We thank the reviewers for their helpful comments and suggestions. We have taken these into careful consideration and changed the manuscript accordingly, which has improved it in content, clarity and presentation. Our responses to the comments are outlined below:

**Short Comment 1 (SC1):**

As outlined on [https://www.geoscientific-model-development.net/about/manuscript_types.html](https://www.geoscientific-model-development.net/about/manuscript_types.html) program code and data need to be made available in an open and persistent way. If this is not possible, e.g. due to copyrights or license issues, reasons need to be stated code availability section. Contacting one of the authors is not seen as persistent form of availability. You may consider to use the DOI service of your university and reference the DOI in your paper.

Thank you for highlighting this. We had forgotten to provide the links to access the model code (for both the physical and biogeochemical components). We have changed the code availability section accordingly. The most current official versions of the code to run the model are made available at: http://www.myroms.org.

**Referee Comment 1 (RC1):**

**General comment**
The authors presented a new high-resolution biogeochemical (BGC) model for the East Australian Current (EAC) system. The challenge of this work relies on providing a tool able to:
1. explore the BGC dynamics in the selected region
2. understand the EAC system dynamics as a whole
To address these objectives, the authors coupled ROMS and bio_Fennel, obtaining a model able to explore the complex BGC dynamics of the selected area at a regional and finer scale. The simulated surface chlorophyll-a dynamics were compared with a 10-years dataset of remotely sensed chlorophyll-a product observations (i.e., Copernicus-GlobColour). To assess the model performance several statistical metrics were used. Furthermore, the simulated vertical distribution of the nitrate was assessed against the CARS dataset. The high-resolution model presented here represents a powerful tool to explore the impacts of oceanic features and associated biological responses in the off shore East Australian waters. As stated by the authors in the manuscript text, this would not be possible simply analysing climatological fields by their own. Overall, aims and results of the work are well presented, as well as the different statistical analyses used to assess the simulations. In my opinion, only few sections require clarifications, as detailed below.

**Specific comments:**

*Page 2, line 14:* I do not think there is a need to start a new paragraph here.
*Page 2, line 27:* Same as above, I believe the topic is still the EAC.
We agree, and both have been modified accordingly.

*Page 3, line 6:* Insert the Internet link for CARS
We refer to this in the Code and Data availability section (http://www.marine.csiro.au/~dunn/cars2009/).

*Page 3, line 11-13:* here would be useful to insert the phytoplankton response to these physical factors in both cyclonic and anticyclonic eddies
Added the information that cyclonic/anticyclonic eddies are usually associated with high/low chlorophyll concentrations (page 3, line 15-20 of the revised version):
“Their contrasting (cycloic/anticyclonic) dynamical regimes create different biogeochemical environments: cycloic eddies present low sea level anomalies, doming isopycnals and a shoaling nutricline, while anticyclonic eddies are associated with high sea level anomalies, isopycnal depression and a deepening nutricline (McGillicuddy, 1998). Cycloic eddies are usually associated with elevated chlorophyll, while anticyclones present chlorophyll suppression (Everett et al., 2012, Gaube et al., 2014). Eddies close to the shelf may entrain biomass-rich shelf waters which are then transported offshore (Tranter et al., 1986, Everett et al., 2015, Macdonald et al. 2016).”

Page 5, line 25: the link of GlobColour would be helpful for the reader. This is provided in the Code and Data availability section: http://marine.copernicus.eu/services-portfolio/access-to-products/; ID: OCEANCOLOUR_GLO_CHL_L4_REP_OBSERVATIONS_009_082.

Section 2.4: moving the description of the quantitative metrics below every equation can help to better understand the analyses performed and how to read the different panels in Fig. 5 and 6. This is a great suggestion for better readability, thank you. The section 2.4 has been modified accordingly.

Page 8, line 21: EAC nutrient-poor water affect the phytoplankton growth and, as a consequence, the chlorophyll fields. Modified from “Nutrient-poor EAC water impacts the chlorophyll fields as visible by (…)” to “Nutrient-poor EAC water impacts phytoplankton growth, and associated chlorophyll fields, as visible by(…).”

Page 9, line 8-9: I think the sentence should be reworded. Indeed, rather than ‘aggravate the model misfit to observations’, the remote sensing biases described earlier can explain the satellite observations vs simulations inconsistencies. Rephrased to “these remote sensing biases can partly explain the inconsistencies between simulated and observed chlorophyll” (page 9, line 25 of revised version).

Page 9, line 19-20: I do not understand the meaning of the sentence. The inconsistencies do not derive uniquely from the physic processes, to which corresponds biological responses?
Both parts (physics and biology) play their role in creating these inconsistencies. Partly because the physical structures are not in the right place at the right time (as it is a free running model), but also because the biological model is an oversimplification of reality, not capturing its complexity, and so it is unable to fully reproduce the observed variability. We have edited the text as follows to clarify this point: “Inconsistencies at high frequencies likely derive from a combination of physical and biological processes. In a free run such as this, dynamical features like the EAC physical position, mesoscale eddies and small-scale fronts, are generally offset in space and/or time from those in nature. In addition, the modelled N₂PZD₂ system, with only one phytoplankton and one zooplankton compartment, is an oversimplification of reality. By. not capturing the complexity of the natural biological community, it is unable to fully reproduce the observed variability” (page 10, lines 4-9 of revised version).

Page 10, line 5-26: the authors here accurately describe the features of figure 7. Therefore, would be helpful to insert mode 1, 2, etc… on the top of Fig. 7, to help the reader to quickly refer to the panels while reading the text.
Thank you, Fig. 7 has been modified accordingly.

Figure 1a: It would be useful to shows the EAC and the separation zone and possibly the formation/occurrence of CE and ACE eddies off East Australia. At least a schematic image
would be important as I think these information are more relevant for this study rather than the depth.

We have modified Fig.1a to represent the model domain through an 8-day average of the simulated surface chlorophyll (from mid-October 2008), where we highlight the EAC position, one cyclonic eddy and one anticyclonic eddy. We highly appreciate the suggestion as it led to a much more relevant illustration. Thank you.

**Figure 8:** I imagine the top row is from ROMS and bottom row from CARS. Please double check that, as the figure and caption are not consistent.

Yes, that was a mistake, thank you for detecting it. Corrected.

**Referee Comment 2 (RC2):**

**Summary:**

I enjoyed reading this paper and it has the potential of adding to our understanding of BGC dynamics in the EAC separation region. My major criticism at this point relates to the interpretation of the results. I’d like to encourage the authors to consider adding additional interpretation and analysis on a number of fronts that are suggested below. Whilst the comments may appear critical, I think they would strengthen the study.

**General Comments:**

The initial slope of the P-I curve and half-saturation coefficients have been tuned to recreate the observed Chl-a concentrations. In section 3.1 the model is assessed against observed Chl-a using modelled surface concentrations of Chl-a. The Remotely sensed Chl-a could be considered a some form of “depth weighted” averaged concentration over the optical depth (which can be quite deep in this region). Therefore by taking into account “difference in kind” error between modelled Chl-a and the CMEMS GlobColor Chl-a products, combined with the comparison of a modelled surface Chl-a being compared with a “depth weighted” averaged, there is scope for the “tuning” to be biased. Would it not be better to average the modelled Chl-a over an optical depth?

This is a great point and one that we had considered. We could have attempted to identify an optical depth based on what the satellite “samples”, by using the diffuse attenuation coefficient at 490 nm \((1/K_d490)\), for instance, and then integrating the Chl-a variable over such depth. It is also possible that doing the inverse - converting the 3D Chl-a variable into a 2D field - would be a better approach. The 2D ocean colour product could be determined from the 3D field using (Gordon and Clark, 1980):

\[
C_{2D} = \frac{\int_0^z C(z) f(z) dz}{\int_0^z f(z) dz}, \quad f(z) = \exp \left( -\int_0^z 2K_d dz \right)
\]

where \(f(z)\) is the exponential weighting function that accounts for the arriving irradiance having been attenuated differentially at each depth and returned to the surface by the same factor. This approach has been used to calculate remote-sensing reflectance from depth-resolved model inherent optical properties (IOP) fields (Baird et al., 2016) and depth-resolved chlorophyll fields (Moline and Prezelin, 2000). However, such approach is also not without limitations - particularly in assuming that the downwelling and upwelling path-lengths are the same.

It is also worth noting that our model covers quite a large area, occupied with different water masses defined by widely contrasting optical properties, which makes the definition of an optical depth far from trivial. Taking all of this into account, we have decided to use the simplest and most common approach (eg. Matear et al., 2013, Cetina-Heredia et al, 2017),
which is to assume that the surface model field is equivalent to the ocean colour product and allows for a direct comparison of the two. Thus, for these reasons our results remain unchanged.

A majority of the results and discussion focus on Chl-a and Nitrate, yet there are 4 other non-observed state variables that influence the dynamics. What do these distributions look like? Are they sensible? Do they qualitatively behave as one would expect? We have focussed on Chl-a and Nitrate as our interest is in realistically simulating the region’s phytoplankton variability. We see the other state variables acting as “closure terms” in helping us reach that goal. However, we qualitatively verified them throughout the model calibration stage. Please refer to the surface means of the remaining state variables, as well and their domain-averaged surface time series, bellow:

Figure S1: Daily surface fields of: a) Zooplankton; b) Ammonium; c) Small detritus, d) Large detritus, averaged over the full study period.
Apart from not being our focus, we have opted to not include the analysis of these variables in the manuscript because they lack an observational dataset that allows a quantitative validation. There is no equivalent to the expansive coverage of ocean colour datasets, and climatologies such as CARS do not extend to these variables. Discussing them would inevitably be subject to a high degree of speculation.

There is little discussion of the interaction of physical processes with BGC? For example, the vertical supply of nutrients into the photic zone. What is the typical flushing time of water in the mixed layer?

We agree that this topic is very interesting, and we have already started investigating the regional nutrient replenishment mechanisms into the euphotic layer. However, we plan to publish these results separately as a standalone piece of work as we believe that exploring such dynamics here would steer the focus away from what we are trying to accomplish: to evaluate the model. Moreover, it is potentially outside the scope of this journal.

In areas of the domain where the flushing time is short (through horizontal advection), the BGC dynamics will be dominated by the prescribed boundary conditions. Whereas in areas where the flushing time is comparatively long, BGC dynamics will be dominated by internal model processes. Such an analysis would help explain the discrepancies in PCA mode 1 as mentioned below.

This would hold true for models encompassing smaller regions or with much lower resolution. However, our model spans more than 15° in latitude and extends to almost 1000 km offshore - with 317 grid cells across latitude and 272 across longitude. Even with the significant flux through the northern boundary created by the East Australian Current, the BGC dynamics in most of our domain are generally dominated by internal model processes. Calculating the flushing time (inside the MLD, for instance) would be very interesting and is something that we will explore in the next study on the regional nutrient replenishment mechanisms. Thank you for the suggestion.

The colorbars on many of the figures are such that it is really hard to look quantitatively at the results. It is really difficult to pick discernible differences in color between 0.2 and 0.5 mg Chla m-3. More attention needs to be given to the colormaps used to generate the figures. The addition of a shelf contour to the plots will allow the reader to discriminate the deep ocean, from the shelf and shelf-break.

Thank you for highlighting this point. We have revised all our figures accordingly in order to make the values more distinguishable (decreased the smoothness of the colormap and modified colorbar min. and max. where needed). We have also increased the thickness of the shelf-break line (200m isobath) to increase its visibility. Figures 3, 5 and 6 are updated.
As it stands the paper is descriptive of observed phenomena, but the power of a model is that it allows you to explore unobservable quantities. There is little if any discussion about the dynamics of the unobserved state variables nor derived quantities like primary production etc.

If the model is to be used to quantify and interpret the 3D time evolving state of the EAC, then the authors must assess the model in a way that presents evidence to the reader that the model is fit for purpose. Broad statistics are used to show that there is reasonably good correlation between the model and remotely sense observations, but in many cases a detailed interpretation of the results is not presented.

We agree with this point, however the focus of this paper is to evaluate the model, as per a GMD model evaluation paper. Here we show that the model depicts reality with sufficient skill that it can be then used to explore those unobserved quantities and to study their underlying dynamics. What we set ourselves to do here is to validate the model’s ability to reproduce the overall patterns of variability of the most observed variable (Chlorophyll) as stated in the introduction. This provides the foundation needed for further exploration of the dynamics that underlie such variability.

Furthermore, the model has only been assessed against remote sensing and in-situ climatology (nitrate), there exists a rich set of BGC observations from gliders and research cruises for the area. I would strongly encourage the authors to undertake an assessment of the BGC model against in-situ data. Why not try a comparison of the model fields against Schaeffer et al., (2016). Whilst it is close to the shelf break, it may assist with providing an additional in-situ dataset for which to assess the model against.

We had considered this. However, we have decided that it wouldn’t be a sensible approach at this stage because a free-running model at the regional scale is not supposed to be able to reproduce the type of variability (high frequency and very local) associated to these observations. Such an assessment may still have its place, but at a later iteration, on a nested model with higher focus on shelf dynamics. Such an analysis on the current configuration would take the focus away from the regional variability patterns the model was configured to reproduce.

We appreciate the reviewer’s thoughtful suggestions above, all of which are good ideas for further research. However, we feel that they are mostly beyond the scope of this initial GMD model evaluation paper, which is focussed on describing the application of the model to the EAC System - hence creating the first high-resolution BGC model of the region, and not on drawing substantial conclusions about its dynamics.

Specific Comments
Page 1, Lines 24-30: There are varying complexities of BGC models ranging from highly parameterised through to extremely complex. The parameter identifiability problem associated with additional complexity is discussed in Friedrichs et al., (2007) with further suggestions on how to adequately represent uncertainty in Parslow et al., (2013). As for using Chl-a as a variable to assess model skill, Baird et al., (2016) show that observed OC3M Chl-a from satellites can at times be very different to simulated Chl-a from a model. This is confirmed in Jones et al., (2016). These “difference in kind errors” are important in the interpretation of the results later in the manuscript.

Thank you for that. We have added this additional information and relevant references by changing the Introduction slightly. The text now reads:

“Uncertainty intrinsic to BGC models does not derive from parameter estimation alone but also extends to the choice of equations used to describe the targeted ecosystem (Franks, 2009). These models generally aggregate plankton populations into broadly defined trophic compartments and track the flow of a chemical element, such as nitrogen or carbon, among these compartments. Variations in this model structure, i.e. model complexity, include the use of additional limiting nutrients (such as silicate, phosphate, and iron), the division of the
planktonic groups into multiple functional types or size classes, and inclusion of additional state variables such as bacteria or detritus. Model complexity is usually decided based on the targeted ecosystem and/or ecosystem function and has been one of the core topics in BGC modelling; discussed in Anderson (2005) and thoroughly explored in Friedrichs et al. (2007), for example. The selection of a particular model complexity and set of parameters is usually made by evaluating how well these are able to reproduce the available observations of state variables; due to its continuous acquisition and spatial coverage, remotely sensed chlorophyll is the most abundant BGC data set for marine ecosystem model evaluation. However, the use of this data set is not without uncertainty of its own. Remotely sensed chlorophyll is not always directly comparable to the simulated chlorophyll fields, as shown in Baird et al., (2016) and confirmed in Jones et al., (2016). We note that we have also added Anderson (2005) to complement Friedrichs et al., (2007) discussion on model complexity.

Page 2, Lines 14 - 21: References needed. We agree and have added them to the text. It now reads: “Therefore, coupled model configurations that attempt to realistically resolve ocean features and their impact on phytoplankton provide invaluable insight for a diverse range of topics including fisheries (Blanchard et al., 2012), water quality and ecosystem health management (Rombouts et al., 2013), carbon sequestration (Blain et al., 2007), and climate change (Matear et al., 2013).”

Page 3, Lines 16 - 25: Can you comment as to the suitability of this N2PZD2 model for this particular area? There are other choices available, both more complex and simpler. Is a single P group suitable for this region? This is a good question, and the answer depends on the scale and focus. If the focus is on the general patterns of variability - we hope we have demonstrated that a single P group allows the model to capture these satisfactorily. However, a single Phytoplankton group is likely not enough for a more detailed approach, in which both offshore and on-shelf communities need to be adequately represented. In that case, we would suggest using our application of the model as a basis on which to build upon in trying to reach increased complexity. We have stated this on page 5, lines 15-21 (lines 24-30 of the revised version): “Shelf phytoplankton species and community structure are expected to be different from the phytoplankton community found offshore (Armbrecht et al., 2013). For this study, we chose to apply an established, relatively simple biogeochemical model with only one phytoplankton functional type. In part, this decision reflected the overall emphasis of the physical model on the EAC and vast offshore region; the model has limited ability to resolve critical physical dynamics on the shelf due to model and forcing resolution and omitted freshwater inflow. Our overall focus is on the larger scale BGC dynamics, their seasonal variability, and local impacts of offshore mesoscale processes. Future modelling efforts will address shelf processes and more complex biogeochemical interactions”.

Page 6, Lines 1-8: You mention here that the model is initialised with Nitrate from CARS. How are the other model variables initialised, especially those that are unobserved? Thank you for asking, we meant to provide this information. The other variables are initialised with a seeding population of 0.01 mmol N m\(^{-3}\). We have added this to the sentence in section 2.2 as follows: “The model is initialised with seeding populations of 0.01 mmol N m\(^{-3}\) for all state variables except nitrate (NO\(_3\)), which is derived from the CSIRO Atlas of Regional Seas climatology (CARS, described in Sect. 2.3.2).”

Page 8, Line 12: What is the likely cause for the bias? It appears that the model is overestimating the Chl-a by a factor of 2 for substantial periods of time. This relates to my question posed above given that you are comparing a surface value with an observed value calculated over an optical depth. That is a good observation. The optical depth may indeed play a role in it, but so would a conjugation of different factors such as model structure and the loosely constrained parameters, which may not be “perfect” in describing the region’s phytoplankton community.
Moreover, the remote sensing inherent biases are possibly aggravating the inconsistencies between this data set and the model. We refer to these limitations throughout the text, such as in page 5, lines 15-21 and page 9, lines 5-9 (page 5, lines 24-30, and page 9, lines 28-34 in the revised version). The fact that the onset of the spring bloom is slightly shifted northward in the model also contributes to this bias, which is larger precisely in the northern area (as discussed throughout the text and illustrated by the differences in Fig 2.b, discrepancies in the patterns for months of August and September in Fig.4, larger values in Fig.5b, and, discrepancies in patterns in Mode 1 of Fig. 7).

Page 9, Line 3: If you use a 200m depth contour on the plots, it will help denote the region you are discussing.

Thank you. We have increased the visibility of the 200m isobath line in Figures 3, 5 and 6.

Page 9, Line 4: Is the model parameterised to simulate large or small phytoplankton?

This is a good question and one that is deceptively hard to answer due to the parameters being so loosely constrained in the literature. Based on Geider et al. (1997), for example, the parameters used are suitable to describe a community akin to microphytoplankton. This is in accordance to the results of the taxonomic investigation, under different oceanographic conditions, of Armbrecht et al. (2013). However this is highly speculative: the initial slope of the P-I curve and the half saturation concentrations for uptake of NO3 and NH4 are very poorly constrained for most species and the cited taxonomic study was developed on the shelf.

Figure 7: Suggest adding column titles to denote modes 1-4. Top row - y-axis needs explaining in caption

Thank you, we have added mode numbers and y-axis units/explanation to the caption.

Page 10, lines 21-26: This section is very light on the analysis and interpretation of the PCA analysis shown in Fig. 7. Whilst the correlation coefficient might be high, the are very obvious differences in the spatial structures of the model and obs. Interpretation is needed to explain these differences beyond just that relating to correlation. e.g. is the model over or under predicting the spring bloom, and in what areas? This may assist in determining why there are discrepancies in the northern section of the domain.

This is an important point and something that we have overlooked in our attempt to not expand the analysis to the underlying dynamics. So thank you for bringing our attention to it. We have added our interpretation on the difference in spatial structure of the first EOF mode. We attribute it to the differences in the patterns (slight northward displacement) of the onset of the spring bloom, with reference to Figure 4 where it is most easily discernible (months of August and September). The text now reads:

“Mode 1 captures the spring bloom, reaching its annual maximum around the beginning of October of each year. The discrepancies on the spatial structures of this mode are better interpreted with the help of Fig.4, specifically the anomalies for the months of August and September. As mentioned before, these show a northward displacement and slight overestimation of the onset of the spring bloom in the model, which leads to the same differences in the patterns captured by Mode 1”.

Page 11, lines 1-2: This transect lies so close to the eastern boundary of the model domain that there is a risk that what is being seen in Figure 8 is influenced by the climatological boundary conditions prescribed at the boundaries. What does a transect from the central domain look like?

The transect location was chosen in a way that it wouldn’t be under the influence of the EAC as we wanted to illustrate the “background” latitudinal gradient. Its graphical representation on Figure 1.a made it look like it was quite close to the eastern boundary but it is more than 40 grid cells (approx. 250 km) from it. This is a justified concern, however, and so we have shifted the transect to the middle of the domain. Fig.8 has been updated and this has not caused a substantial change in the patterns:
Figure 8: Seasonal nitrate transects (mmol N m\(^{-3}\)) along a transect through the middle of the domain, South to North (orange line in Fig. 1). Top row: CARS climatological values; Bottom row: model.

Figure 9 would benefit from an additional row showing the difference between the model and CARS, such a plot would assist in the interpretation of subtle differences including showing the differences in the supply of nitrate to the surface waters which is important for primary production.

Thank you for this suggestion. We have added an extra row with the difference to Figure 9 and a brief discussion on it to the text:

Figure 9: Monthly nitrate profiles (mmol N m\(^{-3}\)) at three different locations (A, B and C in Fig. 1). The solid line and the shaded area in the top and middle rows represent the mean value and first standard deviation, respectively. These were calculated over the study period (for model data, in blue) and climatology (for CARS, in red). Bottom row illustrates the difference (ROMS-CARS) for each average month (coloured) and average difference (thick black line).
“The bottom row of Fig.9 illustrates the difference in monthly averages (coloured lines) and average difference (black line) between model and observations. This is obtained by subtracting CARS nitrate concentrations to the simulated nitrate fields at the same three locations (A, B and C) as depicted above. It is worth noting that at these locations the model generally overestimates nitrate concentrations in the upper 800 m, except for the austral winter months (June, July and August) of the central profile, when it underestimates concentrations by approximately -1.5 mmol N m\(^{-3}\). The central and southernmost profiles are characterized by an inversion from average overestimation to average underestimation at around 1000 m, with the central profile showing the highest average difference of -2.2 mmol N m\(^{-3}\) at 1600 m. The northernmost profile presents the highest overestimation of the upper 250 m, due to a difference of approximately +2.5 mmol N m\(^{-3}\) during the spring months of September, October and November. This is of particular significance as an overestimation of the nitrate concentrations within the model’s euphotic layer, albeit low, may contribute to the slight northern shift of the simulated spring bloom pattern (observed in sections 3.1 and 3.3)”.

New references: