We thank the reviewer for their comments, and respond to each point below in blue text, and proposed altered text in italics.

The paper 'Modelling Ocean Colour Derived Chlorophyll-a' by Dutkiewicz et al. uses a modified version of the MIT general circulation model that incorporates scattering and absorption properties of water, detritus, coloured dissolved organic matter (CDOM) and 9 phytoplankton types. 'Actual' chlorophyll-a concentration (Chl-a) is determined by summing the variable chlorophyll-a concentration of the 9 phytoplankton types. Remote sensing spectral reflectance, determined from the resulting modelled upwelling and downwelling irradiances are used in a NASA OCx type algorithm to compute satellite-like 'derived' chlorophyll-a concentration which is compared to 'actual' chlorophyll-a concentration. Firstly, the model output is used to test assumptions used in the derivation of chlorophyll-a concentration using the standard OCx ratio algorithm. Secondly, bloom initiation timings, determined from the 'derived' and 'actual' chlorophyll concenrations, are compared. Lastly, the impact that other optically important parameters may have on 'derived' chlorophyll concentration is explored. The authors conclude that (a) applying a single set of coefficients in the OCx algorithm globally is not as accurate as applying regionally variable coefficients, (b) there is a temporal mismatch between the initiation of the spring bloom defined using 'actual' Chl-a compared with that de-fined using 'derived' Chl-a and (c) that this mismatch may be caused by the optical influence of other substances such as CDOM and/or detritus.

I found this paper to be generally well-thought out and well-written. For colleagues who are not regular users of these products, this paper provides important caveats for the use of satellite derived data. For those of us who work with ocean colour derived products more often, the results may not be unsurprising, but it serves as a timely reminder of the limitations associated with satellite derived data.

We thank the reviewer for these positive comments and for recognizing that though some of the results are not surprising to those who are experts in ocean colour products, these results offer a reminder of the limitations. We do believe that for non-experts of ocean colour products, these results will be very informative.

Whilst interesting and worthy of publishing in Biogeosciences, I have some concerns with some of the quite sweeping conclusions that appear to be derived from comparison with a single dataset or single datapoint (referred to in more detail below). In addition, I have made other specific comments that I believe should be addressed be-fore this manuscript is suitable for publication.

The reviewer's comment relates to Fig 8, and the discussion of the results shown in it for one location and one year. This figure was only supposed to be illustrative, not the main basis for our conclusions. We have significantly rewritten this section to make this obvious, and include several more panels in Fig 9 that show that the inferences we gain from that location are relevant for much of the globe. See below.

SPECIFIC COMMENTS

P 1, L27 – I am unclear whether the term '. . .real world Chl-a. . .' used here refers to actual in-situ chlorophyll-a concentration or satellite derived chlorophyll-a concentration. The term 'real' appears to be used interchangeably throughout the manuscript.

We agree that this incident of the use of "real" was confusing and will change this to be "to real world satellite-derived Chl". We have also gone through the manuscript being careful how we use "real world" in the text. We also add a sentence in the introduction to define "real world" for the rest of the paper:

"*"(In this article "real-world" will be used to refer to the real ocean and the real derived ocean colour products that are provide by space agencies. The "real world" is thus different to the numerical biogeochemical/ecosystem/optical model output and the products derived from it.)"*

P5, L15 – Would it make more sense to swap Figures 1 and 2 so that the comparison of model and OC-CCI reflectance appears first as Fig 1 and is followed by the product comparisons as Fig 2. If this were the case, would there then be anything gained by comparing model actual and derived Chl-a to the OC-CCI Chl-a product? As I under-stand it, this paper is more about using the model output as a test ground to compare model 'actual' to 'derived' Chl-a rather than testing how well the model replicates the real world' values. Just a thought.

We like the "thought" – it would make a cleaner paper not getting into the aspect of evaluating the model. However, one of the points we would like to make is that comparing model "actual" or model "derived" Chl-a to "real-world" satellite derived Chl-a leads to different results. There are therefore important implications of our analysis for model validation (i.e. that comparing model chl-a to real-world satellite-derived Chla is not comparing like-for-like), so we feel this comparison is an important part of the paper. We have considerably rewritten the relevant paragraph (last one of section 3.1) to make this point clearer and to emphasis that this is not a "evaluation", but a demonstration (underlined is added or altered text).

"Finally in this section, we ask: Which model Chl-a (derived versus actual) best matches real-world satellite-like derived Chl-a (here the OC-CCI product)? We do not do this not for model validation purposes (see evaluation in Dutkiewicz et al., 2015), but rather to re-emphasis that the derived Chl-a product is only a proxy for actual Chl-a: the two are not the same thing. We compare model climatological monthly model derived Chl-a and model actual Chl-a to OC-CCI monthly climatology regridded to the model configuration (1 degree resolution). We find that the model derived Chl-a has global RMSE of 0.2867mg/m^3, which is significantly lower than 0.6370mg/m^3 found when comparing model actual Chl-a to OC-CCI. Comparisons are particularly better for the Southern Ocean and North Pacific (Fig 1). Consequently, some (though certainly not all) of the biases noted when comparing model actual Chl-a (Fig 1b,e) to real world satellite derived Chl-a products (Fig 1a,d, section 2 and in the model evaluation done in Dutkiewicz et al. 2015) are due to the real world Chl-a derived product bias and not a deficiency in the biogeochemical/ecosystem/optical model. It follows that a model satellite-like derived products (Fig 1 c,f) might be a better evaluation tools for comparing to ocean colour products derived with the same algorithm (Fig 1a,d) than the model actual Chl-a fields themselves."

Though we agree that it would be nice to have figure 2 (reflectance) before figure 1 (product: Chl-a) we could not find a way to do this without complicating the paper. So we have left as is.

P6, L13 – The authors talk about comparing '. . .locations and dates similar to those in NOMAD.' What is their definition of similar?

The model has 1 degree resolution – so we are using the degree box within which the actual is situ measurements are taken. This is unnecessarily confusing though and we now state differently:

".. at locations and dates nearest those in NOMAD"

P6, L15 – Again they use 'similar' to describe the resulting relationship between model 'actual' chlorophyll, model X, and real world in situ observations without really defining what similar means.

Here we had meant to reference Figure 4, where we show the OC4 and OC3M-547 functions. We now change this sentence to reflect this:

"The resulting relationship between model blue/green reflectance ratio (X) and Chl-a from subsampling the model (Fig 3a) is similar to that found for real-world algorithms (Fig 4)."

P6, L20 – The authors make no mention of the discrepancy between model reflectance wavebands (blue – 450 nm, 475 nm or 500 nm, green – 550 nm) and those used in the OC4 (blue – 443 nm, 490 nm or 510 nm, green – 555 nm) or OC3M-547 (blue – 443 nm or 488 nm, green – 547 nm) algorithms when comparing coefficients in Table 1. It might make it clearer to those not familiar with the derivation of these algorithms that they are not comparing like with like.

Yes, thank you. This is an important point and in the revised version we will add the following to the text and to the figure caption.

We will add where we compare the model RRS to OC-CCI Rrs:

"We note that the model does not have the exact same wavebands as any of the ocean colour satellites, and as such here we compare to the nearest bands: 450nm model to 443nm for the OC-CCI product, and 550nm model to the 555nm OC-CCI product."

Also when we compare the model GS coefficients to OC4 and OC3-457 (Fig 4 and Table 1) we now add the following sentence:

"Some of the differences between real-world and model coefficients is likely to come from the use of different exact bands in the blue and green (e.g. 550nm for model green versus 555nm for OC-CCI)."

We also add the following to the figure caption in Figure 2:

"We compare the model wavebands against the nearest OC-CCI wavebands, but note that they are not identical."

P7, L1-5 – The authors compare the OC-CCI Chl-a product to model derived Chl-a (although see my second comment). Where are the plots to support the statistics? What monthly climatologies are used to generate these statistics (is it a combination of Jan and Jul or all months?) Over what period are the January and July OC-CCI mean values determined? Are the OC-CCI output averaged to 1 degree by 1 degree similar to the model output? What version of the OC-CCI product is used? Are these OC-CCI products just OCx type output or do they include data from the Hu CI algorithm? The OC-CCI output is just one product. The statement at L5 seems to be quite a bold statement to make when only one product has been compared.

The product from OC-CCI was OC4 without the Hu CI component. We discuss the Hu et al adjustments now in the introduction and text (see comments by Reviewer 1).

The RMSE is determined from daily values over the full time period (not just Jan and Jul) averaged globally. We have regridded OC-CCI onto the same grid as the model for these comparisons.

We would prefer not to add any additional figures, especially for this point. However, we do agree with the reviewers' concern on this section and have considerably rewritten it. We add some additional text (also in the figure caption) to explain the statistics better and state what OC-CCI product we are using. Additionally we add text in both introduction and summary to highlight that OC4 is just one product, and that there are newer (and better) products also available. We have stuck here to the OC4 as it was the one that compared to the version of OC-CCI that we downloaded (a new version (V3) has just been released). We do specifically mention the newer OCx+CI versions. We are not convinced that the statement on L5 is too bold. However we do add a caveat in it.

The paragraph questioned here is provided above for the discussion on Fig 1 (P5, L15) above. And we add the following sentence to caption of Figure 1:

"We use version 2 of the OC-CCI, which uses an OC4 algorithm for determining the Chl-a product, and thus comparable algorithm as used in our model derived Chl-a shown in e,f."

In the introduction we have the following (underlined is added or altered text):

"There is significant ongoing work to improve algorithms. For instance, the newest National Aeronautics and Space Administration (NASA) reprocessing of Chl-a products has included a merged approach which uses different combination of reflectance bands at low Chl-a (Hu et al., 2012). There have also been many attempts to develop more mechanistically derived algorithms (e.g. using known relationships between absorption, scattering and reflectance). Here we focus on the Chl-a estimated from the blue/green reflectance as it is still the most commonly known product, and until very recently used in products downloaded from both NASA and the European Space Agency (ESA) data portals, as well as merged products such as the Ocean Colour Climate Change Initiative (OC-CCI). However we note that similar techniques used in this paper could help inform on other algorithms."

The final sentence of the summary now reads (underlined is new text):

"We also hope that the ocean colour community will see the potential of model approaches such as this for deriving sampling strategies, further studies on newer Chl-a algorithms (e.g. NASA Reprocessing 2014.0, and OC-CCI V3 release), other ocean colour products, and will help with algorithm developments for current and future ocean colour measurements."

P7, L20 – Perhaps it's my eyesight but I'm not convinced that Figs (b) and (e) show '. . .much lower biases at high latitudes. . .'. (I assume you are comparing Fig 6 (b) and (e) to Fig. 6 (a) and (d))

Yes we were comparing b,e to a,d – will make clearer in the revised version. And we agree that this is not obvious from the figure – the improvement is mostly just at the edge of the data before moving into the white areas. We now restate this sentence (underlined is new/altered text):

"Though there is some improvements in some regions in the higher latitudes, there is actually decrease in skill at lower latitudes (Fig 6b,e compared to a,d). There is in fact a slight increase in the mean % absolute bias (23%) between this and the GS estimates: When transformed into percent errors the increased biases at low Chl-a, low latitude regions become more prominent."

P8, L3-5 – Are grid cells with depths less than 1000m also excluded?

Yes, will be stated in the revised version.

P8, L22-24 – These statements appear to be derived from data taken from one point in the North Atlantic. Is this a fair representation of the global pattern or is it just representative of this location?

We use the one location as an illustration. The results shown in figure 9 show that same lag in initiation of the bloom occur across much of the high latitudes. We make clear in the text and Figure 8 caption that this is just an illustration. To further bolster the relevance of this discussion in space and time, we now include other panels in figure 9 to show the mismatch in the peak timing, as well as in the initiation of increase in CDOM and detrital matter relative to actual Chl-a.

Section 4 now reads (underlined is altered or added text):

"*We have noted that in all approaches, though even more obvious in RA, there is a seasonally altering pattern between the derived and actual model Chl-a (Fig 6). The amplitude of the peak of spring blooms is often underestimated in the products derived using global coefficients (GS and GA) in high latitude, especially in the subsampled algorithm (GS) (Figs 6). , Derived Chl-a values were also often higher than model actual Chl-a outside of bloom peaks. We consider the phenology here, using a single location (in the subpolar North Atlantic) for a single year as illustration (Fig 8a). Though the derived products show similar (though smaller) peaks to the actual Chl-a, and sometimes similar peak timing early in the season (see for instance the first distinct peak in this illustrative location), there are noticeable lags for the maximum peak (shown with a dotted line) and other mismatches later in the season. We also find that*

the bloom period last later into the year. The actual Chl-a also starts its sharp increase in spring (the initiation of the spring bloom, shown with dashed line) considerable before all three derived products (Fig 8a). We follow the approach of Cole et al (2012) for determining the "initiation of the spring bloom" as the time when the Chl-a first increases 5% above the annual median (horizontal dashed line, more description in Appendix A).

Figure 8 shows just one location for 1 year. To consider the large scale patterns, we determine the lag in the spring initiation (Fig 9a) and maximum bloom timing (Fig 9b) for each location averaged over all years. We find that in almost all locations the derived Chl-a shows the bloom starting later than the model actual Chl-a (Fig 9a). This offset is typically by about 5-10 days but can be as much as 30 days. The maximum Chl-a from the derived product also lags the actual Chl-a in most locations, though by only a few days (Fig 9b). These result indicates that temporal as well as spatial biases occur as a result of deriving Chl-a from X and suggests care should be taken when calculating phenology from satellite products or when evaluating phenology in models using satellite-derived Chl-a. We discuss the reason for the lags in the next section."

Figure 8 is now more clearly described as just an illustration with the connection to Figure 9 made to make obvious that Fig 9 has the global results. We also add additional lines to Fig 8 to show the initiation of the bloom. We attach the new Fig 8 and 9, and captions below.

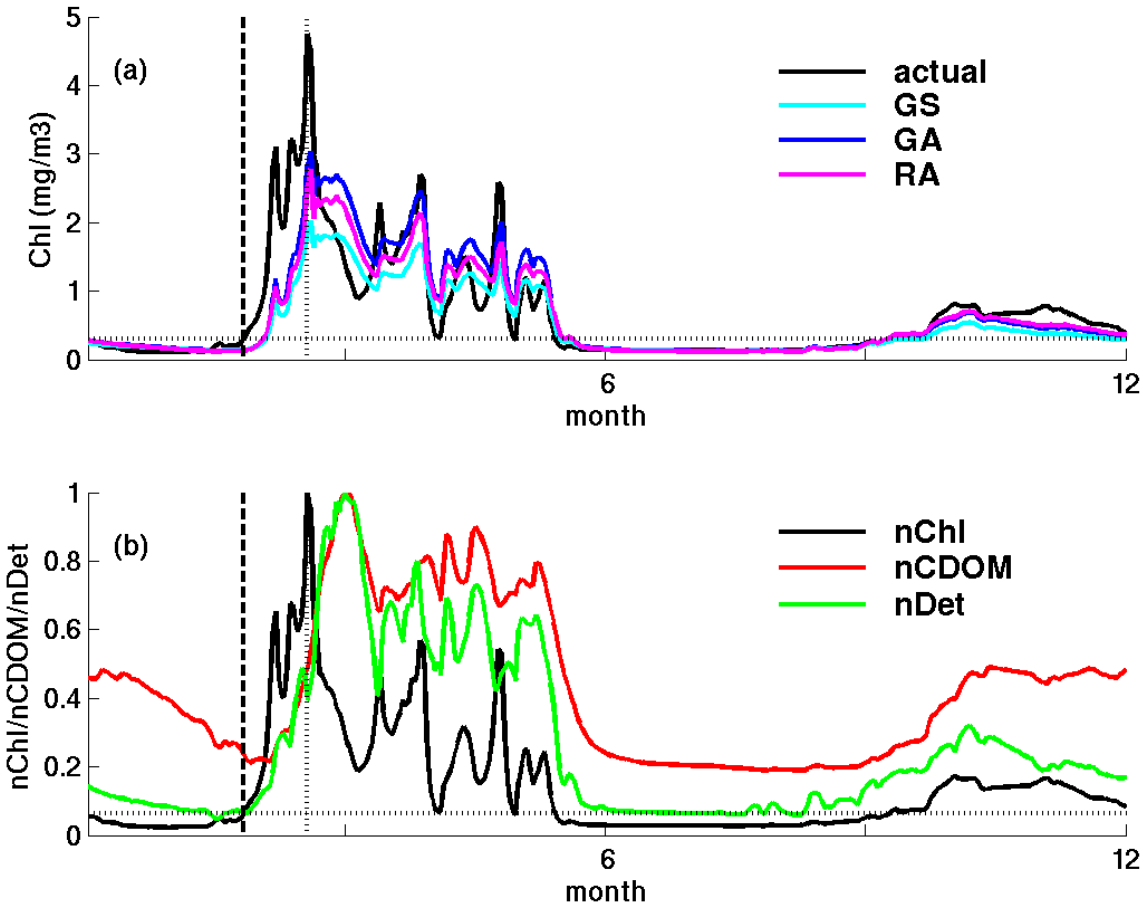

Figure 8: Illustrative timeseries for one year from a single location in the North Atlantic (shown as x on Fig 9). (a) "actual" Chl-a (black), derived Chl-a using subsampled output (GS, light blue), derived Chl-a using all output (GA, dark blue), and the Chl-a product derived using a regional specific algorithm (RA, purple). (b) actual Chl-a (black), CDOM (red) and detritus (green), all normalized to their peak value. Dashed vertical line indicates the "initiation of the bloom" which is taken to be when Chl-a reaches 5% above the annual median value (dotted horizontal line shows this value for the model actual Chl-a), following Cole et al (2012) and discussed further in Appendix A. The vertical dotted line indicates the peak of the bloom. Shown here is only a single year and location, however for larger scale perspective, the difference in initiation and peak timing between model actual and derived Chl-a averaged over all years are shown for the globe in Figure 9.

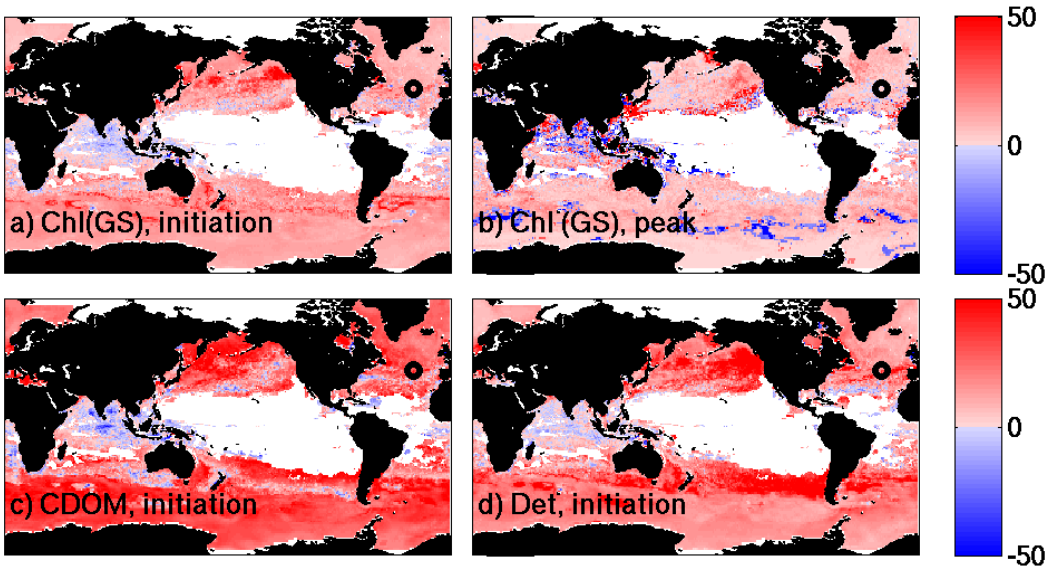

Figure 9: Lag in phenology. Number of days between a) the initiation of the spring bloom from model actual Chl-a and that for the model derived Chl-a (GS) ;b) yearly maximum of model actual Chl-a and that for the derived Chl-a (GS); c) initiation of the spring bloom from model actual Chl-a and the initiation of the CDOM increase; d) initiation of the spring bloom from model actual Chl-a and the initiation of detrital particle increase. Bloom initiation is defined as when Chl-a, CDOM or detrital particles reach 5% above their annual median value (see Appendix A). Lack of output indicate did regions with no significant seasonal cycle or are not resolved by the model (e.g. Arctic Ocean).

P9, L4 – The authors could reference the Dutkiewicz et al. (2015) paper again here.

Good idea, we will do so in the revised version.

P9, L5 – The authors refer to 'studies' then reference a single instance.

We add "e.g" to the text and additional references (e.g. Loisel et al., 2010; Brown et al., 2008, Siegel et al., 2005a, 2005b)

Brown, C.A., Huot, Y., Werdell, P.J., Gentili, B., and Claustre, H.: The origin and global distribution of second order variability in satellite ocean color and its potential applications to algorithm development, Remote Sensing of Environment, 112, 4186-4203.

Loisel, H., Lebac, B., Dessailly, D., Duforet-Gaurier, L., and Vantrpotte, V.: Effect of inherent optical properties variability on chlorophyll retrieval from ocean colr remote sensing: an in situ approach. Optics Express, 18,

Siegel, D.A., Maritorena, S. Nelson, N.B., and Behrenfeld, M.J.: Independence and interdependencies of global ocean color properties; Reassessing the bio-optical assumption. J. Geophys. Res., 110, C07011, doi:10.1029/2004JC002527, 2005a

Siegel, D.A., Maritorena, S., Nelson, N.B., Behrenfeld, M.J. and McClain, C.R.: Colored dissolved organic matter and its influence on the satellite-based characterization of the ocean biosphere, Geophys. Res. Letters, 32, L20605, doi:10.1029/2005GL024310, 2005b.

P9, L17-21 How do the authors support this statement? If it is the timeseries data in Figure 8, then these are data from just one point in the North Atlantic. I don't think that data from one location and for one year is sufficient to warrant these conclusions.

See our comments above for P8, L22-24 above. The results in Figure 9 show the global results averaged over 13 years. However we do see the reviewers point, and now add the extra panels (c,d) to Figure 9. We modify the text in Section 5 (original P9, L17-19) as such (underlined is added or altered text):

"However, we find that though linked, there are noticeable lags in the sharp increase in accumulation (Fig 8b, Fig 9 c,d) and peak timing and decline(Fig 8b) between CDOM and detrital matter and the model actual Chl-a."

Figure 3 – How do you differentiate between zero bias and lack of data? Could I suggest that lack of data is coloured differently to zero bias?

I assume you mean Figure 6? Good idea. We will do so in the revised paper.

Figure 4 – Not sure whether the figure order works. The first mention of Fig. 4 that I can find occurs on P11 after reference to all the other figures. Again, if the authors are comparing the polynomials it might make it clearer to the reader if they acknowledge that different wavelengths have been used in the derivation in the legend.

Figure 4 is mentioned first on pg 6 (line 22), but we now also reference it earlier in the newer version of the paper. Since two of lines in Figure 4 are those in Figure 3, it does quite naturally belong here.

We add the acknowledgment of different wavelengths in the caption of Figure 4:
"Note that the algorithms for the model come from band ratio of 425nm/450nm/475nm and 550nm. For the real world algorithms the band ratios are different and specific for the satellite sensor (SeaWifs or MODIS)."

Figure 8 – I don't think the x axis matches the label in Fig 8 (b). I assume the vertical dotted line marks the peak in 'actual' Chl-a?

Thank you for catching this – we have altered the axis to match Fig 8a. Yes, the vertical dotted line is the peak – we now mention this in the figure caption. We have also added a line to show the initiation of the spring bloom for clarity. See new figure 8 and caption above.

TECHNICAL CORRECTIONS

P1, L16 – Should read '. . .Chl-a to the actual. . .'

Thank you – we will fix this.

P1, L25 – Should read 'This result indicates. . .'

Actually, I think should read "These results indicate…", so will change to this instead. Thank you for catching this inconsistency.

P2, L15 – I think this is the first use of the acronym CDOM and so it should be defined here.

Yes, will add.

P2, L18 – Should read 'There have been. . .'

Yes, thank you.

P2, L20 – Should read 'product' instead of 'products'

Yes, will fix

P2, L24 repeats L7

We will remove the text at L24 from the revised text.

P3, L28 – In situ is italicised here but nowhere else.

Thanks – we will remove the italics to be consistent.

P4, L14, 15, 17, 19 - Repeated uses of 'explicit'.

Yes, will redo to remove excessive use of “explicit”. Thank you.

P5, L15 – ‘Fig.2’ is italicized

We will remove the italics. Thank you.

P7, L4 – Missing figure number

Will correct.

P7, L19 - Should read ‘. . .lead to a better. . .’

Yes, thank you.

P9, L3 – Should read ‘lead’ rather than ‘leads’

Yes, thank you

P9, L21 – Should read ‘. . .remains relatively high. . .’

Yes, will fix

P9, L29 – Don’t think there should be a comma after ‘pigments’.

Indeed not, we have removed.

P9, L31 – However, I think there should be one after ‘reflectance’.

Yes, have fixed this in revised version.

P12, L22 Should read ‘. . .by-products. . .’

Yes, thank you.