Dear reviewer,

We appreciate your comments. In the following text I address your considerations.

Referee comments are reproduced here, in italic, to facilitate reading.

Anonymous Referee #1 Summary

The topic of this manuscript is the response of the oceanic plankton to inputs of iron and other nutrients from the deposition of airborne dust. The topic itself is interesting and relevant. While Patagonia is a small dust source area compared to the large desert regions mainly located in the northern hemisphere, small changes in dust deposition may make large impacts. The authors claim that they find a response in plankton concentration to changes in Patagonian dust emissions, indicating an important role of airborne dust deposition. This finding is derived from satellite observations of different parameters (NDVI, AAI, Chlorophyll, phytoplankton carbon, …) which are correlated with each other for a time period from 2003 to 2010. Unfortunately the conclusions are based on many assumptions without direct evidence. Therefore I cannot recommend it for publication in the current form.

Author: I will present here further evidences in favor of our assumptions.

Specific topics:

1) Anonymous Referee #1

- The methods are not made entirely clear in the manuscript. How were the correlations computed? Were they just based on annual means, monthly or daily fields? Did I overlook some information on this? Seasonal correlations may provide additional information.

Author: Although the text sets the spatial and temporal resolution of the analysis (annual), it does not present which correlation index was used for the biological proxies. We used Spearman's ρ for all correlations in the article. This information will be added in the revised text.

Different spatial or temporal scales could be used to add information, as the same or different relations may exist over different scales, but we believe that the information on the proposed scales is sufficient in the scope of our analysis.

2) Anonymous Referee #1

- 8 Years is not a great basis to provide statistically significant information from correlations. Please provide information about the statistical significance of the correlations of dust and plankton signals.

Author: There is no other data set with longer span that could be used. Data sets that discriminate iron bearing aerosol as the AAI were available for the Coastal Zone Color Sensor (CZCS) era, which, unfortunately for this study, was focused on the northern hemisphere coastal areas. In the modern era (1996 onwards), Total Ozone Mapping Spectrometer (TOMS) data is available until 2005; it covers a period of the same size (in rounded years), but It includes the strong ENSO transition of 1997/1998 that could affect the analysis. The Ozone Monitoring Instrument (OMI) onboard Aura began in late 2004, and as it is still functional, it could represent a longer time span in the future. However, as the data for 2011 and 2012 cannot be used due to the strong Puyehue eruption (June/2011) – which covered a substantial part of Patagonia with thick ash deposits that have been remobilized latter by wind erosion – the OMI useful data for this region would still have less than eight complete years. Therefore we decided to use SCIAMACHY data and it is still the longer suited data for this region.

As for the statistical significance, there are several important discussions on the overuse of the statistical hypothesis testing. A particular point in this discussion that is valid for the matter at hands is that of its arbitrariness (Cohen, 1994; Johnson, 1999; Nester, 1996). By specifying a significance level (e.g., the recommended 5% of the Neyman-Pearson lemma) the magnitude of the effect for a statistically significant result is set by the number of observations (in fact, the degrees of freedom) and by the variance – for a large sample size and/or a low variance even the smallest effect size would be statistically significant; while in the opposite case, a very high effect size would be required. This "automatic" procedure gives the false sense of objectivity, especially because the sample size and sampling procedure for this type of environmental study are not set by the researcher, which uses whenever data is available, nor it is required that he justify or sets the minimum biologically or environmentally significant effect size; therefore such a use of the hypothesis test does not add meaningful insight, especially in the case of point null hypothesis (e.g., effect size equal to zero). That is why, within the framework of classical statistics, many have argued for the use of the confidence interval centered on the observed values or simply the variance, as it shows explicitly the uncertainty of the estimation in contrast to merely presenting a dichotomous, arbitrary result.

Another important question is of the spatial structure of the result, which is very coherent as discussed in the article. This somewhat coherent structure is expected as a result of spatial correlation emerging from large scale processes that determine the biological variability. In our view, what is important in this case than is not peculiarly if this or that pixel is statistically significant, but if there is evidence of correlation with some area. This context again reduce the usefulness of a hypothesis test, for which the interpretation is that the observed values would or not be less than 5% frequent in a long run of the same data collection setup if the null hypothesis were true. It is important to bear in mind that this line of reasoning evolved from an earlier (and somewhat different) perception by Fisher as a method to prevent against random sampling fluctuations. To quote one of his passages in Fisher (1929): “Their function is to prevent us being deceived by accidental occurrences, due not to the causes we wish to study, or are trying to detect, but due to a combination of many other circumstances which we cannot control.” In synthesis, it would be a method to dismiss any data for which chance alone could not be disregarded as the primary generating process. Therefore, if large areas of the ocean are spatially correlated due to the scales of the processes that determine their properties what is the interpretation if all pixels in an area have the same type of response, but some “fail” hypothesis test and others don’t? I would suggest that the results for all the area should receive the same amount of support (or belief), but for all the considerations listed above I would rather argue against its computation.

Please note that we have been coherent about our statistical approach, as we only used the critical values for the spearman correlation to apply a spatial constrain in the source areas definition, noting that it was an arbitrary decision and result (in terms of spatial extent). This was done in hoping to aid comparison with model results.

Therefore, to be more transparent about the uncertainty associated with the analysis, we suggest adding the lower and upper limits for the confidence interval for each of the images in the article in a supplementary resource. Readers interested only in the statistical significance can then look for regions were the lower limit of the positive correlation areas does not include the zero, which is approximately the same.

As the Fisher transformation can only be applied to correlation coefficients if the variables could be assumed to come from a bivariate normal distribution, confidence intervals will be calculated with bootstrap. The bootstrap is a numeric procedure and therefore presents statistical fluctuations about its nominal confidence level. As a general rule, the fluctuations are lower for higher bootstrap samples and lower confidence levels. Confidence bounds of the 90% confidence interval will be calculated by the bias-corrected and accelerated (BCa) bootstrap method with 20.000 replications using functions from the R package “boot”. We are performing algorithm checks and optimizations and these images will be uploaded in a second response as soon as their calculations are completed.

3.a) Anonymous Referee #1

Some assumptions that were insufficiently supported by the information given in the manuscript:

(A) Dust microphysical properties and transport altitude exhibit lower variability than dust optical thickness;

Author: The specific question is what process has greater influence on the AAI interannual variation. Our associative analysis would only be valid if the main cause for the interannual variability in the AAI is the dust load in the atmosphere. In lack of external information, the suppositions needed are that aerosol properties and layer heights are the approximately the same for the same year period.

The fact that we are working with averages over large spatial and temporal scales provide some logical support to this assumption, as the sources are the same (mineralogy) and also broad meteorological patterns. But also because that we are extracting AAI data exactly over the defined source regions; the dust properties and layer height variation should be more affected by environmental variations with distance from source. These suppositions are found as conclusions from observations by Hsu et al. (1999) in comparison of AAI with sun-photometer data.

As cited in the manuscript, Chiapello et al., (1999) show a strong correlation on dust concentration and AAI. We can add to this the study of Deroubaix et al. (2013) who found a better correlation between dust load and AAI for low transport altitude, characteristic of Patagonian dust (Li et al., 2008). And also the study of Hsu et al. (1999), who found linear relation of AAI with sun-photometer data.

3.b) Anonymous Referee #1

Some assumptions that were insufficiently supported by the information given in the manuscript:

(B) Coherence between dust sources and deposition;

Author: It was not clear what was meant by "coherence between dust sources and deposition". But I will answer assuming that is related to the nature of the analysis, i.e., correlation of emission, not deposition, with biological response.

The life time of dust in the atmosphere is of a few days, and much of the coarse fraction will deposit over land or near shore (e.g., Johnson et al., 2011). Li et al. (2010) have shown that less than 13% of the trajectories from Patagonia reach Antarctica continent in less than 10 days which means that the majority of the dust transported is deposited over the Atlantic sector of the Southern Ocean (Li et al., 2008). It also should not be expected that the correlation patterns resembled that of estimated deposition from several models, mainly because ocean currents can redistribute the dust load and because different regions may have different sensibilities to dust additions (Erickson et al., 2003).

3.c) Anonymous Referee #1

Some assumptions that were insufficiently supported by the information given in the manuscript:

(C) Ratio between dust load and oceanic biomass low enough such that the interannual signal is not disturbed.

Author: Unfortunately no research team have so far reported a Patagonian dust storm over the ocean, only background concentrations. This renders it difficult to access the dust concentration over those exceptional conditions were it is more likely to occur the so called artifact effect. Considering the calculations of Wozniak and Stramski (2004) dust concentrations above 0.1 g m^{-3} could produce a measurable artifact effect for the average chlorophyll concentration of 0.5 mg m^{-3}. A dust flux greater than 5 g m^{-2} would be required to achieve this concentration in a 50 m mixed layer depth. In contrast, we discuss in the article that a flux of only 0.4 g m^{-2} would be necessary to promote biological response.

It should also be taken in account the large longitudinal span of the positive correlation, covering areas very distant from source, and would not be reasonable to assume such a high dust flux or in water concentration at that distance from a small source. In comparison, a dust flux of ~1 g m^{-2} was measured by Herut et al. (2005) in the Mediterranean fallowing a large emission event from the Sahara.

There is a large uncertainty in models to use them with confidence to settle an estimated flux, what is expected due to the scarcity of measurements and the lack of dust storm observations. Nevertheless, two recent estimates offer evidence in favor of our assumption. Li et al. (2008) estimate an average deposition of ~0.3 to 3 g m^{-2} $year^{-1}$ for the region of positive correlation, a value much lower than 5 g m^{-2} in a single occasion. Estimates of single dust storm events are provided by Johnson et al. (2011). Inverting their formulations is possible to calculate values in the range of ~0.2 to 2 g m^{-2} of dust flux, also less than twice lower than would be required for a measurable artifact effect.

4) Anonymous Referee #1

- The point that atmospheric dust load does not interfere with the retrievals of oceanic biomass (chlorophyll-a, diatoms, phytoplankton carbon) is crucial for the credibility of the results, but not sufficiently supported in the manuscript, apart from stating some references (section 2.2). Which are the ratios of the relative concentrations that would be low enough so that the signals do not interfere with each other? What concentrations are present in the area, are they below that ratio? In particular when looking at the interannual variability, the high dust years may disturb the phytoplankton signal, which would make the presented results meaningless. Please provide further evidence that this is not the case. The comparison of the correlation with the chlorophyll concentration is not conclusive – as far as I understand the total ocean area below 40 deg S was used for this analysis, however the correlation maps show that there appear to be different controls in different parts of the ocean. To be conclusive only the values in the region with the significant positive correlations should be taken into account in figure 4.

Author: These subjects have been addressed in comments 2 and 3.c above.

5) Anonymous Referee #1

- A possible misinterpretation of biomass burning aerosol as dust aerosol was not discussed, but could potentially greatly impact the results. As far as I see by excluding areas with NDVI>04 the authors claim to reduce the biomass burning impact, however biomass burning smoke is frequently transported over great distances. One approach may be to use information from the MODIS 'fine mode' optical thickness product to exclude biomass burning influence.

Author: To have an influence on our estimates, biomass burning plumes would have to pass over the defined source regions and have a significant optical effect on the AAI retrieval (similar or higher than dust influence), which is highly unlikely due to the semi-desert condition. Nevertheless, it could be that the interannual variability in dust and biomass burning emissions from Patagonia are correlated – that is, even if the AAI signal is dominated by dust, any fertilization effect observed could have contribution of biomass burning if it has similar temporal pattern. In that case, the magnitude of the contribution would be hard to separate. This is especially so because validations of the MODIS fine mode fraction algorithm <u>overland</u> reveled a very poor performance (Anderson et al., 2005; Remer et al., 2005). For instance, mixed aerosol type criterion is almost never met, resulting in values of fine fraction that are either 0 or 1 (Remer et al., 2005). It also present a positive bias and it incorporates some coarse fraction aerosol types - and therefore this "fine fraction" could be better defined as "nondust fraction" (Anderson et al., 2005). But again, this possible confounding effect would only occur if the biomass burning interannual variability is correlated with the dust emission.

To evaluate the above effects that could affect our analysis (impact on the AAI retrieval and/or correlation with dust interannual variability), we provide counts of biomass burning spots identified in the World Fire Atlas algorithm 1 from the Along Track Scanning Radiometer (Arino et al., 2012). Fire spots were aggregated in grid boxes equivalent to the SCIAMACHY spatial resolution. A map of sum of events (2003 - 2010) is provided in Figure 1 and interannual time series of fire spot counts inside source areas and for all Patagonia (> 40ºS) are provided in Figure 2. The spearman correlations with the AAI data and respective 90% confidence intervals (BCa for 20.000 bootstraps) are -0.107 (-0.848 to 0.899) for fire counts over source areas and -0.309 (-0.854 to 0.538) for fire counts over all Patagonia. There is a high uncertainty in the estimates but they don't support a biomass burning influence on the correlation of AAI with the biological proxies.

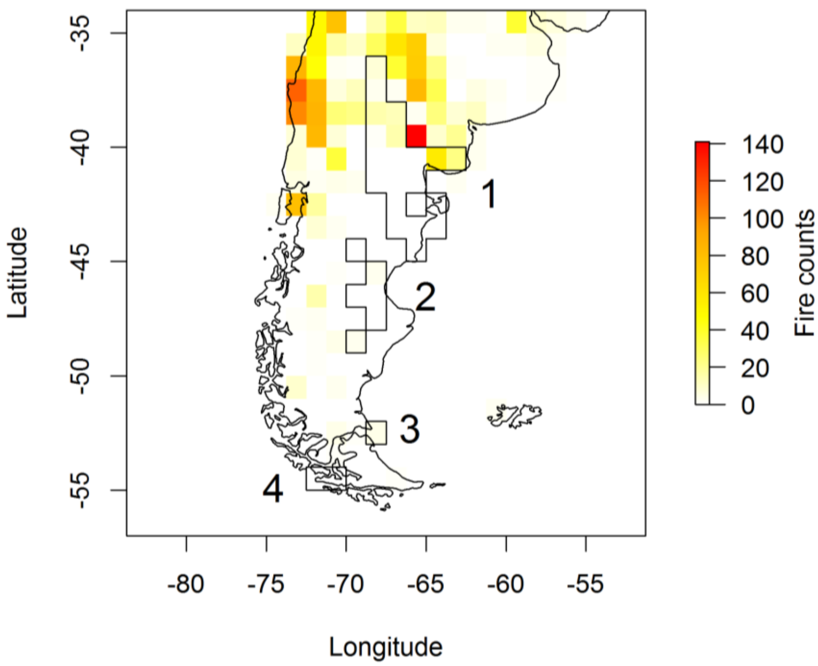

Figure 1 – Distribution of fire count sums from 2003 to 2010.

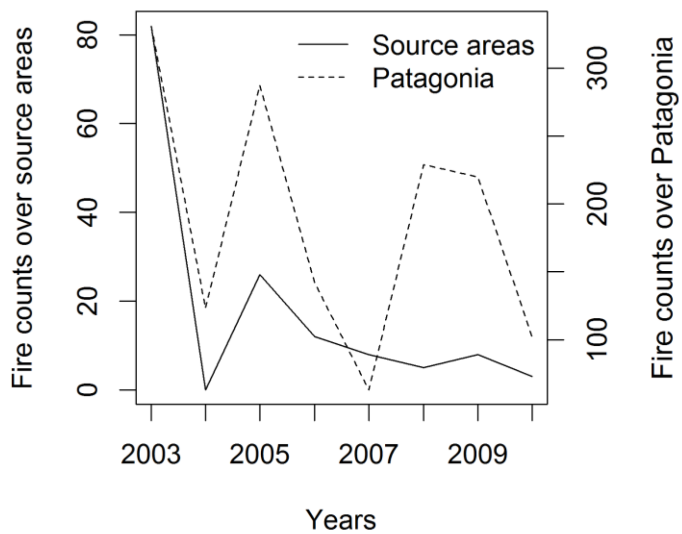

Figure 2 – Annual sum of fire counts over source areas and over Patagonia.

6) Anonymous Referee #1

- *Section 3.1: That area 4 is a misrepresented as dust source casts doubt on the method to determine dust sources.*

Author: We could not find a reasonable explanation for the area 4 to be identified as a source region by the method employed. Nevertheless, as cited in the article, with the exception of the area 4, the source areas are in excellent agreement with other estimates and reports. Therefore the results would be approximately the same using a previously published estimate of source areas. Also, we reinforce that due to small extent, the addition or exclusion of the area 4 does not result in any noticeable difference in the results.

7) Anonymous Referee #1

- *What is the motivation of using '10% trimmed AAI' as individual strong events may also induce strong ocean deposition events and plankton growth? If you are after some background signal, you should use the median values instead.*

Authors: As explained in the article, our motivation for the trimmed mean was to use a robust estimator as the simple mean is very sensitive to even one single event. As the study focused on annual variations, a large effect on the simple mean caused by only a single large event could bias the resulting time series and disconnect it from the biological series. Nevertheless, the results are approximately the same using simple mean.

8) Anonymous Referee #1

- *Minor remark – some language improvement would be needed.*

Author: The text will be revised to improve understanding.

References

Anderson, T. L., Wu, Y., Chu, D. A., Schmid, B., Redemann, J., and Dubovik, O.: Testing the MODIS satellite retrieval of aerosol fine-mode fraction, Journal of Geophysical Research, 110(D18), D18204, doi:10.1029/2005JD005978, 2005.

Arino, O., Casadio, S., and Serpe, D.: Global night-time fire season timing and fire count trends using the ATSR instrument series, Remote Sensing of Environment, 116, 226–238, doi:10.1016/j.rse.2011.05.025, 2012.

Chiapello, I., Prospero, J. M., Herman, J. R., and Hsu, N. C.: Detection of mineral dust over the North Atlantic Ocean and Africa with the Nimbus 7 TOMS, Journal of Geophysical Research, 104(D8), 9277–9291, doi:10.1029/1998JD200083, 1999.

Cohen, J.: The Earth is Round ($p < .05$), American Psychologist, 49(12), 997–1003, doi: 10.1037/0003-066X.49.12.997, 1994.

Deroubaix, A., Martiny, N., Chiapello, I., and Marticorena, B.: Suitability of OMI aerosol index to reflect mineral dust surface conditions: Preliminary application for studying the link with meningitis epidemics in the Sahel, Remote Sensing of Environment, 133(0), 116–127, doi:10.1016/j.rse.2013.02.009, 2013.

Erickson, D. J., Hernandez, J. L., Ginoux, P., Gregg, W. W., McClain, C., and Christian, J.: Atmospheric iron delivery and surface ocean biological activity in the Southern Ocean and Patagonian region, Geophysical Research Letters, 30(12), 1–4, doi:10.1029/2003GL017241, 2003.

Fisher, R. A.: The statistical method in psychical research, Proceedings of the Society for Psychical Research, 39, 185-189, 1929.

Herut, B., Zohary, T., Krom, M. D., Mantoura, R. F. C., Pitta, P., Psarra, S., Rassoulzadegan, F., Tanaka, T., and Thingstad, T. F.: Response of East Mediterranean surface water to Saharan dust: On-board microcosm experiment and field observations, Deep Sea Research Part II: Topical Studies in Oceanography, 52(22-23), 3024–3040, doi:10.1016/j.dsr2.2005.09.003, 2005.

Hsu, N. C., Herman, J. R., Torres, O., Holben, B. N., Tanré, D., Eck, T. F., Smirnov, A., Chatenet, B., and Lavenu, F.: Comparisons of the TOMS aerosol index with Sun-photometer aerosol optical thickness: Results and applications, J. Geophys. Res., 104(D6), 6269–6279, doi:10.1029/1998JD200086, 1999.

Johnson, D. H.: The insignificance of statistical significance testing, The Journal of Wildlife Management, 63(3), 763–772, doi: 10.2307/3802789, 1999.

Johnson, M. S., Meskhidze, N., Kiliyanpilakkil, V. P., and Gassó, S.: Understanding the transport of Patagonian dust and its influence on marine biological activity in the South Atlantic Ocean, Atmospheric Chemistry and Physics, 11(6), 2487–2502, doi:10.5194/acp-11-2487-2011, 2011.

Li, F., Ginoux, P., and Ramaswamy, V.: Distribution, transport, and deposition of mineral dust in the Southern Ocean and Antarctica: Contribution of major sources, Journal of Geophysical Research, 113(D10), D10207, doi:10.1029/2007JD009190, 2008.

Li, F., Ginoux, P., and Ramaswamy, V.: Transport of Patagonian dust to Antarctica, Journal of Geophysical Research, 115(D18), 1–9, doi:10.1029/2009JD012356, 2010.

Nester, M.: An applied statistician's creed, Applied Statistics, 45(4), 401–410, doi: 10.2307/2986064, 1996.

Remer, L. A., Kaufman, Y. J., Tanré, D., Mattoo, S., Chu, D. A., Martins, J. V, Li, R.-R., Ichoku, C., Levy, R. C., Kleidman, R. G., Eck, T. F., Vermote, E., and Holben, B. N.: The MODIS Aerosol Algorithm, Products, and Validation, Journal of the Atmospheric Sciences, 62(4), 947–973, doi:10.1175/JAS3385.1, 2005.

Wozniak, S. B., and Stramski, D.: Modeling the Optical Properties of Mineral Particles Suspended in Seawater and their Influence on Ocean Reflectance and Chlorophyll

Estimation from Remote Sensing Algorithms, Applied Optics, 43(17), 3489, doi:10.1364/AO.43.003489, 2004.