Responses to Referee #1

He et al use MODIS and SeaWiFS satellites to derive chlorophyll-a concentration within the eastern China seas to identify seasonal and inter-annual variability. A tremendous amount of effort seems to have gone into this manuscript and the authors are off to a reasonably good start. However, I have many concerns regarding the presentation of the data and manuscript. This manuscript requires major revision before it can be considered for acceptance. First the manuscript presents little if any quantitative information or statistical results. The data is presented and described in a qualitative fashion, which is inappropriate for a scientific article. Second, I find the discussion and conclusions to be relatively boring and, as presented, provide little new information. Finally, would benefit from minor overall editing to tighten up some sentences.

Response: Thanks for these general comments. According to reviewer's comments, we have added more quantitative information, revised the discussion and conclusions in the revised manuscript. Also, we have edited the manuscript by a native English speaker. We think the revised manuscript is much improved after the major revision.

Abstract: The abstract provides no quantitative information. The authors state they validate and calibrate chlorophyll algorithms, it doesn't seem like they successfully did any type of rigorous calibration effort. Instead they used a previously calibrated algorithm and performed a validation using in-situ data. The results of the validation are never discussed, other than the algorithms overestimate in highly turbid waters. This has repeatedly been the case with MODIS and SeaWiFS OC3 and OC4 algorithms.

Response: Thanks for these comments. We have added more quantitative information in the abstract in the revised manuscript. In the paper, the purpose of validation is to correct the systematical bias between satellite-derived Chla and in situ Chla, and to ensure a proper data set to exam the bloom variation, but not to calibrate the coefficients of Chla algorithms (OC3M or OC4) nor to develop a new regional Chla algorithm. To make it clearer, we changed the word "calibration" into the "correction" in the revised manuscript.

In addition, we have added more details of the validation results. According to the suggestion from the reviewer, we carried out more rigorous validation following the method of Bailey and Werdell (2006) in the revised manuscript, using native high resolution L2 satellite data and narrow time window (less than 3h) with considering the homogeneity test (5×5 box) and L2 flags. Using the new

validation results, we recalculate the bloom statistics in the revised manuscript.

It isn't clear what the authors mean by the concentration caused by high water turbidity was less than 10 µg/L. How is this justified in being a threshold for bloom identification?

Response: Thanks for this comment. Generally, satellite remote sensing overestimates Chla in turbid water using the empirical algorithm based on blue-to-green bands ratio, such as the OC3M and OC4 algorithms. In this paper, we found that such overestimation had upper limit value (taken as the threshold for bloom identification in this paper) as the increase of water turbidity. And, we evidenced that the upper limit value was generally less than 10µg/l from both in situ and satellite data. The purpose of finding the upper limit value is to eliminate the "false bloom" caused by the satellite overestimation in turbid coastal oceans. We have added these discussions in the revised manuscript.

In addition, we also carry out a sensitivity analysis by adjusting the thresholds as 8µg/l and 12µg/l as well as 10µg/l to check the effects threshold variation in the revised manuscript. Comparing with the bloom intensity index (BI) derived by 10µg/L, the mean differences of the BI derived by 8µg/L and 12µg/L are 9.8% and 6.4% for the region A in the Changjiang Estuary, and 22.7% and 14.7% for the region B in the Yellow Sea and Bohai Sea. Although the BI slightly varies with the threshold values from 8 to 12µg/L, the inter-annual variation and long-term trends are consistent well. The results prove that the influence of small variation of our chosen 10µg/L as threshold for large boom identification is quite less.

What are the statistical results from the ENSO and PDO cycles? Is it possible to really assess the ENSO and PDO decadal cycles if the authors have just over 1 decade of data? Seems like a stretch to me.

Response: We agree with the review's comment and remove the last sentence about the ENSO impacts in the abstract. It is difficult to assess the impacts of ENSO and PDO on the blooms using only 14 years dataset. Yet, we think it is still possible to analyze the relationship between ENSO, PDO and bloom intensity index. In the revised manuscript, we have added more statistical results and focused on the relationship analysis instead of the impacts.

Introduction: The authors mention previous work that used CZCS and OCTS but state the algorithms over estimated chlorophyll, how is that any different than the MODIS and SeaWiFS performance?

Response: Thanks for this comment. We have rewritten this part in the introduction. We want to present two opinions to show the importance of our

study. First, although there have been some studies to investigate the temporal and spatial distribution of satellite-derived chla, the seasonal and inter-annual variations of phytoplankton blooms have never been investigated in the East China Sea. Second, chla concentration can be used as a practical index for phytoplankton bloom, but the satellite algorithm always overestimate chla concentration in high turbid coastal water (for all the CZCS, OCTS, MODIS and SeaWiFS data), and false booms would be counted from such overestimated chla data. Therefore, in our study, we will first correct the systematic error between satellite chla and in situ chla to ensure a good dataset for further study, find a credible chla threshold for bloom identification in the turbid coastal oceans, and then present and analysis the bloom distribution.

The authors also mention Yamaguchi and Shi and Wang looked at seasonal climatology and interannual variations of chlorophyll. Then they state their work is different because seasonal and interannual variations have never been investigated. This is contradictory.

Response: Thanks for this comment. In the manuscript, we mean that the seasonal and inter-annual variations of phytoplankton blooms have never been investigated in the ECS, although both Shi and Wang (2012) and Yamaguchi et al. (2012) have analyzed the spatial and temporal variability of chla in the eastern China seas.

Data: What do the authors mean "preserve statistical rigor"?

Response: Thanks for this comment. In this paper, we need to count the numbers of bloom events in each year. If data is lacking in several months in one year, the statistical results would be affected by the total number of valid samples. Although SeaWiFS worked successfully until the end of 2010, data is unavailable for several months after 2008 due to technical difficulties with the sensor. So, for the validness of statistical results, we used the continued SeaWiFS data from 1998 to 2007. We have added this discussion and deleted the words "preserve statistical rigor" in the revised manuscript.

It doesn't seem like the authors undertook any calibration effort, so this should be removed from the text.

Response: Thanks for this suggestion. In the paper, we used the validation results to correct the systematical bias between the satellite retrieval and in situ Chla to ensure a proper data set for the study of bloom variation, but not to calibrate the coefficients of Chla algorithms (OC3 or OC4) nor to develop a new regional Chla algorithm. We agree with the review's comment, and change the word "calibration" into the "correction" in the revised manuscript. Since the accuracy of satellite product is very important for the confident of satellite

oceanography application, we think this part is important and necessary for this paper.

What method was used to match in-situ data with satellite data? Bailey and Werdell provide a standard method which is the accepted approach. They recommend using only L2 data at the native resolution of the sensor. Instead the authors used daily binned data, which is problematic because multiple pixels may have been spatially composited into a daily data bin. The authors never discuss the box extraction, do they use 1x1, 3x3, or greater? Do the authors filter for viewing angle at all? They should follow the method of Bailey and Werdell.

Response: Thanks for these comments and suggestion. In the original manuscript, we matched the daily binned satellite data with the in-situ data according to same-day and closest pixel. According to review's suggestion, we followed the validation method proposed by Bailey and Werdell (2006) in the revised manuscript. We used the native high resolution L2 satellite Chla data and narrow time window (less than 3h) with considering the homogeneity test (5x5 box extraction) and L2 flags. Then, we recalculate the bloom statistical results in the revised manuscript.

What is meant by "heavy cloud masking"? Do the authors know all of the masking was due to clouds using the L2 flag array?

Response: Thanks for this comment. We have deleted this sentence in the revised manuscript.

Delete the words "powerfully organized".

Response: We have deleted these words in the revised manuscript, thanks.

What is meant by having "almost the same sampling stations?"

Response: To be cleared, we have revised this sentence into "The two cruises had almost the same sampling locations and used the same measurement method for the chla", thanks.

Where the in-situ chlorophyll samples corrected for phaeopigments, what method was used, please provide a reference.

Response: The water samples for chlorophyll a (chla) were filtered onto 25mm Whatman GF/F filters with 0.7μm pore size. The filters were preserved in liquid nitrogen before being processed. Each filter was analyzed with a Turner-Design10 fluorometer to obtain the chla concentration using the non-acidification fluorometric method (Welschmeyer, 1994). We have added this method in the revised manuscript, thanks.

Remove the sentence about "a valuable and treasured dataset."

Response: We have removed this sentence in the revised manuscript, thanks.

How do the authors handle multiple samples with depths less than 5m? Are they averaged? Weighted?

Response: If there are multiple samples within 5m, we use the topmost surface layer sample in this study. Generally, there are two layers within 5m, one is the topmost surface layer, and the other is 5m depth layer. We have added this in the revised manuscript, thanks.

2.2.3: How is the "surface layer" defined?

Response: In the serial oceanographic data sets, there are 0m, 10m, 20m, … layers, and in this study, we define the "surface layer" as the zero depth layer. We have added this in the revised manuscript, thanks.

How are "serial stations" defined?

Response: That means the time series stations. We have changed the "serial stations" into the "time-series stations" in the revised manuscript, thanks.

How is a "serial oceanographic survey" defined?

Response: That means the regularly oceanographic survey. For the KODC data sets, both the nitrate and phosphate concentrations were regularly measured in February, April, June, August, October and December in every year. For the JMA data sets, both nitrate and phosphate concentrations were usually measured at least in January, April, July, and October to represent four seasons of a year. We have added these in the revised manuscript, thanks.

3: When the annual means were calculated, how was the spatial extent addressed? Was this the annual mean for the entire eastern China sea?

Response: The annual mean Chla was calculated for each pixel of SeaWiFS and Aqua/MODIS data. Thus, there were 5 annual-mean Chla maps for both Aqua/MODIS and SeaWiFS during their overlapped observing time from 2003 to 2007. We have added this in the revised manuscript, thanks.

It seems odd the authors discuss figure 2 then jump back to figure 1. Figures should be discussed sequentially.

Response: Thanks for this comments. We have corrected this in the revised manuscript.

What is meant by SeaWiFS overestimates more?

Response: Thanks for this comment. The purpose of section 3.1 is to remove the systematic difference between SeaWiFS and Aqua/MODIS data, instead of

evaluate them. In the revised manuscript, we have removed this sentence.

3.2: What is meant by "same daily time window"? What hours were used, typically +/- 3hrs of satellite overpass are retained for match-ups. What solz and senz angles were filtered for each satellite? What variance was used to excluded satellite extracted points during the match-up process?

Response: Thanks for these comments. The "same daily time window" means that satellite observation and in situ measurement are within the same day. In the revised manuscript, we followed the validation method proposed by Bailey and Werdell (2006) in the revised manuscript. We used the native high resolution L2 satellite Chla data and narrow time window (less than 3h) with considering the homogeneity test (5x5 box extraction with CV less than 0.15) and L2 flags.

How is "heavy cloud cover defined"?

Response: Thanks for this comment. We have removed these words in the revised manuscript.

How do the authors justify "good performance"? Instead of qualitatively stating this, it should quantitatively be presented.

Response: Thanks for this comment. We have presented the linear relationship coefficients in the revised manuscript. The correlation coefficients between the satellite retrieval and in situ chla values in log transformation scale are 0.85 and 0.72 with SD of 0.30 and 0.40 for the SeaWiFS and Aqua/MODIS data, respectively.

It isn't clearly present how the authors choose Rrs555>0.005 as an indicator of high turbidity and where the overestimation of chlorophyll begins. Even when chlorophyll associated with Rrs555<0.005 are considered the linear regression follows the same trend of "overestimation." So this doesn't seem like a justified approach.

Response: Thanks for these comments. We agree with reviewer's comments. In the revised manuscript, we use the linear regression for the whole matched-up data sets instead of the sub-dataset with Rrs555<0.005.

Again, the authors don't clearly justify their selection of 10ug/L. A selection of 17.9 or 5.7 could have also been selected (authors discuss these numbers in 5.4) and would have similar results, so why 10?

Response: Thanks for this comment. We need to set a threshold for bloom identification to ensure a uniform criteria for the assessment of bloom inter-annual variations. In this paper, we have found that there were upper limit values for overestimations of satellite-derived Chla caused by extremely high turbidity. Thus, we used this upper limit value (threshold for bloom identification) to avoid the "false blooms" in the turbid coastal waters. We figure

out that the upper limit value was generally less than 10μg/l from both in situ and satellite data. In addition, we have carried out sensitivity analysis by adjusting the threshold values of 8μg/l, 10μg/l and 12μg/l to check the effects.

What do the authors mean, "Thus, the statistics of blooms are expected to be reasonable and meaniful"? This is an ambiguous statement and no statistics are presented, so it is impossible to tell what is being presented.

Response: We agree with reviewer's opinion, and remove this sentence in the revised manuscript, thanks.

5 Discussion: Overall, I think the discussion section tends to ramble on, and in many parts reads more like a results section. The discussion seems to lack the expected comparison of other systems. How does this compare to other areas around the world, such as the Gulf of Mexico.

Response: Thanks for these comments. We have reorganized the discussion section in the revised manuscript, which focus on the comparisons of nutrient supply in two different systems of Region A and Region B since these two region show difference long-term trend of bloom intensity during the past 14 year. Region A is the estuary area influenced by the large Changjiang river plume, and Region B is the semi-enclosed marginal sea with less influenced by large river plume.

Pg 127, section 25 is awkward and unclear. What are the authors talking about when they say the SeaWiFS plumes were inversed for CDOM?

Response: Thanks for this comment. This section is not very important, and we have removed this section in the revised manuscript, thanks.

The section on nutrients reads much like a result section and only provides qualitative discussion. Some statistical analysis should be conducted, perhaps a multiple linear regression to prove there is a relationship instead of just guessing.

Response: Thanks for these comments. Due to two major reasons, we currently can only give the trend analysis instead of the relationship regression. First, the spatial-temporal coverage between bloom intensity index (BI) and in situ nutrient are quite different. The satellite-derived BI is a regional and annual integrated value, while in situ nutrient is sampled at a specific station and time. Second, the responses of the phytoplankton bloom to nutrients loading are quite complicated. The phytoplankton blooms depend not only on the rich nutrients, but also on the water temperature, vertical mixing, solar radiation, and so on. Thus, in this study, we only focus on the long-term trend analysis. Of course, the mechanisms and processes analysis of the response of phytoplankton bloom on nutrients increase is an interesting subject worthy of study in the future, but it is beyond the scope of this study. We have revised the discussion part to make it clearer.

The connection to ENSO and PDO seems very weak, and requires quantitative analysis and statistics.

Response: Thank for this comment. We have added the discussion about this link, and presented the statistical results in the revised manuscript.

6. Conclusion: The authors state, "despite significant inter-annual variation" but the significance is never presented.

Response: Thank for this comment. The magnitude of the inter-annual variation of BI in the CJE (region A) (annual bloom intensity index varying from 64336 kg/m to 194253 kg/m) was presented at the section 4.4 in the revised manuscript.

In the previous sections the author highlight the overestimation of 10 ug/L now they discuss an overestimation for chla less than 3 ug/L.

Response: Thank for this comment. There are two different concepts. 10ug/L is the upper limit value that water turbidity can potentially induce the satellite overestimation of the Chla retrieval. Specifically, in the extreme turbid water with zero chlorophyll concentration, the false satellite-retrieved Chla can be up to 10ug/L. However, if the in situ Chla is very high (e.g. 20ug/L), then the water turbidity could cause the underestimation of satellite-retrieved Chla than the in situ value. To avoid confusing, we delete the 3 ug/L in the revised manuscript.

Back to my original point, using Rrs>0.005 seems arbitrary and the overestimation is the same if all data is included, only data above Rrs>0.005, or Rrs<0.005.

Response: Thank for this suggestion. In the revised manuscript, we have used all match-up data instead of using the sub-dataset with Rrs>0.005 or Rrs<0.005.

Figures 2 and 3 present no statistics, what is the regression equation, R2, RMSE, and significance?

Response: Sorry for the unclear description. The statistical results (regression equation, R, RMSE, and significance) were already given in the Eq.(1) for Fig.2 and the line 15 (Page 120 of the original manuscript) for Fig.3.

Figure 5: This figure isn't clearly explained in my opinion, I'm not sure what computation was undertaken.

Response: Sorry for the unclear explanation. In Fig.5, we computed the maximum Chla value for each pixel from all 8-day composite Chla data sets from SeaWiFS (1998–2007) and Aqua/MODIS (2003–2011). Similarly, we compute the maximum and minimum of Rrs555 values from all 8-day composite Rrs555 data sets from SeaWiFS (1998–2007) and Aqua/MODIS (2003–2011). The purpose of Fig.5 is to evident that the satellite overestimation of Chla caused by high water turbidity is less than a threshold value (such as 10μg/L).

Figure 9 and 13: What is the slope? Is it significantly different from zero?

Response: Thank for this comment. The slopes in Fig.9 are -0.204% and -0.297% per-year for SeaWiFS (1998–2007) and Aqua/MODIS (2003–2011), respectively. The multiple-station averaged slopes in Fig.13 are 0.37μmol/L and 0.05μmol/L per-year for the nitrate and phosphate concentrations, respectively. We have added these in the revised manuscript.