We really appreciate the anonymous referee #1 for your very positive and constructive comments. Almost all suggestions were accepted and the manuscript has revised as below.

RC = Referee's comments; AR = Authors' Response (written in blue); RS=reconstructed sentences (written in green)

First of all we changed the manuscript title from "Influence of timing of sea ice retreat on phytoplankton size during marginal ice zone bloom period **in** the Chukchi and Bering shelves" to "Influence of timing of sea ice retreat on phytoplankton size during marginal ice zone bloom period **on** the Chukchi and Bering shelves".

**RC1:** Page 12613, Line 1-4: This sentence needs to be revised. How TSR plays crucial roles under seasonally ice-covered environments? You clearly have to mention one of expectable reasons. For example, TSR plays crucial roles in controlling distribution of phytoplankton community size structure in the seasonally ice-covered marine ecosystem (or) in affecting food-web structure (You mentioned that TSR is crucial to understand to evaluate bottom-up effects of PP on the food web: in line 13-15, page 12616).

AR: Thank you for pointing it out. I reconstructed the first two sentences as below adding a sentence.

RS: The size structure and biomass of a phytoplankton community during the spring bloom period can affect the energy use of higher-trophic-level organisms through the predator-prey body size relationships. The timing of the sea ice retreat (TSR) also plays crucial roles in seasonally ice-covered marine ecosystem, because it is tightly coupled with the timing of the spring bloom. Thus, it is important to monitor the temporal and spatial distributions of a phytoplankton community size structure.

**RC2:** Page 12615-12616, Line 29-Line 1-2: Change "ecosystem changes" to "phytoplankton community changes". Application of these regional ocean color algorithms is expected to contribute to comprehension of phytoplankton community changes (not ecosystem changes) in Arctic ecosystems.

AR: Thank you for pointing out. I revised the sentence as below.

RS: The application of these regional ocean colour algorithms to long-time-series satellite data is expected to contribute to our comprehension of phytoplankton community changes.

**RC3:** Page 12617, Line 19-20: I agree that the development of thermal stratification is an important controlling factor of nutrient supply into the upper layer. In the section 4.3, you supposed that the nutrient conditions in the surface are mainly related to the thermal stratification. However, the local wind filed could be important for affecting the nutrient vertical distribution as well as horizontal distribution in the Bering and Chukchi Sea.

AR: Thank you for the very valuable comment. I agree with you that local wind is important to determine nutrient conditions as several studies have already reported. I wish I could analyze wind data in this study, too. But unfortunately there is no time to obtain and analyze 16-year time series wind dataset. However, I added several sentences to the top of last paragraph of discussion 4.3 to describe about the importance of taking into account of local wind field for spring bloom as below.

RS: There are several studies that referred to the importance of the local wind field to determine nutrient conditions and chla concentration near the ice edge instead of TSR. For example, Niebauer et al. (1995) showed the effect of the wind velocity on the initiation and duration of the ice-edge bloom in the southern Bering shelf. Arrigo et al. (2014) also reported the wind direction in the northern Chukchi shelf has large impact on the upwelling of nutrients from shelf break. Although inter-annual variability of such wind field can also affect the phytoplankton community size structure during the MIZ bloom period in the local scale, our spatial statistical analysis revealed a general trend of the bloom time $F_L$ toward TSR. It has been reported that the sea ice retreat timing can alter the seasonal succession of phytoplankton in Svalbard (Leu et al., 2011) and the northern Chukchi Sea (Fujiwara et al., 2014). Our results also suggest that the phytoplankton community size composition can change with the timing of the sea ice retreat in the shelf region of the Bering and Chukchi Seas during the MIZ bloom. These results provide important information to help understand the efficiency of energy transfer for organisms that depend on the MIZ blooms as a food source. However, it should be noted that the relationships between the TSR and phytoplankton size might not be applicable to the entire Arctic, because our region of focus is characterized by a shallow bathymetry, sea ice retreat near the summer solstice (Fig. 4a), and the dominance of thin first-year ice. These unique environmental characteristics can tightly couple the phytoplankton bloom and sea ice dynamics.

**RC4:** Page 12622, Line 6-7: How did you measure in situ primary production? Using 13C or 14C? You have to mention the used method at the end of the sentence.

AR: I should provide the information about primary productivity measurement. We measured PP using 13C method. The sentence was reconstructed as below with an additional citation.

RS: PP was measured using the stable $^{13}C$ isotope method (Hama et al., 1983) for several optical depths, and then $PP_{eu}$ was calculated by integrating the PP values from the surface to $Z_{eu}$ (see Hirawake et al., 2012 for more details of the sampling and analysis).

**RC5:** Page 12622, Line 18-20: Based on relationships between in situ FL and satellite FL, you found a slight overestimation in the low FL and underestimation in the high FL range. Could you explain the reason for it?

AR: Thank you very much for pointing out such an important point. I tried to explain why the slope differed from 1:1 line, and I concluded that it is because of underestimation of Rrs from MODIS. Fortunately, we were able to obtain the 13 match-ups between in situ and MODIS-measured Rrs during the cruise of 2013. Figure R1b (same as Figure 2b in revised manuscript) indicates the relationship between in situ and MODIS-Rrs. Although there are linear relationships between the Rrs, we found that MODIS underestimated Rrs for every band I compared (412, 443, 469, 488, 555 and 667 nm). The slopes and intercepts between them are listed in Table R1 (same as Table 2 in revised manuscript). Using these relationships, we converted in situ Rrs to MODIS-Rrs using the dataset used in appendix B. Then, we compared the FL derived from original in situ Rrs and in situ Rrs converted to MODIS-Rrs (Figure R2b and Fig. 3b of revised manuscript). $F_L$ calculated from MODIS-converted Rrs also showed significant underestimation compared to $F_L$ calculated from in situ Rrs (the slopes ranged from ~0.33 to ~0.46), which is very similar to middle to high range of $F_L$ of Figure R2a. Although we understand that the number of dataset is not enough to illustrate the characteristic of MODIS-Rrs and the relationship may vary with seasons and regions, such underestimation of Rrs significantly caused the underestimation of $F_L$ via the calculation steps (i.e. QAA and SDM), at least during summer of 2013. The description about this point has added to result 3.1 as follows.

RS: The accuracy of the SDM-derived $F_L$ was evaluated by comparing the $F_L$ values from the in situ measurements and daily matched MODIS level-2 dataset. Twenty-five data points were available for this examination, which were collected over a wide area of the Bering and Chukchi Seas during different seasons (Fig. 1). Fig. 2a compares the satellite-derived $F_L$ and in situ $F_L$. The SDM successfully retrieved the $F_L$ values for 17 of the 25 data points (68% of the data) within a ±20 % $F_L$ range (Fig. 2a). The RMSE was 25%. The satellite validation was very similar to the results of Fujiwara et al. (2011), who showed that the $F_L$ derived using the in situ measured $R_{rs}(\lambda)$ had a 69% accuracy and an RMSE of 22.7%. However, it should be taken into account that there was a slight overestimation in the low $F_L$ range and relatively large underestimation in the high $F_L$ range (slope = 0.48 and intercept = 0.18). We also evaluated the performance of $R_{rs}(\lambda)$ retrieval of MODIS comparing with in situ-measured $R_{rs}(\lambda)$ at 13 data points during the GRENE cruise of 2013. Fig. 2b shows the comparison of in situ- and satellite-derived $R_{rs}(\lambda)$, and we found that the MODIS significantly underestimates $R_{rs}(\lambda)$ at every wavelength (the slopes ranged from to ~0.34 to ~0.46). The slopes and intercepts between the two $R_{rs}(\lambda)$ are listed in Table 2, which was used as the factors to convert in situ $R_{rs}(\lambda)$ to MODIS-$R_{rs}(\lambda)$. The conversion factors were applied to the $R_{rs}(\lambda)$ dataset used in appendix B, and then, we compared the SDM performance between in situ $R_{rs}(\lambda)$ and $R_{rs}(\lambda)$ converted to MODIS-$R_{rs}(\lambda)$ to assess how the retrieval errors in the MODIS-$R_{rs}(\lambda)$ affect estimation accuracy of the $F_L$. The relationship between the two derived $F_L$ is shown in Fig. 2c, and we found similar relationship with Fig. 2a. It can be said that the underestimations of $R_{rs}(\lambda)$ significantly caused the underestimation of $F_L$ especially in the middle to high range of $F_L$. Although we conducted the optimization of the QAA to derive more accurate input parameters of the SDM (appendix B), retrieval error in the $R_{rs}(\lambda)$ significantly affected the SDM accuracy in the study region. Nevertheless, the high determination coefficient indicated that the correlation was sufficient to address the spatio- and temporal variability of $F_L$. Here, we conclude that the SDM is applicable to satellite remote sensing in the Bering and Chukchi Seas, and is effectively optimized for what is known as optically complex water (Matsuoka et al. 2007, Naik et al. 2013).

**RC6:** Page 12624, Line 17: How did you determine the open water period? Is this meant SIC reach zero? Define the open water period in the materials and methods section.

AR: I should provide the information. I defined the period as the number of day of SIC < 10%. The sentence was added describing the definition as below in materials and methods section.

RS: The length of the open-water period for each pixel was defined as the number of days where the SIC was less than 10%.

**RC7:** Page 12626, Line 19-25: You found that the timing of the MIZ bloom occurrence was not tightly coupled with the TSR in the western side of the Bering Strait and coastal Alaska. Plausible reasons (nutrient conditions or CDOM effects) are suggested for explaining it.

However, the coastal Alaska is the region affecting by nutrient-poor Alaska Coastal Waters not nutrient-rich Anadyr Water.

AR: Thank you for the comment. I should explain separately about Bering Strait and coastal Alaskan area. The related sentences were reconstructed as below.

RS: However, our results also indicated that the CMAX date is delayed in the Bering Strait and along the coast of Alaska (Fig. 4c). This suggests that the timing of the annual maximum phytoplankton biomass does not fully depend on the sea ice retreat timing and the subsequent nutrient conditions in these areas. In the Bering Strait, the CMAX delay would be because the large nutrient supply continues even in summer as a result of the inflow of nutrient-rich Anadyr Water (Springer and McRoy, 1993, Springer et al., 1996). On the other hand, in coastal Alaska, optical contamination of the CDOM during the chla retrieval can also affect the CMAX values, especially in the area where the Alaskan Coastal Water (ACW) flows. Matsuoka et al. (2011) suggested that the ACW contributes to the high value of CDOM absorption ($a_y(\lambda)$), which can cause an overestimation of satellite chla values, because the CDOM absorbs light at some of the same wavelengths as chla (Matsuoka et al., 2007; Naik et al., 2013).

**RC8:** Page 12629, Line 5: Change "phytoplankton cell size composition" to "phytoplankton community size composition"

AR: Thank you for pointing out. I revised the sentence as below.

RS: Our results also suggest that the phytoplankton community size composition can change with the timing of the sea ice retreat in the shelf region of the Bering and Chukchi Seas during the MIZ bloom.

**RC9:** Page 12629, Line 7-9: Could be different food quality depending on phytoplankton size composition? If so, you have to mention reference. For example, since the food quality as food source could be related to phytoplankton size composition (reference), these results provide important information to help understand food quality. I think that the changes in phytoplankton species composition might be related to food quality change.

AR: Thank you for pointing out. It was wrong sentence. I would like to use "food quality" same as meaning of "efficiency of energy transfer".

RS: These results provide important information to help understand the efficiency of energy transfer for organisms that depend on the MIZ blooms as a food source.

Table R1. Slope and intercept between MODIS and in situ measured Rrs.

| $\lambda$ | 412 | 443 | 488 | 555 | 667 |
|---|---|---|---|---|---|
| Slope | 0.36185 | 0.33849 | 0.34321 | 0.42559 | 0.46137 |
| Intercept | 0.00106 | 0.00154 | 0.00163 | 0.00048 | 0.00019 |
| $r^2$ | 0.34 | 0.48 | 0.59 | 0.75 | 0.73 |
| N | 13 | 13 | 13 | 13 | 13 |

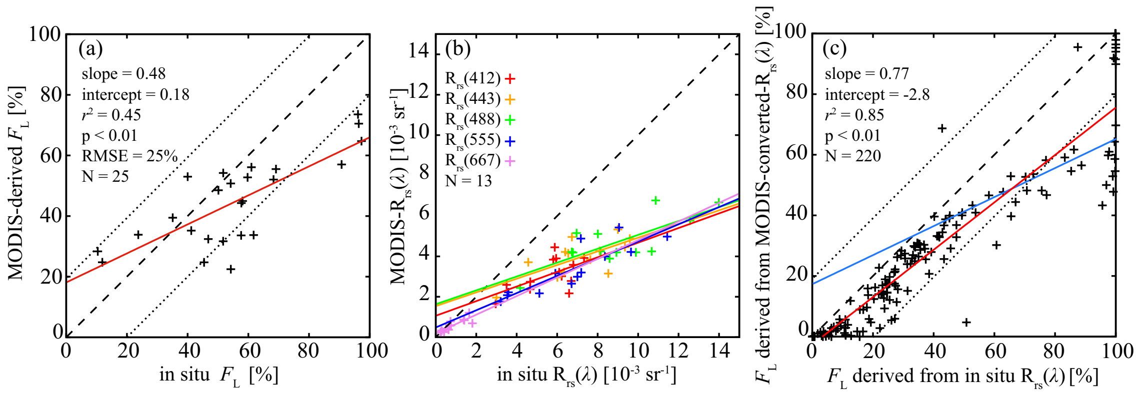


Figure R1. Scatter plots of (a) satellite-derived $F_L$ vs. in situ observed surface $F_L$ ($r^2 = 0.45$, $p < 0.01$, N = 25, RMSE = 25%), (b) satellite-derived $R_{rs}(\lambda)$ vs. in situ measured $R_{rs}(\lambda)$ (N = 13, slopes and intercepts for each $\lambda$ are listed in Table R1), and (c) $F_L$ derived from MODIS-converted-$R_{rs}(\lambda)$ vs. $F_L$ derived from in situ $R_{rs}(\lambda)$ ($r^2 = 0.85$, $p < 0.01$, N = 220). Dashed lines are 1:1 lines, doted lines indicate ±20% from 1:1 (Fig. R1a and c), and coloured solid lines represent regression line, respectively. The regression line of Fig. R1a is also shown in Fig. R1c with the blue solid line for comparison.

We thank the anonymous referee #3 kindly for careful reading our manuscript and for giving useful comments. Almost all suggestions were accepted and the manuscript has revised as below.

RC = Referee's comments; AR = Authors' Response (written in blue); RS=reconstructed sentences (written in green)

First of all we changed the manuscript title from "Influence of timing of sea ice retreat on phytoplankton size during marginal ice zone bloom period **in** the Chukchi and Bering shelves" to "Influence of timing of sea ice retreat on phytoplankton size during marginal ice zone bloom period **on** the Chukchi and Bering shelves".

**RC1**: Page 3, line 5. Food webs are simple, not short.

AR: Thank you for pointing out. I revised the sentence as below.

RS: Thus, the marine ecosystems in this region have been characterized by rather simple food webs and efficient energy transport to higher trophic levels through tight pelagic-benthic coupling (Grebmeier and Dunton, 2000).

**RC2**: Page 5, line 5. The statement that "Since phytoplankton grazers efficiently use the high phytoplankton biomass produced during bloom periods for their growth and production: : :" directly contradicts the statement on Page 2, line 23-24 that "The high primary production in the region is not completely consumed by the grazers in the water column due to low grazing pressure".

AR: Thank you for pointing out. The sentence was reconstructed as below.

RS: Because zooplankton strongly depend on the timing and magnitude of the spring bloom for their growth and production (e.g., Hunt et al., 2002, 2011; Søreide et al., 2011), the influence of the TSR on the phytoplankton size composition during the MIZ bloom period is crucial to evaluate the bottom-up effects of the primary production on the food web.

**RC3**: Page 8, lines 10-13. The statements that "…but ABPM retrieves optimal values of chla normalized productivity (PBopt) from aph(λ) instead of from SST and chla…" and "… because PPeu by ABPM is independently derived from temperature…" are inconsistent. The first indicates that PPeu is independent of temperature and the second states that PPeu is derived from SST. Both statements cannot be true.

AR: Thank you for pointing out. The sentence was reconstructed as below.

RS: The use of $a_{ph}(\lambda)$ is suitable to discuss the effect of ocean warming on $PP_{eu}$, because the $PP_{eu}$ value derived by ABPM is independent of the temperature (Hirawake et al. 2011).

**RC4**: Page 9, lines 17-19. The substitution of annual median PPeu for missing pixels seems dangerous considering that PPeu increases significantly over the course of the 16-year study. The authors should attempt to evaluate the consequences of this correction.

AR: Thank you for the comment. It is important to show the consistency of applying annual median to missing value. Firstly, we performed spatial and temporal interpolation to $P_{opt}$ data to increase valid pixel (Page 12620, line 4–7 BGD). After the interpolation, the fraction of the number of missing days to whole open water period was ~40% in maximum. We also evaluated the expected error about the substitution method. We chose two stations from the Chukchi and Bering shelves and retrieved times series of $PP_{eu}$ for 2007. Here, temporal integrated $PP_{eu}$ was defined as the "true" annual primary production. We conducted sensitivity tests that randomly replaced the time-series to the annual median of PPeu and calculated the simulated APP. The percentages of the number of replaced values were changed from 0 to 100%, and we evaluated how relative percentage error from true APP varies with the number of replaced pixels. The simulation was continued 10 times for each station. Figures R1A and B indicate the results of the simulation. Although the relative error increases with the number of replaced pixels, we found that the substitution of the annual median $PP_{eu}$ causes only for several percent error (~10% in maximum) to calculate the APP both for the Bering and Chukchi shelves. Therefore, we concluded that substitution of the annual median $PP_{eu}$ for missing pixels is effective way to derive the accurate APP. We revised the sentences describing the calculation method of the APP as below.

RS: The annual median $PP_{eu}$ was calculated for each pixel, and was substituted to accurately compute the APP when the $PP_{eu}$ was missing as a result of cloud cover. We conducted sensitivity tests and found that the substitution of the annual median $PP_{eu}$ to the missing values causes only a small error (~10% in maximum) to calculate the APP (data not shown).

**RC5**: Page 11, section 3.1. The authors describe the accuracy of FL estimated from satellites by comparing it to in situ values. They conclude that the performance is acceptable, but only focus on a few metrics. They never give the slope of the relationship, which differs greatly from the hoped-for 1.0. This metric should be added. Also, the satellite overestimates FL by 40% at high values, hardly a "slight overestimate." The authors should be more candid about the shortcomings of the algorithm. I was actually encouraged that there was a significant relationship between the satellite-derived and in situ FL. I wish the authors had tried to explain why the slope differed from 1.0, rather than just glossing over this point.

AR: Thank you very much for pointing out a very important point. We received very similar comment from the other referee. We provided the statistical information about how estimated $F_L$ differed from in situ $F_L$. In addition, we tried to explain why the slope differed from 1:1 line. I concluded that it is because of the underestimation of Rrs from MODIS. Fortunately, we were able to obtain the match-ups between in situ and MODIS measured Rrs during the cruise of 2013 (there is no Rrs measurement during the BASIS cruises). Figure R2b (same as Figure 2b in the revised manuscript) indicates the relationship between in situ- and MODIS-Rrs. Although there are linear relationships between the two Rrs, we found that MODIS underestimated Rrs for every band I compared (412, 443, 488, 555 and 667 nm). The slopes and intercepts between them are listed in Table R1 (same as Table 2 in revised manuscript). Using these relationships, we converted in situ Rrs to MODIS-Rrs using the dataset used in appendix B. Then, we compared the FL derived from original in situ Rrs and in situ Rrs converted to MODIS-Rrs (Figure R2b and Fig. 3b of revised manuscript). $F_L$ calculated from MODIS-converted Rrs also showed significant underestimation compared to $F_L$ calculated from in situ Rrs (the slopes ranged from ~0.33 to ~0.46), which is very similar to middle to high range of $F_L$ of Figure R2a. Although we understand that the number of dataset is not enough to illustrate the characteristic of MODIS-Rrs and the relationship may vary with seasons and regions, such underestimation of Rrs significantly caused the underestimation of $F_L$ via the calculation steps (i.e. QAA and SDM), at least during summer of 2013. The description about this point has added to result 3.1 as follows.

RS: The accuracy of the SDM-derived $F_L$ was evaluated by comparing the $F_L$ values from the in situ measurements and daily matched MODIS level-2 dataset. Twenty-five data points were available for this examination, which were collected over a wide area of the Bering and Chukchi Seas during different seasons (Fig. 1). Fig. 2a compares the satellite-derived $F_L$ and in situ $F_L$. The SDM successfully retrieved the $F_L$ values for 17 of the 25 data points (68% of the data) within a ±20 % $F_L$ range (Fig. 2a). The RMSE was 25%. The satellite validation was very similar to the results of Fujiwara et al. (2011), who showed that the $F_L$ derived using the in situ measured $R_{rs}(\lambda)$ had a 69% accuracy and an RMSE of 22.7%. However, it should be taken into account that there was a slight overestimation in the low $F_L$ range and relatively large underestimation in the high $F_L$ range (slope = 0.48 and intercept = 0.18). We also evaluated the performance of $R_{rs}(\lambda)$ retrieval of MODIS comparing with in situ-measured $R_{rs}(\lambda)$ at 13 data points during the GRENE cruise of 2013. Fig. 2b shows the comparison of in situ- and satellite-derived $R_{rs}(\lambda)$, and we found that the MODIS significantly underestimates $R_{rs}(\lambda)$ at every wavelength (the slopes ranged from to ~0.34 to ~0.46). The slopes and intercepts between the two $R_{rs}(\lambda)$ are listed in Table 2, which was used as the factors to convert in situ $R_{rs}(\lambda)$ to MODIS-$R_{rs}(\lambda)$. The conversion factors were applied to the $R_{rs}(\lambda)$ dataset used in appendix B, and then, we compared the SDM performance between in situ $R_{rs}(\lambda)$ and $R_{rs}(\lambda)$ converted to MODIS-$R_{rs}(\lambda)$ to assess how the retrieval errors in the MODIS-$R_{rs}(\lambda)$ affect estimation accuracy of the $F_L$. The relationship between the two derived $F_L$ is shown in Fig. 2c, and we found similar relationship with Fig. 2a. It can be said that the underestimations of $R_{rs}(\lambda)$ significantly caused the underestimation of $F_L$ especially in the middle to high range of $F_L$. Although we conducted the optimization of the QAA to derive more accurate input parameters of the SDM (appendix B), retrieval error in the $R_{rs}(\lambda)$ significantly affected the SDM accuracy in the study region. Nevertheless, the high determination coefficient indicated that the correlation was sufficient to address the spatio- and temporal variability of $F_L$. Here, we conclude that the SDM is applicable to satellite remote sensing in the Bering and Chukchi Seas, and is effectively optimized for what is known as optically complex water (Matsuoka et al. 2007, Naik et al. 2013).

**RC6:** Page 12, lines1-6. It would be useful here to have some vertical profiles of FL and PP to see how they vary with depth.

AR: I understand that the information of vertical profiles is important. I provided vertical profiles of $F_L$ and PP in Figure R3b (Fig. 3b and e in the revised manuscript), and described about them in result 3.1 and discussion section 4.1. Please sea the reconstructed sentences provided in RS of RC8.

**RC7:** Page 14, line16-17. The authors need to clarify that over most of the Arctic Ocean, omission of the SCM results in only a small error in PPeu (Ardynya et al. 2013, Arrigo et al. 2011).

AR: Thank you very much for the comment. I should cite the important papers. Please sea the reconstructed sentences provided in RS of RC8.

**RC8:** Page 14, lines 20-21. I don't understand the basis of their assumption that just because there is a good relationship between surface PP and PPeu, then surface FL can be used to infer FL throughout the water column. I see no reason why this should be true

AR: Thank you for the comment. I didn't intend to mean like the sentence. I separately described about the vertical profile of FL and PPeu in the revised manuscript. In addition, related to the comment from another reviewer, I changed the Figure 2 (Figure R3a and Fig. 3a in the revised manuscript) to comparison of surface FL with column integrated FL (defined as percent contribution of column integrated $chla_{>5\mu m}$ to column integrated $chla_{total}$). Reconstructed sentences (2nd paragraph of 3.1 and 4.1) are as bellow.

RS:

**Second paragraph of subsection 3.1**

To confirm how the satellite-derived $F_L$ and $PP_{eu}$ represent a water column's phytoplankton size structure and productivity, we compared the surface and vertically integrated $F_L$ values (calculated using Eq. 3 with water-column-integrated $chla_{total}$ and $chla_{>5\mu m}$) using an in situ dataset. The vertically integrated $F_L$ showed a significant relationship with the surface $F_L$ (Fig. 3c, $r^2 = 0.67$, $p < 0.01$) in spite of the presence of the subsurface $F_L$ maximum at the many of the sampling sites, especially in low and middle range of the surface $F_L$ were observed (Fig. 3b). Similarly, $PP_{eu}$ was also significantly correlated with the surface PP (Fig. 3c), and vertical distribution of PP also represents the large contribution of surface PP to $PP_{eu}$ (Fig. 3e). Although the surface values of both $F_L$ and PP explain the variation of water-column-integrated values, the depth of the maximum PP (6.1 ± 8.9 m) was significantly shallower than the depth of the maximum chla (20.1 ± 13.7 m) (Fig. 3d).

We also reconstructed subsection 4.1 as below

**4.1. Evaluation of performance of the satellite algorithms**

The validation of the satellite-derived $F_L$ showed a sufficient correlation with the in situ $F_L$, although the vertical distribution of the phytoplankton should be confirmed, because the SCM is commonly distributed in the study area and high Arctic Ocean (e.g. Hill et al. 2005, Ardyna et al., 2013). The omission of the SCM sometimes causes a large error in the satellite estimation of $PP_{eu}$ in the high Arctic (Hill et al., 2013). On the other hand, Arrigo et al. (2011b) and Ardyna et al. (2013) showed that the omission of the SCM can lead to only a small error in the $PP_{eu}$ retrieval over most of the Arctic Ocean. Our results supported the latter idea that surface PP explained the variation of $PP_{eu}$ well (Fig. 3c). Similarly, surface $F_L$ showed a very good relationship with $F_L$ retrieved from water-column-integrated chla in spite of the presence of subsurface $F_L$ maximum (Fig. 3a and b). A similar relationship was also reported between the surface and water-column-integrated chla in the Bering shelf (Lomas et al. (2012). Thereby, we suggest that the surface $F_L$ can also be reasonably used to predict the upper layer phytoplankton size structure. This might be due to the shallow bathymetry and SCM depth (~20 m) associated with the nitracline depth (Brown et al., in-press) compared to the higher Arctic Ocean where a deeper SCM (>40 m) is commonly found (Ardyna et al., 2013). Furthermore, the significantly shallower depth of the PP maximum compared to the depth of the SCM (Fig. 3d) causes the PP at the SCM to make a smaller contribution to $PP_{eu}$. This is also consistent with the results shown by Brown et al. (in-press) for the Chukchi shelf. Hence, we believe that ocean colour remote sensing is applicable to discuss the temporal and spatial relationships between the distribution of sea ice and phytoplankton variables (i.e. $F_L$, $PP_{eu}$, and chla), at least in the study area.

**RC9:** Page 16, lines 1-5. This statement is true mainly because the dates of CMAX and ice retreat vary by such a small amount in the northern Bering Sea.

AR: I'm afraid but I don't agree with this point. Because annual maximum values of chla often do not follow sea ice melt. Figure R4 indicate an example of chla seasonality of the Bering Strait area (168˚N, 170.2˚W). CMAX actually appeared in August or September and it did not depend on timing of sea ice retreat. Therefore, I would like to remain the sentences.

**RC10:** Page 16, lines 15-18. This statement is very vague and seems out of place in this paper. Add specifics or make its relevance more clear.

AR: Thank you very much and I agree with the comment. I reconstructed the sentence as below.

RS: Although the MIZ bloom timing is important for the growth and/or reproduction of zooplankton and more higher trophic organisms in the seasonally ice covered sea (e.g. Hunt et al., 2002, 2011, Leu et al., 2011), it is also crucial to comprehend how the phytoplankton community size structure during the MIZ bloom varies with the yearly change in the TSR, considering the energy use through predator-prey body size relationships

**RC11:** Page 17, lines 11-13. I don't understand the logic of this statement, assuming that nutrients are mixed uniformly throughout the water column on the Chukchi Shelf in winter. Surface nutrients will remain high until phytoplankton draw them down. This could happen early because of under-ice blooms or because of early ice retreat.

AR: Thank you for pointing it out. The sentence was reconstructed as follows.

RS: Hence, high surface-nutrient concentrations are expected to remain for a longer period after the sea ice melt in early-ice-retreat years compared to the late years, and conversely, the nutrients can immediately be consumed by large phytoplankton in late ice retreat years.

**RC12:** Page 17, lines19-22. Sufficient light can penetrate first year ice for phytoplankton to grow only if ice has melted and ponds have begun to form.

AR: Thank you very much for the comment. The sentence was modified as follows.

RS: Because thin first-year ice dominates in the shelf areas of the Chukchi and Bering Seas (Comiso et al., 2008), sufficient light can penetrate into the water column through a melt-pond or fragile ice when the solar radiation is strong enough during the melting season (Arrigo et al., 2012).

**RC13:** Page 18, line 7-8. This statement is true mainly because the dates of CMAX and ice retreat vary by such a small amount in the northern Bering Sea.

AR: As I described at RC9, I'm afraid but I would like to remain this sentence.

**RC14:** Page 19, line 19 and Page 20, lines 17-18. The idea that FL controls APP is probably wrong. Both of these variables are controlled by nutrient availability. More nutrients leads to both larger cells and higher production.

AR: Thank you very much for the comment and I agree with phytoplankton size changes related to the nutrient condition and after that primary production will change. I have already referred to nutrient condition and/or grazing pressure primarily control phytoplankton size (Page 12630, line 9–13 BGD), but in addition, I modified the sentence (Page 12630, line 24–27 BGD) as below.

RS: Because the Arctic Ocean is predicted to become a more ice-free ocean by many models (Perovich and Richter-Menge, 2009), it is suggested that the contribution of the nutrient conditions and the subsequent phytoplankton size structure to the APP can be larger in the future.

**Tables**

Table R1. Slope, intercept and $r^2$ between MODIS and in situ measured Rrs.

| $\lambda$ | 412 | 443 | 488 | 555 | 667 |
|---|---|---|---|---|---|
| Slope | 0.36185 | 0.33849 | 0.34321 | 0.42559 | 0.46137 |
| Intercept | 0.00106 | 0.00154 | 0.00163 | 0.00048 | 0.00019 |
| $r^2$ | 0.34 | 0.48 | 0.59 | 0.75 | 0.73 |
| N | 13 | 13 | 13 | 13 | 13 |

**Figures**

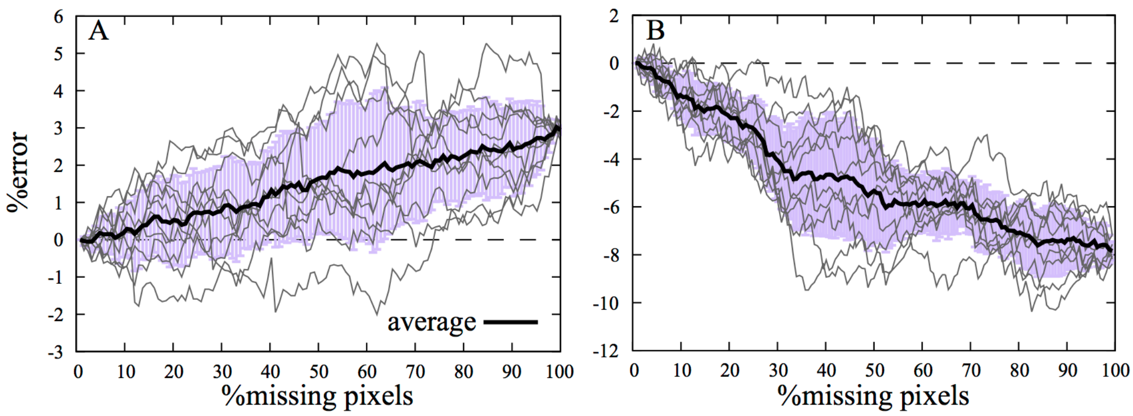


Figure R1. Relationships between the percentage of replaced pixels from original $PP_{eu}$ to annual median $PP_{eu}$ and percent error from "true" APP. Grey lines indicate the each simulation and thick black lines represent the average of the error. Error bars coloured in violet represent standard deviations.

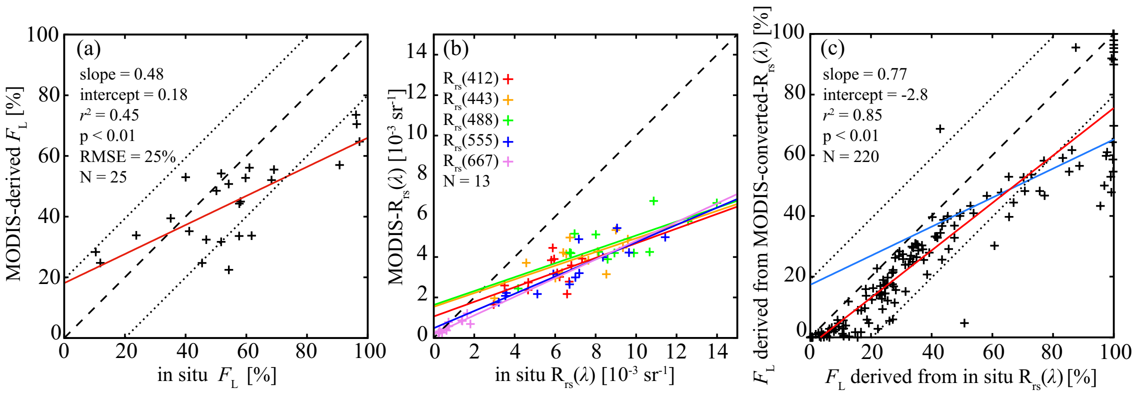


Figure R2. Scatter plots of (a) satellite-derived $F_L$ vs. in situ observed surface $F_L$ ($r^2 = 0.45$, $p < 0.01$, N = 25, RMSE = 25%), (b) satellite-derived $R_{rs}(\lambda)$ vs. in situ measured $R_{rs}(\lambda)$ (N = 13, slopes and intercepts for each $\lambda$ are listed in Table R1), and (c) $F_L$ derived from MODIS-converted-$R_{rs}(\lambda)$ vs. $F_L$ derived from in situ $R_{rs}(\lambda)$ ($r^2 = 0.85$, $p < 0.01$, N = 220). Dashed lines are 1:1 lines, doted lines indicate ±20% from 1:1 (Fig. R2a and c), and coloured solid lines represent regression line, respectively. The regression line of Fig. R2a is also shown in Fig. R2c with the blue solid line for comparison.

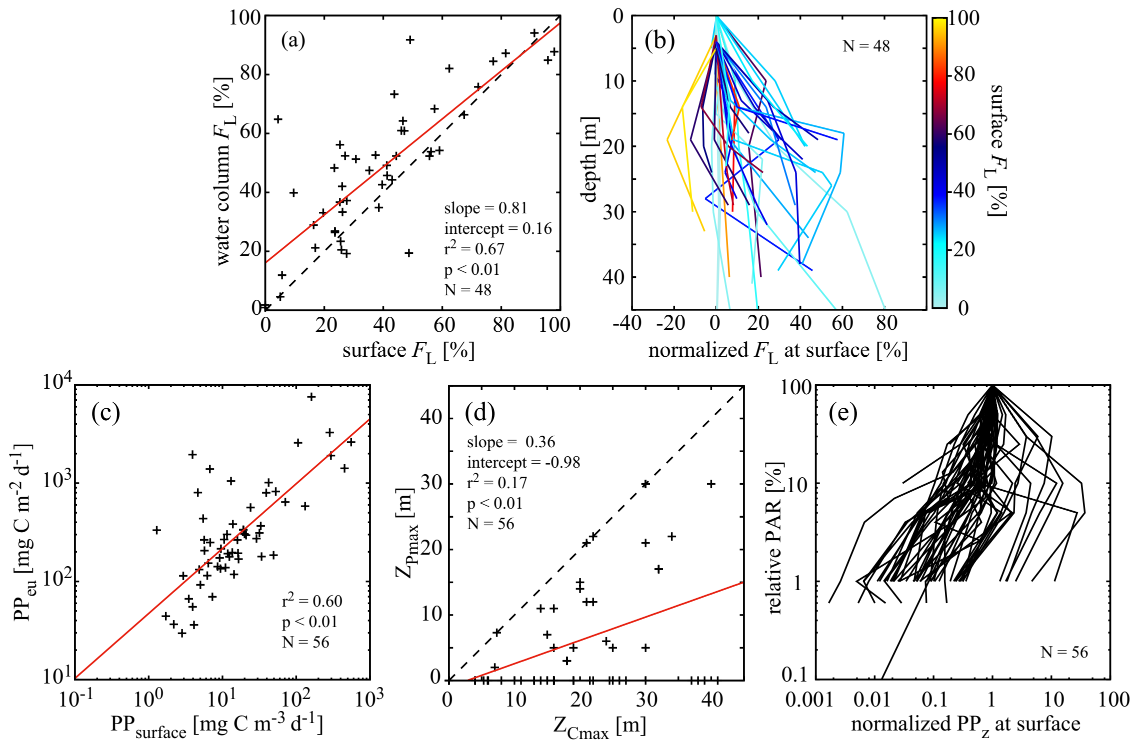


Fig. R3. (a) In situ water-column-integrated $F_L$ vs. surface $F_L$ ($r^2$ = 0.67, p < 0.01, N = 48), (b) vertical profiles of $F_L$ which are normalized at surface $F_L$, (c) in situ $PP_{eu}$ vs. in situ surface PP ($r^2$ = 0.50, p < 0.01, N = 56 in log scale), (d) depth of PP maximum ($Z_{Pmax}$) vs. depth of chla maximum ($Z_{Cmax}$), and (e) vertical profiles of PP which are normalized at surface PP, respectively. Note that vertical axis of Fig. 3e indicates %PAR relative to surface PAR value. Vertical profiles of $F_L$ are coloured by the surface $F_L$ value (Fig. R3b). Dashed lines indicate 1:1 line, and red solid lines represent regression lines, respectively.

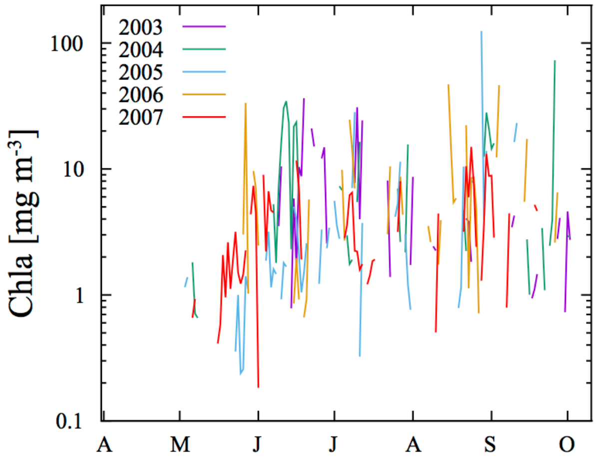


Figure R4.Time series (2003–2007) of chla for the point of Bering Strait (168˚N, 170.2˚W).