Towards a merged satellite and in situ fluorescence ocean chlorophyll product by Lavigne et al., present a new, exciting method for "calibration" of chlorophyll fluorescence profiles using remote sensing data. Method discussed here takes the correction approach presented by Boss et al [2008] a step further. With some modifications and really nice thought-through parameterization and validation; they show that this exciting new method that can be used globally; not only for gaining meaningful data from single chlorophyll profiles, but also to build climatological maps of chlorophyll concentration. This method is highly welcomed in this time when large datasets of chlorophyll fluorescence are collected by instruments deployed on autonomous vehicles, and calibration and intercalibration present an emerging issue.
I am of an opinion that Biogeosciences is an appropriate journal for this manuscript, and I do believe oceanographic and modeling community could benefit from the findings presented here. I am in favor of publication of this manuscript, however I would recommend that the authors address the comments and suggestions listed below because that might, to my opinion, improve the impact of the paper.

General comments:

Word *significant* (used in the paper numerous times) should be accompanied with some statistical method that is showing the significance of the comparison. For example, Figure 7, I can see that there are differences between method presented here and Boss et al [2008] method. How significant those differences are, I could see only if authors compare these two datasets, and use a specific statistical method (t-test or something similar) to show the significance of these differences.

I see that authors are using a ±30% cutout line when comparing their results with the remote sensing chlorophyll. Is that 30% a random number or it has something to do with the fact that remote sensing derived chlorophyll has a *30*% uncertainty level. If it does, it would be nice to state that somewhere in text; it makes your case stronger.

Specific comments

11901-11.
Total Chlorophyll-a is not a pigment. It is a pigment sum.

"...in <u>all autotrophic</u> marine organisms."
This is not true. What about chemolithoautotrophs? I presume you meant photoautotrophs. I would recommend re-writing this sentence. This is an opening an opening sentence and it should have a strong statement and lead reader into the story.

11901-16.
Word Total is un-needed. Total chlorophyll-a has been previously defined (line 11) as Chl-*a*, so no need for word total.

11901-18.
"...as with several other biological parameters,"

I would be free to say that most, if not all biological parameter measurements are scarce, and probably the most abundant biological oceanic measurement that is available is Chl-*a.*

11902-3

This whole paragraph is poorly written. Especially since one of techniques mentioned here (fluorometry) is a base for in-situ measurement of chlorophyll fluorescence. Maybe a mention of one of the numerous reviews and comparison between these three techniques, that would make the reader realize why is the HPLC technique important.

11902-8&9

"...which <u>additionally</u> provides the concentrations of a large spectrum of phytoplankton accessory pigments in <u>addition</u> to Chl-a."
Un-needed repetition.

11906-8. and Table 1

Increasing temporal and spatial resolutions does not significantly modify the similarity between the HPLC and satellite estimations. However, the number of match-ups strongly decreased.
As I read this table, it is true that increase in temporal resolution results with the decrease in number of matchups. However, I would not agree with authors that increase in spatial resolution (from 0.25 to 0.1) but keeping the same temporal resolution (8-days) causes the strong decrease number of matchups (80.5 to 77.5). This is a table that is showing nicely that Boss et al [2008] approach is lowering down the regression strength, however from the results presented in table 1, I really see no reason for not using 8 day/0.1 deg boxes. It would be interesting to see how would the overall results look like if the spatial resolution is increased, and how does the increase/decrease in size of the pixel impact the overall matchup performance.

11906 - 12.

"For the three stations, only the HPLC and fluorescence data available for the 1998–2007 period were retained (i.e. the period of activity of the SeaWiFS ocean color sensor)."

This should me mentioned in the beginning of the Data section.

11907-11 .. or fluorometric...

11908 -12. Non-photochemical quenching is a serious issue when it comes to fluorescence. I am of an opinion that it is something that is should be mentioned in the introduction to this paper. Authors here use correction proposed by Xing et al, however - this paper has not been published yet and I am not familiar with the details of this correction. I am, on the other hand familiar with the work of Sackman et al [2008], and one developed for North Atlantic Bloom Experiment 2008 (see http://data.bco-dmo.org/NAB08/Chlorophyll_Calibration-NAB08.pdf) where auxiliary data are used to make this correction successful. Authors are stating that highest value encountered within the mixed layer was used as a reference point and for extraction to the surface. There are several reasons why such a simple correction is making me uncomfortable. Primarily, it is based on assumption that mixed layer is constantly mixed - vertical distribution of the phytoplankton in the mixed layers can be non-uniform and variable (from Sackman et al [2008] paper and personal experience). What if the increasing vertical pattern of the chlorophyll fluorescence (as a function of depth) is a reflection of the either changing abundance of the phytoplankton, changing community structure or just simple acclimation to the lower irradiance levels? How does one distinguish between "quenched" profiles and "non-quenched" profiles when one does not use other measurements to ground truth it? I have done a fast calculation using your approach and came up with up to 8% error in MLD integrated chlorophyll for highly mixed open ocean waters.

I am aware that this is not the paper focused on chlorophyll quenching correction, but it is an important part of the story and I would like to see more elaboration on methods, potential errors, etc.

11910-7 What kind of linear regression was used here? Type I or Type II. That would be nice to see. Also - for the sake of significance of these results - I would recommend

stating p-values of each of these linear correlations, not only when you need it to make a point (like in line 26 on page 11916)

11910-24 here authors are stating that scattering of the data around the 1:1 line is relatively homogeneous, and later in discussion (11914 - line 6) they state that dispersion of datapoint under 0.05 mg m^{-3}. Maybe mentioning here that scattering was relatively homogenous on values higher than 0.05 mg m^{-3}.

11910-22 Paragraph starting here, discussing on the results presented in the Table 4. It seems to me that DYFAMED is an outlier, when compared with the rest of the dataset (HOTS and BATS). Although this paper is focusing mostly on performance of the correction approach, it would be interesting to see some "real" discussion here. For example, authors are stating that r^2 is highest in DYFAMED since largest range of chlorophyll was observed. But - other performance measurements (MPD and SIQR) are much worse for DYFAMED than for HOTS and BATS. Is it maybe because DYFAMED is seeing largest phytoplankton biodiversity, therefore there are some other drivers of Chl F variability that could drive this error higher? It would be interested to say something later in discussion.

11911 - 8-18. This paragraph doesn't read clearly. I suggest re-writing since the main points are getting lost.

11914-4 to 7 First sentences - I don't understand - please re-write. Second sentence, maybe cite figure 1, just so reader can follow it clearly.

11915-20 As stated in general comments - if authors use word significant, make sure that statistics are present here - fact that average difference was 0.15 mg m^{-3} is an interesting finding but does not have any statistical significance.

11916-3. I really don't understand what is more relevant within certain localized areas? I suggest re-writing this sentence to make sense to the reader.

Table 1 needs a number (n) of HPLC/sat matchups.

Table 3, I am not sure if this is really needed here since it is a duplicate of the data shown in [*Uitz et al.*, 2006]

Figure 1. - panel d - keep the limits of the x and y axis the same in all the panels. Log-log space is not an easy format to think in, keeping all the axis same will allow reader to explore results easier.

References:

Boss, E., D. Swift, L. Taylor, P. Brickley, R. Zaneveld, S. Riser, M. J. Perry, and P. G. Strutton (2008), Observations of pigment and particle distributions in the western North Atlantic from an autonomous float and ocean color satellite, *Limnology and Oceanography*, *53*(5), 2112-2122.

Sackmann, B. S., M. J. Perry, and C. C. Eriksen (2008), Seaglider observations of variability in daytime fluorescence quenching of chlorophyll-a in Northeastern Pacific coastal waters, *Biogeosciences Discuss.*, *5*(4), 2839-2865.

Uitz, J., H. Claustre, A. Morel, and S. B. Hooker (2006), Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll, *J. Geophys. Res.*, *111*(C8), C08005.