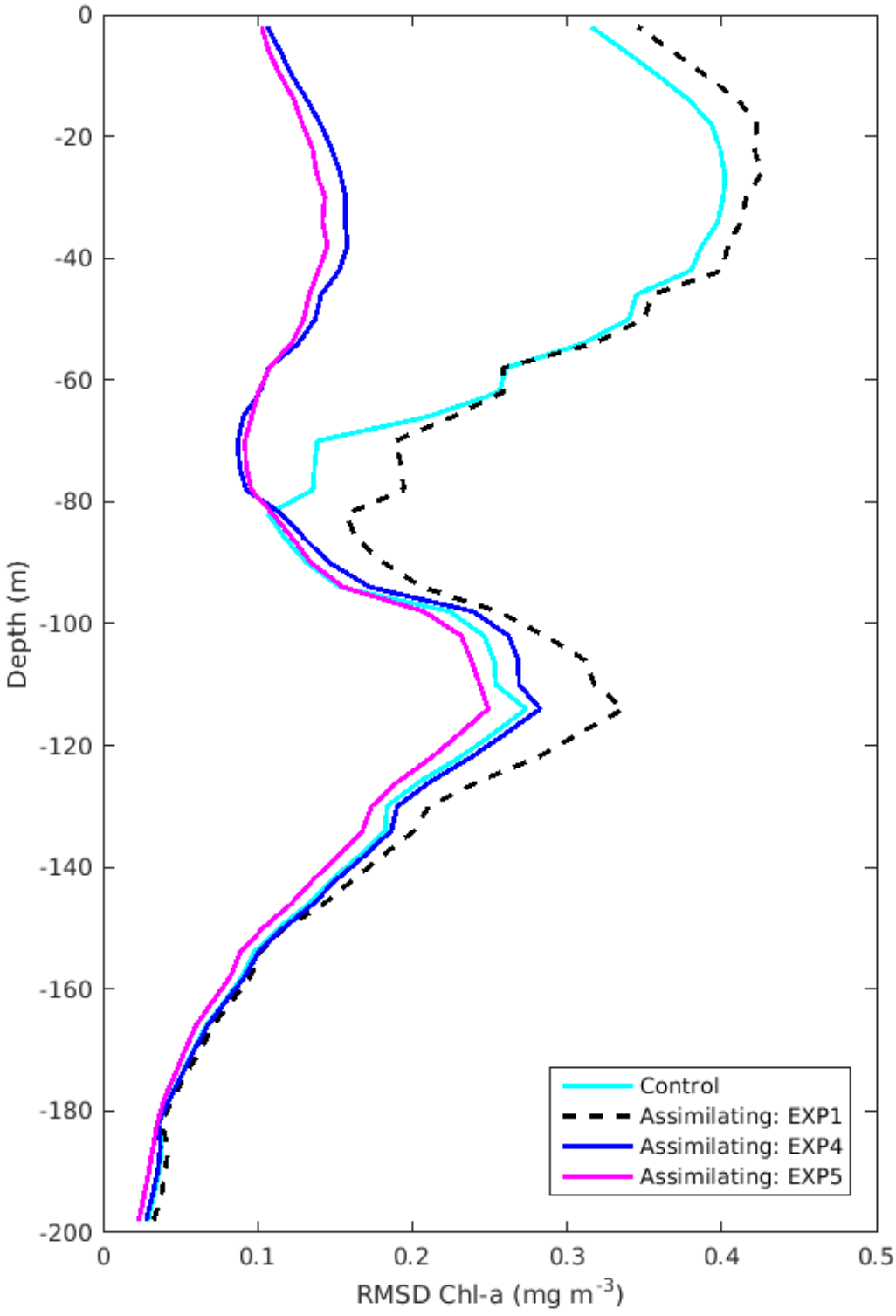

Figure 9: A profile of the temporal mean Chl-a RMSD between the glider observations presented in Figure 7 and the non-assimilating control (cyan) EXP1 (black dashed, note that the mean RMSD is calculated using a short time period), EXP4 (blue) and EXP5 (magenta).

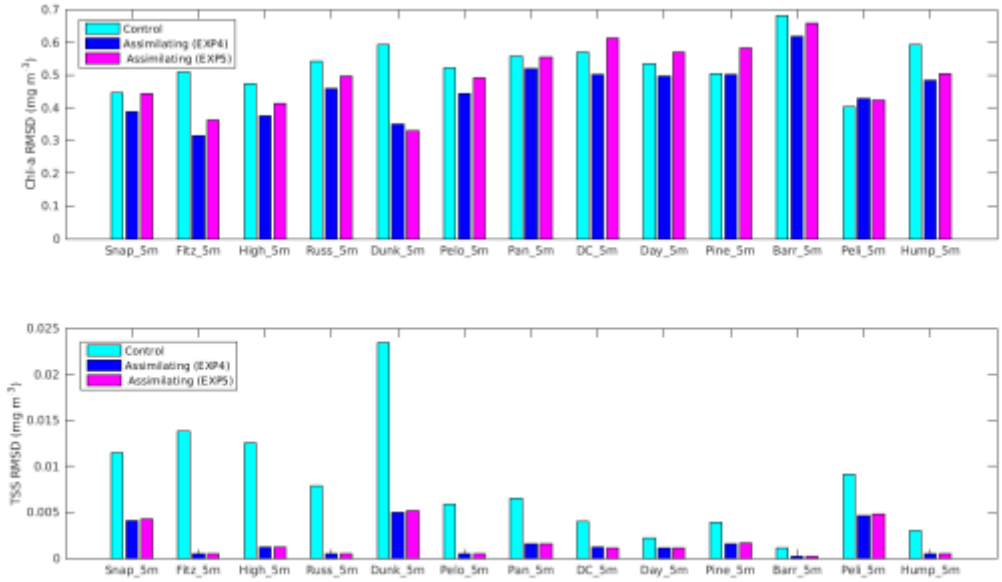

Figure 10: A comparison of Chl-a and TSS RMSDs between the in-situ GBRMPA MMP moorings for the non-assimilating (cyan) and assimilating runs of EXP4 (blue) and EXP5 (magenta), the GBRMPA MMP sites are denoted by the yellow circles in Figure 1.

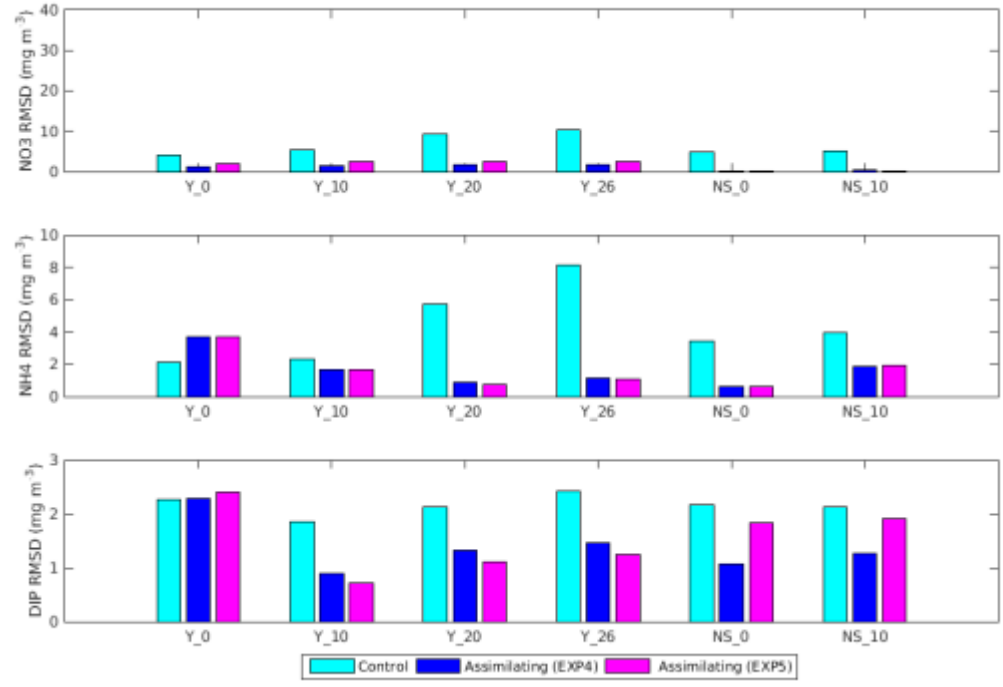

Figure 11: A comparison of RMSD of simulated nutrients with in-situ bottle samples for the non-assimilating (cyan) and assimilating run of EXP4 (blue) and EXP5 (magenta). Observations are obtained at the Queensland IMOS site (yellow triangles in Figure 1) at Yongala (Y) and North Stradbroke (NS) Island. Y_0 are the Yongala surface samples, while Y_26 are the samples taken from 26 m depth.

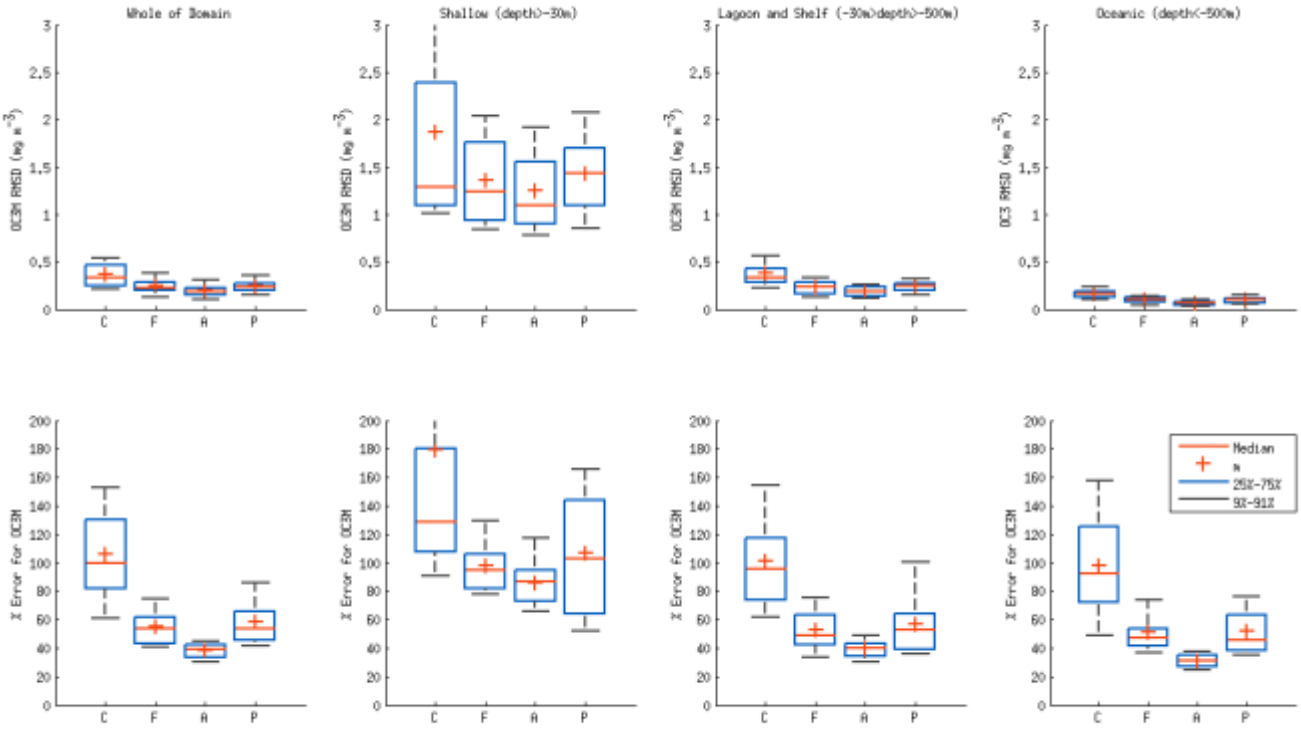

Figure 12: Box and whisker plots of RMSD (top row) and MAPE (bottom row) of the mis-match between simulated OC3M and ANN-observed OC3M. Each panel contains the control run (C) and EXP4 showing forecast (F), persistence (P) and analysis (A). Presented are statistics for the whole domain (left column), and regions as defined by bottom depth depth (three rightmost columns), for which the mean and range of values are presented in Tables 2 and 3.

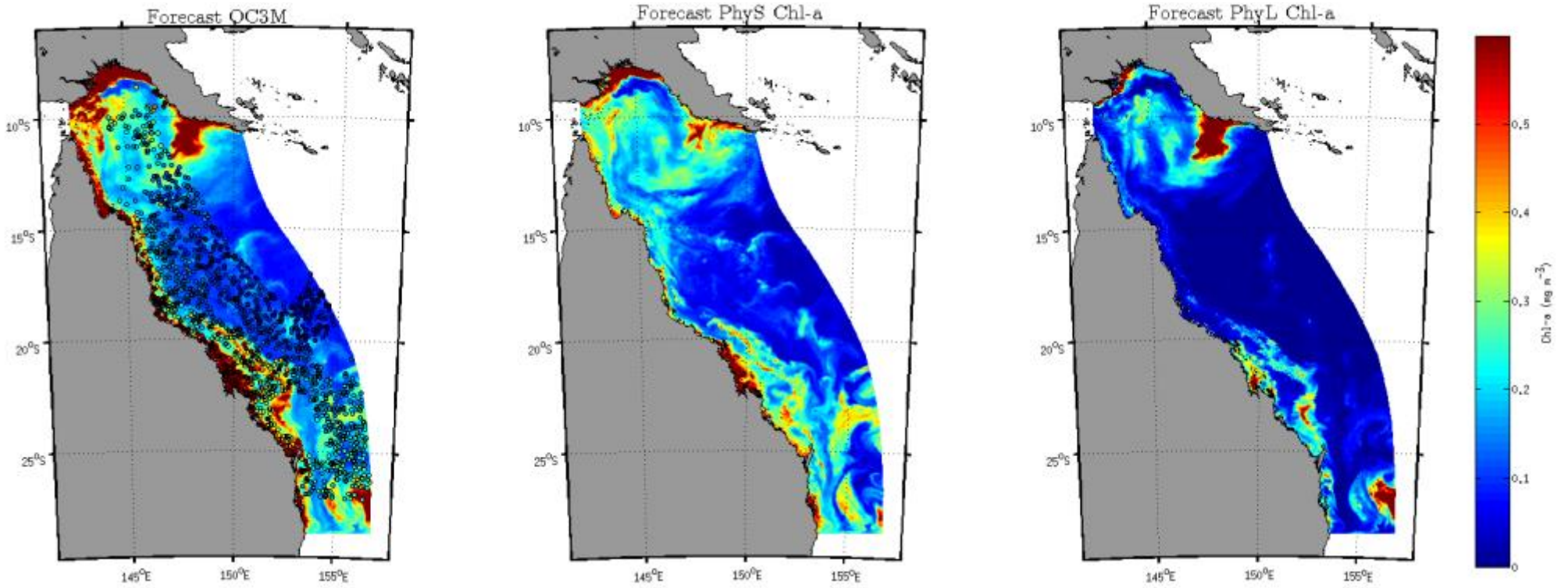

Figure 13: Simulated (forecast) OC3M (left) for cycle 22 (12th September 2013) with observatns overlaid, surface Small Phytoplankton (PhyS) Chl-a (centre) and surface Large Phytoplankton (PhyL) Chl-a (right).

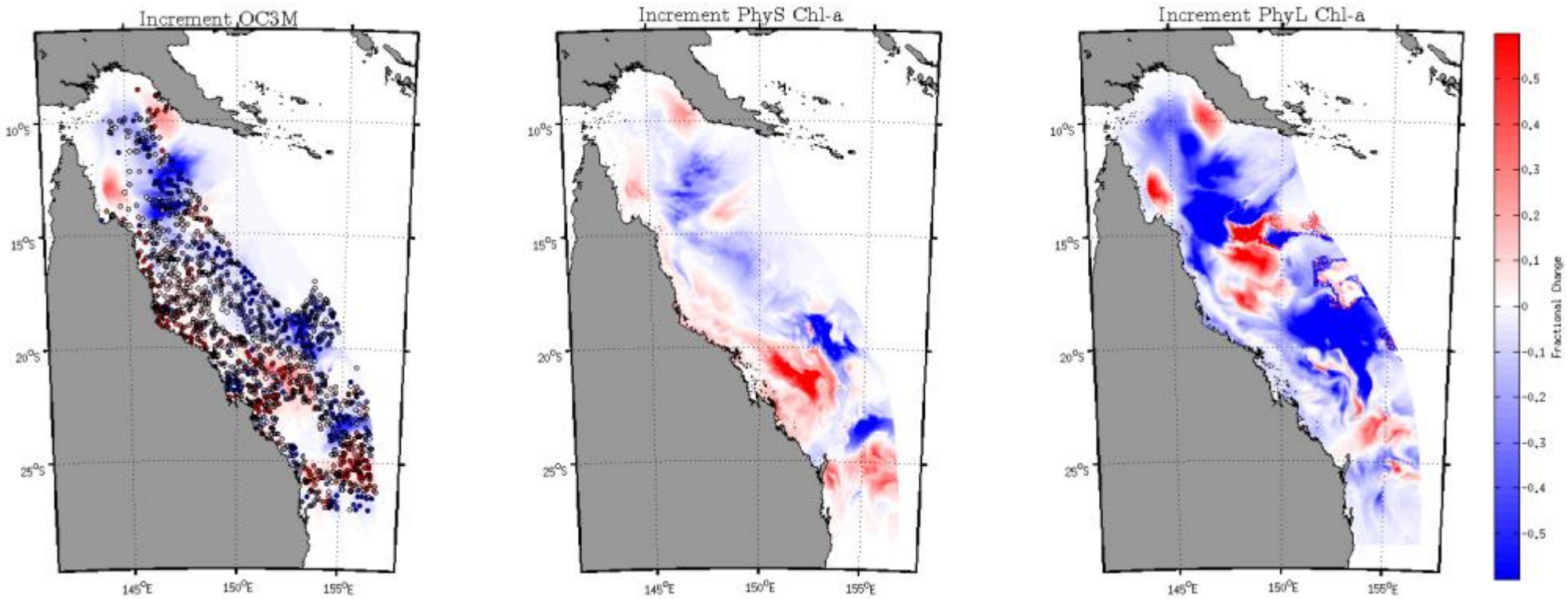

Figure 14: Increments that are added to the forecast fields generated by the assimilation system for simulated OC3M with innovations overlaid (left), and the prognostic variables of surface small phytoplankton Chl-a (centre) and surface large phytoplankton Chl-a (right).

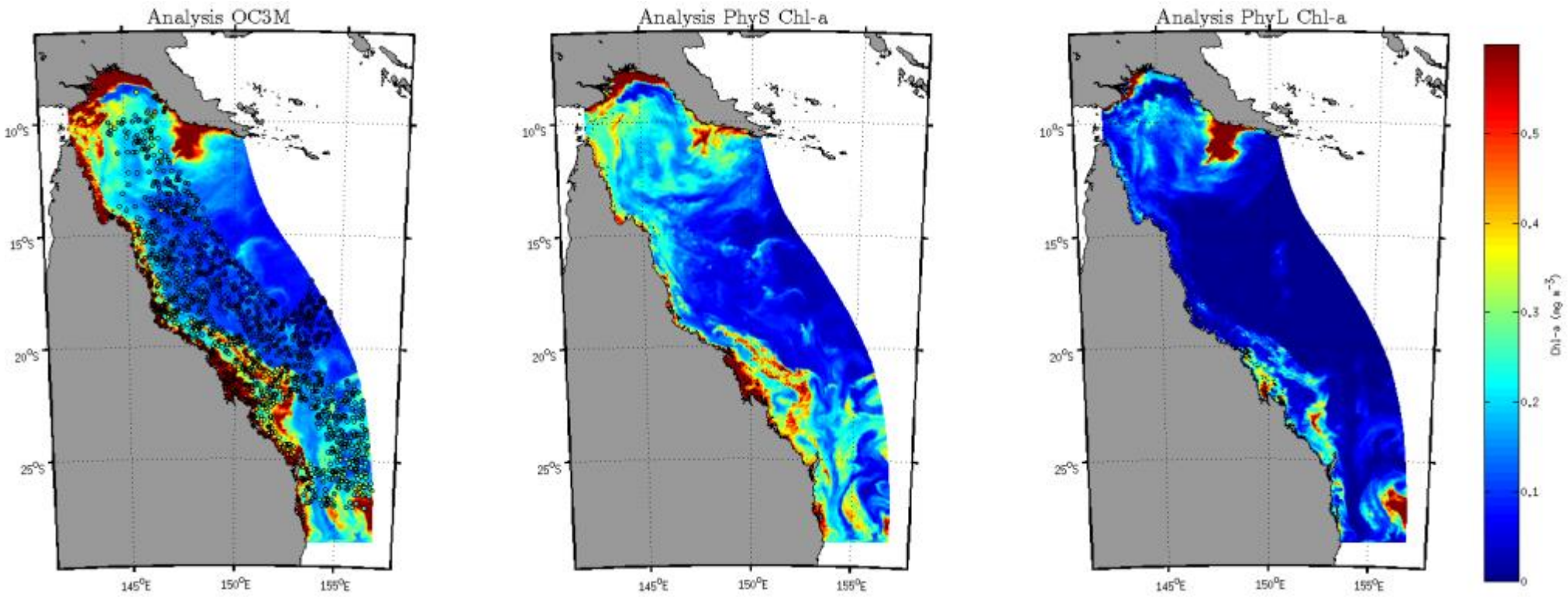

Figure 15: The resulting analysis fields for simulated OC3M and withheld ANN OC3M observations overlaid (left), and the analysis fields for the prognostic variables of small phytoplankton (centre) and large phytoplankton (right) for the 12th September, 2013.

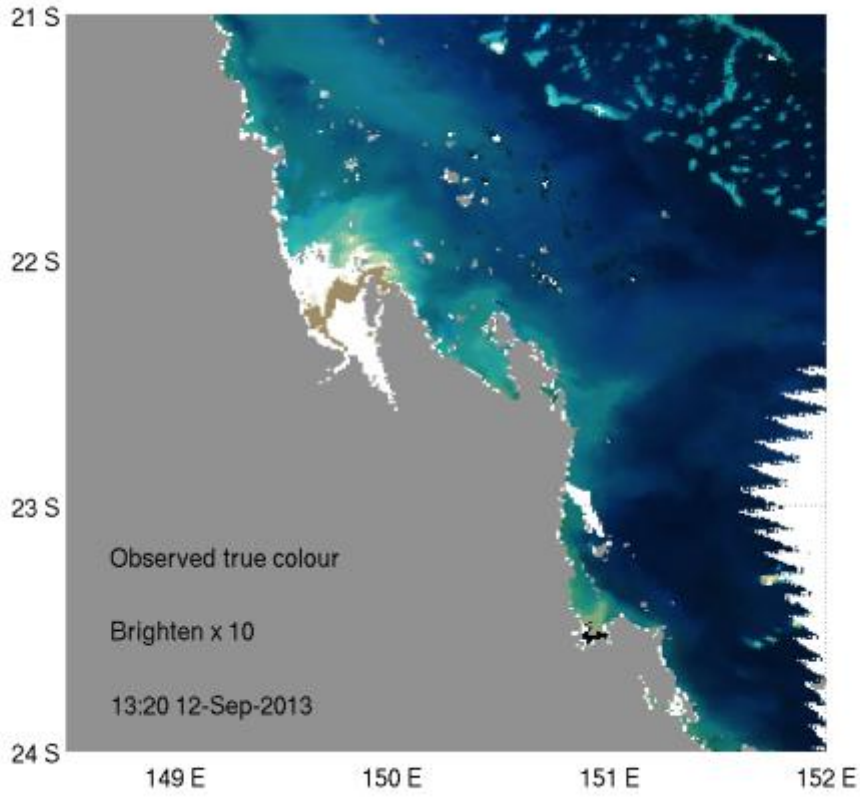

Figure 16: Observed true-colour image on the 12 Sep 2013 obtained from 1 km resolution, atmospherically-corrected ANN remote-sensing reflectance. The RGB wavelengths used were 667, 551 and 488 nm and processed using the MODIS true colour algorithm (Gumley et et al., 2010; Baird et al., 2016). The white pixels are clouds, grey is land.

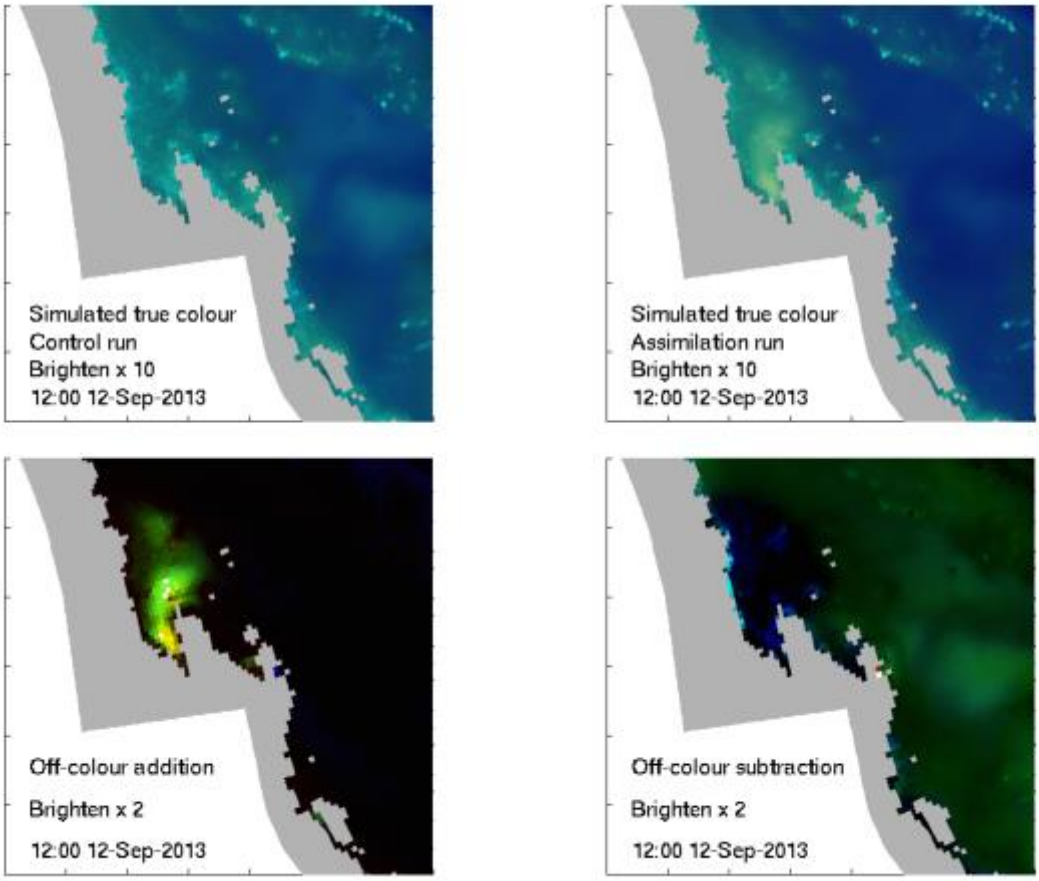

Figure 17: The simulated true-colour image on the 12 Sep 2013 of the control run (top left) and assimilation run (top right). The difference between the remote-sensing reflectance in the control and assimilated runs was used to quantify the colour (referred to as off-colour) added (i.e. greater surface expression, bottom left) and subtracted (i.e. less surface expression, bottom right) due to the updating of optically-active constituents in the assimilation run (see Fig. 13 for more details). Note that the off-colour images have a smaller brightening factor as the MODIS true colour stretch saturates the features that are of most interest. Simulated true colour images are not falsely-coloured, thus do not require a colour map, nor are they 2D as they have a depth of field, being based on reflectance from multiple depths and the bottom (Baird et al., 2016a). Thus simulated true colour can be considered a photograph of the optical state of the different model runs, and, like observed true colour, a powerful and intuitive visualisation tool for water clarity in biogeochemical models.