

Interactive comment on “On the trophic regimes of the Mediterranean Sea: a satellite analysis” by F. D’Ortenzio and M. Ribera d’Alcalà

F. D’Ortenzio and M. Ribera d’Alcalà

Received and published: 28 November 2008

First of all we would like to thank Emmanuel Boss (EB) and Emmanuel Devred (ED) for their constructive and encouraging comments. After carefully reading them, we concluded that their main suggestions to improve the paper and overcome the present weaknesses, were:

1. to improve the statistical robustness of the analysis through additional, appropriate tests;
2. to integrate remote sensing chlorophyll data with other satellite derivable parameters (e.g., bb, temperature, etc.), in situ data and bathymetry;
3. to opt for a log-normalization of data in place of normalization by maxima.

They also questioned the hypothesis that remotely sensed chlorophyll a distributions,

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



as proxy for phytoplankton biomass, reflect univocally the trophic regime of a certain region. Later in the text we will describe how we improved our analysis, as solicited. In the following we discuss the last point raised in different forms by both of them.

We completely agree that marine food webs are complex, that they generate non linear interactions among components and that emergent patterns might be hard to predict by simple assumptions. This certainly holds true if community structure is the sought unknown. There is a wide consensus that community structure is tightly coupled with a trophic regime. In other words, any changes in the structure will be reflected (or reflect) a change in regime. We believe that this view is substantially true. And we are aware that, even integrating more advanced analyses on remote sensed data, we might refine the information on phytoplankton communities (this is an important task to be addressed in another study, which, by the way, is in progress), but we cannot constrain the community structure. This was beyond the scope of our analysis. Our starting point was the evidence of a wide range of values in phytoplankton standing stocks and steep geographical gradients in the basin. These features had been already highlighted by other studies (see references in the paper), which attributed such variations mostly to the seasonal cycle of atmospheric physical forcing and coastal inputs, with a semi-quantitative subdivision of the basin in areas with different annual mean values of chlorophyll concentration. The possible separation in bioprovinces and the implications on the annual cycle of production were left aside. This is why we addressed the two questions implicitly posed in the study: do gradients in biomass in the tiny Med reflect different regimes, i.e., discontinuities? Is there a quantitative method to assess if it is so, despite the poor information available on in situ processes in most of the basin? Both aspects are, in our view, the backbone of Longhurst's analysis.

The Höevmoller diagrams provide a more intuitive way of gradients and trends than a suite of maps of annual or seasonal averages of chlorophyll concentrations and, in our case, extend to a larger time interval previous analyses. What we believe provide new information on the basin are:

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

1. the statistically robust differences in the seasonal cycle of phytoplankton biomass in different areas;
2. the strong homogeneity of areas with similar seasonal cycle and their weaker, than previously assumed, dependence on latitudinal constraints, especially for areas not connected;
3. the different seasonal cycles displayed by different parts of the basin.

It is worth noting that our regionalization displays strong similarities with historical and recent biogeographical classifications of the basin (see Bianchi C.N., 2007, Biodiversity issues for the forthcoming tropical Mediterranean Sea. *Hydrobiologia*, 580: 7-21 and references therein). We then infer that the seasonal cycle of phytoplankton biomass is tightly coupled with the structure of food web and, in turn, also with the dynamic range of biomass itself. This does not provide, per se, information for mechanistic reconstructions of the dynamics or on the structure of communities, but certainly provides effective directions on where to focus observations, data rescue and analyses.

To make this more explicit in the paper, we added a paragraph in the "Discussion and Conclusions" (page 2972, line 3):

"This point is not trivial and it was not expected in advance. The geographical distribution of the clusters, determined by the seasonal cycle of phytoplankton biomass, is tightly coupled with the dynamic range of biomass itself, as obtained, for example, with a 10 years climatological mean (i.e. Figure 1). Oligotrophic regions, showing very low mean values of chlorophyll concentration, match exactly cluster # 1, 2, 3, whereas, productive regions have seasonal cycles relevantly different (cluster # 4, 5). In other terms, at least in the Mediterranean, accumulations of phytoplankton are observed only where a specific temporal trend is present."

And also in the "Abstract":

"The geographical correspondence between specific clusters and regions showing high

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

values of mean chlorophyll concentration indicates that, at least in the Mediterranean Sea, accumulations of phytoplankton are observed only where specific temporal trends are present."

All the suggestions for a more in depth analysis of the mechanisms were very welcome, because they encouraged what we were already carrying out as a follow-up of the analysis we submitted. We agree that some of the inferences are speculative and/or based on the sparse biological information available to date. We then made more explicit that they are working hypotheses, which could be tested in the next future either with existing data, as suggested by EB, or by focused sampling. Finally, we believe that the promising results of our approach for the Med might be extended to the global ocean, possibly refining the Longhurst synthesis and complementing what ED and colleagues have already done for the Northwestern Atlantic ocean.

Points raised by EB.

1. Chlorophyll is one variable. It only contains a limited amount of information about the ecosystem, and in addition, is not an ideal tracer of phytoplankton biomass as it suffers from physiological variability (e.g. change in chl/cell). Using additional remotely sensed data (e.g. Temperature, CDOM, bb) and in-situ/model data (e.g. mixed layer depth, zooplankton) a better description of the state of the ecosystem is most likely possible.

This comment raises an important point. As anticipated above, the analysis presented in the paper used two approaches. A more "classical" Hoevmoeller diagram, which is often used to determine spatio-temporal variability on satellite data, and a more "innovative" approach, based on a cluster analysis applied to temporal series. Cluster analysis was often used in satellite data processing, although, to the best of our knowledge, it has never been applied to determine similarities between seasonal courses of chlorophyll concentration. As EB pointed out, integration with other remote sensed proxies would definitely improve the results. While this analysis is already ongoing, the

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



amount of time required to finalize it is significant. We rather prefer to get prompt feedbacks from the community about our approach and conclusions, before proceeding to a more integrative analysis. In addition, other ecological data are available for different regions of the Mediterranean Sea, but a large part of them do not have the required spatio-temporal resolution to be compared with satellite observations. The regionalization proposed in the paper will improve the utilization of these data, as the existing observations could be analyzed in a more rational way, for example averaging the in situ data available in the same cluster/ecoregion. We see our contribution as a starting point for additional analysis, which will comprise, along with suggestions received, all the available data.

2. Temporal changes in chlorophyll are the result of many processes (e.g. ML dynamics, grazing, growth, physiological adaptation, species composition changes). Their interpretation without additional data is bound to be speculative at best.

EB is, of course, right (see point 1). But, for the clusters # 1, 2 and 3 we referred to the ML climatology of D'Ortenzio et al. 2005, which partially explain the observed dynamics. For the other clusters, we did not highlight that our comments were mostly inferences. In addition, to better distinguish between data and inferences, the text has been modified, and a couple of too speculative sentences eliminated or modified, e.g.,:

Page 2972 lines 14-15: the sentence "We interpret this pattern as being coupled with the phase of late fall-winter riverine runoff" has been eliminated.

Page 2972 line 17: we added "We advance the following hypothesis:"

Page 2972 line 28: we changed "We interpret this cycle as an overlap of the typical autumnal bloom of temperate regions followed by a progressive deepening of the thermocline and/or the subsequent vertical transport due to cyclonic or mesoscale frontal dynamics." with "We speculate that this cycle is an overlap of the typical autumnal bloom of temperate regions followed by a progressive deepening of the thermocline and/or the subsequent vertical transport due to cyclonic or mesoscale frontal dynam-

ics."

Page 2972 lines 26-29: we changed "We interpret the decrease not only as the result of the biological pump but also as the redistribution of carbon within the food web with an increased ratio of consumers vs. primary produce." with "We advance the hypothesis that the decrease is not only the result of the biological pump but derives also from the effect the redistribution of carbon within the food web with an increased ratio of consumers vs. primary producers."

Page 2974 lines 3-6 : the sentence "Apart from the role played by the new nutrients enriching the photic zone due to the deepening of the mixed layer, it is likely that the biomass reaches higher values also because of a relaxation of the grazing pressure." was eliminated.

3. I do not agree with the statement (abstract) that "The analysis confirmed that the Mediterranean Sea is an ideal area to evaluate the impacts of external physical forcing on the marine ecosystem functioning". To convince me that this statement is true I would need to be shown distribution of physical parameter and distributions of additional biological parameters, not just chlorophyll. The introduction mentions a variety of physical processes occurring in the Med and in the global oceans. Relating them directly to ecosystem response will convince me that indeed the Med has merits as a model for the larger oceans.

This is a subtle point, because our statement aims at highlighting the informative potential of the compact Mediterranean scales. The argument is the following. Mediterranean subregions reflect, to a larger or smaller extent, physical dynamics. They do not univocally follow physical dynamics but covary with it. Previous analyses have stressed this point and our inferences confirm the coupling, which is a basic paradigm of biological oceanography. What we stressed is that different regimes, presumably reflecting also different physical dynamics, are present in a relatively small area. We did not discuss the interannual variability of the areal extent of each regime, which is, in

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



some years, significant. This because we have to develop a better statistical method to allow for a robust intercomparison. We then assume that such variations are first order response to a change in the patterns of physical forcing. If so, the following step would be to monitor in situ the adjustment of biotic component to the change, which could provide insight on response time and mode of biota to physical forcing and, ultimately, to climate. In other words, we give for granted that there is a link between physical and biotic dynamics. Our point is that the Med is a good site to better dissect such link.

4. Why is the data not log-normalized? Why normalize to the maximal value (which is sensitive to outlier and thus less robust than, say, the 90th percentile)?

Clusterizing methods strongly require a normalization of the input data set, in order to avoid any bias introduced by the different ranges of values of the parameters to be classed (see for example Jain et al. 1999, "Data Clustering: a review", ACM Computing Surveys, Vol. 31, N. 3). Moreover, we normalized data for two additional reasons: 1) to avoid any bias due to ocean color algorithms; 2) to focus more on the "shape" of the seasonal courses rather than on the absolute values. The first point is clear. As for the second point, in our approach data are organized in a 52-dimensions space, where the i th variable represents the chl value for the i th week. A weakly time series is then represented as a point in this 52-dimension space. A similarity matrix is obtained calculating the euclidean distances between the 52-coordinates points and clusters are obtained finding "data structures" in the similarities matrix. If data are not normalized (or log-normalized, which, for the computation of the similarity matrix is practically the same) the clusters obtained are representative of the mean values of chlorophyll for a given point and not, as we wanted, of the shape of the seasonal chlorophyll cycle. Concerning the impact of the outliers, we believe that the data are sufficiently smoothed to effectively eliminate outliers. We used 8-days mean of level 3 standard NASA products, with all the QC flags and masks applied. In addition, the 10 years SeaWiFS data are averaged (using a median) to produce a weekly climatology. So, we are enough confident that the maximal value of each time series (which is used to normalize) is

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

representative of the chlorophyll peak for that time series.

5. It is claimed that the cluster analysis is more "robust" than traditional analysis yet no metrics for uncertainties or robustness are provided. Robustness may be demonstrated, for example by adding realistic noise to the data, by showing how well defined the boundaries between clusters are (e.g. by the portion number of points that change groups under changes in the size, geographical extent, fidelity of the data set etc.)

We got the point raised by EB. Therefore, we performed a series of additional tests, which have been added and discussed in the text. A new table summarizes the results of the tests. Figure 5 was also modified.

Text added at page 2971 line 9:

"Moreover, most of the cluster's time series exhibit small dispersion around the mean values (continuous line in Figure 5, evaluated by a +/- one standard deviation), which indicates that the classification is able to group together time series essentially similar. Cluster #7 constitutes an exception, as the spreading of the data encompass most of the dynamic range. To test the relevance and the stability of the regionalization, a series of statistical tests were performed. The original data set has been modified, introducing different degrees of noise (see later) and then creating several data sets "test". The clusterization was then applied to each modified data sets and the results were then compared to the clusters obtained from the original data set. The comparison was performed using as metric parameter, the Jaccard coefficient, (Henning, 2007 and references therein), which indicates the proportion of points belonging to both sets to all the points involved in at least one of the sets. A value of 0.7, or greater, indicates that the cluster is stable (Henning, 2007). Three different types of modified data sets were produced: a "boot strap", which uses the obtained clusters to introduce bias in the data sets, a "noise", which randomly replaces a percentage of points (5% in our case) in the original data set with noise points, and "jittering", which add to every single point in the original data set a noise or error. Noise and errors for the "noise" and "jittering"

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



data sets were calculated using the procedure indicates by Henning (2007), which is based on the covariance matrix of the original data set. For each type of test, 15 data sets were produced and for each data set the Jaccard parameter is calculated. Finally, the average of the Jaccard parameter is retained. The results are summarized in table 1. Only cluster #6 shows a Jaccard parameter below 0.7 for the "Boot strap" and the "Noise" tests, while all the other clusters have high values of the Jaccard coefficient. The three tests demonstrated that the applied clusterization is sufficiently stable and that the obtained clusters, with the noticeable exception of the #6, remain practically unaltered when the original data set is modified."

Boot Strap

Cluster1 0.830 Cluster2 0.821 Cluster3 0.851 Cluster4 0.789 Cluster5 0.884 Cluster6
0.611 Cluster7 0.916

Noise

Cluster1 0.815 Cluster2 0.803 Cluster3 0.876 Cluster4 0.776 Cluster5 0.911 Cluster6
0.676 Cluster7 0.955

Jittering

Cluster1 0.889 Cluster2 0.883 Cluster3 0.914 Cluster4 0.870 Cluster5 0.913 Cluster6
0.826 Cluster7 0.940

Table 1: Mean Jaccard parameter for the 3 stability tests, for the 7 clusters (see text)

6. You avoid defining what you call a bloom (p. 9) yet you write a lot about bloom timing, non-bloom regions etc. Unless you define what you mean by bloom it is not possible to evaluate your statements.

Text as been modified adding a sentence (page 2969 line 19): "Very simply, a "bloom" is a substantial increase (i.e. more then double) of the normalized chlorophyll from its seasonal baseline"

7. Qualitative sentences abound, e.g. "The geographical boundaries between the clusters are reasonably well defined". "The proposed classification, while more statistically robust". It will be useful if these statements were supported by numbers.

See answer to point 5.

8. P.13 is full of speculation about grazers and nutrients which are not supported by data (comment 2 above). The paper will be strengthened a lot if supporting data were presented (I have no doubt the ML depths, temperature, in the least, can be obtained at similar spatial and temporal resolution, e.g. FNOC model. Bb and CDOM fields can help as well, e.g. the analysis of Loseil of bb in the Med, possibly interpreted as in Behrenfeld et al., 2005).

See answer to point 2 and the introductory section.

Points raised by ED

1. The work of A. Longhurst is over simplified in the manuscript. Initially, A. Longhurst spent a great deal of time gathering hydrodynamic (temperature, currents, etc) and biological (chlorophyll profile, photosynthetic parameters) data to define his ecological provinces where physical forcing define the biological traits of each provinces as a distinct ecosystems Often scientists reduce his work to a simple study of satellite maps and/or climatological data.

We feel sorry for having given such impression to ED. Indeed, we consider the biogeography proposed by Longhurst a pivotal contribution to biological oceanography, not just to remote sensing data analysis (Longhurst insights are acknowledged regularly in our paper, i.e., pag. 2962, 2963, 2972, 2973, 2975). In fact our approach, while less rich and detailed, follows the same conceptual framework. We objectively analyzed just one term, the relative variation of biomass over the seasonal cycle, which, by itself, separates different sub-regions of the basin, displaying a similar pattern. Our classification is phenomenological and does not provide the mechanisms behind it, but we

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

consider remarkable the fact that the time course of biomass is such a good descriptor of an area. To avoid any possible misleading and to highlight the relevant differences between our approach and the Longhurst biogeography, we added some text:

Page 2963, Line 20 : "In respect of the Longhurst approach, however, the regionalization proposed here will be obviously much less detailed, as it will be based on a single term of the marine ecosystem (i.e. the surface chlorophyll concentration). However, the identified patterns (i.e. the bio-regions) could strongly improve the comprehension of the Mediterranean ecosystems functioning."

2. In their entire analysis, the authors ignore the bathymetry of the MS, which seems to be highly correlated to the patterns found in the phytoplankton concentration maps. (The use of bathymetry in the cluster analysis might reduce the speckling effect).

We agree only partially. Bathymetry is an important constraint, especially in the Mediterranean Sea, which is a "coastal" ocean. But at a first order of approximation, we do not see such a tight link between bathymetry and provinces. However, we expressly focused our analysis on the seasonal course of surface chlorophyll concentration as a possible proxy for a more complex dynamics, which in our view was also the main conclusion of Longhurst analysis. Therefore, we did not quantitatively analyze all the possible forcing/constraints such as bathymetry, SST, Mixed Layer Depth, etc.. The approach explored by us is a fertile way to extract information from satellite observations. K-mean cluster analysis confirmed the state-of-the-art depiction of Mediterranean Sea production cycle and improved it through coupling spatial patterns and temporal courses of chlorophyll concentration. These information are present in the satellite observations, although that are not directly manifest. Bathymetry, as other important forcings, would be considered in a mechanistic analysis to be performed.

3. The Hoevmoller diagrams give valuable information regarding the development of the phytoplankton bloom and they also reveal a strong North /South and East/West gradient in chlorophyll concentration. Normalisation of the data by the maximum value

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

reduces the information carried in the time series. The reasons given by the authors to use this normalisation are somehow concise. I would suggest to use the natural log-transformed of the data to keep the natural magnitude of chlorophyll concentration and the associated information. This comment would also hold true for the cluster analysis.

When normalizing data, an important piece of the information is lost. As explained in text, however, two main reasons led us to normalize the data set: 1) to avoid any bias due to ocean color algorithms, which have been already observed in the Mediterranean Sea; 2) to focus more on the "shape" of the seasonal courses rather than on the absolute values. Concerning the first point, the normalization doesn't resolve completely the algorithm problem, although it strongly mitigates his effect. Concerning the second point, most of the previous papers on the Mediterranean have already identified the geographical gradients North/South and East/West. Our approach confirmed these results, and tried to go more in depth in the analysis, exploring the relationships between the spatial gradients and the temporal patterns. In particular, we formulated the hypothesis that the shape of the chlorophyll seasonal time course (which is defined more by the timing of the increase and decrease of biomass and less by the absolute values of his peaks and minima) reflects the observed spatial patterns. This hypothesis is not trivial, though we obtained that the areas showing high mean annual values of chlorophyll concentration have temporal cycles significantly different from the areas showing very low values. This result was not expected, and it would not have been possible without the normalization of the seasonal trends (or with a log normalization) as, in this case, the shape of the time series would have been still dependent on the absolute values.

4. The authors performed various test to justify the optimum number of clusters to use in the statistical analysis (i.e., 7 clusters). However, the results show that the development of the bloom follows three main patterns in the Mediterranean basin. On a biological aspect, it seems therefore that three clusters should suffice to describe the trophic regimes of the MS.

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

The separation of the clusters in three main groups was probably misleading, and it was adopted only to better discuss the results. In fact, the seven clusters (as demonstrated by the stability test; see point 5 of the EB comments) are relevantly different: clusters #4 and #5 (bloom and intermittently) and clusters # 6 and # 7 (coastal) have shapes clearly different. The "No bloom" clusters are more similar, but remarkable differences are evident in the duration of the high chlorophyll values, in the timing of the initiation and of the decline of the bloom, and in the differences between the max and minimum values.

5. The authors also normalised the chlorophyll time series to the maximum values but Figure 5 does not show maximum values equal to 1. Does that mean that the normalisation was performed before the "climatological" analysis? If yes, it is interesting to note that cluster 1,2 and 3 have a maximum value close to 1 (?0.9), whereas the remaining clusters (especially 6 and 7) have a very low maximum value. Could that be interpreted as a difference in the timing of the maximum of chlorophyll concentration, which is averaged out over the period of ten years? If yes, it could also play a non-negligible role in the characterisation of the slope of the initiation of the bloom. It would be interesting to plot the standard deviation associated with the "climatological" data because some of the interpretation on the slope of the timing of the bloom in each cluster could be misled by some "outliers" (a very high chlorophyll concentration in a give week) for a given year.

Normalization was computed after the creation of the climatological matrix. In other words, a climatological year was created and each climatological time series was normalized by the climatological maxima for that pixel (see pag 2965, line 19). In the K-means clusterization, the centers of the clusters are simply the mean values of all the elements belonging with a specific cluster. The comment of ED is correct: if all the time series in a cluster have the maximum value during the same week, the maximum of the time-series for that cluster should be exactly one. This is not the case, because the timing of the maximum values shows variability in periods longer than a week.

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

However, for most of the clusters (# 1, 2, 3, 4, 5), maxima are between 0.8 and 0.9, which indicates a general coherence of the timing of the maxima. For the remaining clusters (#6 and 7), the low values of the maxima reflect the more important variability in the time-series belonging for these clusters.

However, we agree with ED that showing the standard deviation associated to the centers of the cluster could better highlight the significance of each cluster. We modified then the figure 5, showing separately the time series of the centers of the clusters and adding the associated curves of the +/- one standard deviation. The standard deviation is calculated on the time-series belonging for each specific cluster.

We also added/replace some text:

Pag 2969 line 25: we substitute "Figure 4 shows the spatial distribution of the clusters obtained with the K-means procedure, while the seasonal evolution of the 7 centers is reported in Figure 5" with "Figure 4 shows the spatial distribution of the clusters obtained with the K-means procedure, while the seasonal evolution of the 7 centers, with the relative +/- one standard deviation, is reported in Figure 5"

Page 2971 line 9: we added: "Moreover, most of the cluster's time series exhibit small dispersion around the mean values (continuous line in Figure 5, evaluated by a +/- one standard deviation), which indicates that the classification is able to group together time series essentially similar. Cluster #7 constitutes an exception, as the spreading of the data encompass most of the dynamic range."

Detailes comments:

Page 2964, line 5: add with between "anomalous" and "respect"

Done

Page 2971, line 6 to line 9: I find all the clusters quiet patchy.

See answer to point 5 of EB

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

A new version of Figure 5 is proposed.

Interactive comment on Biogeosciences Discuss., 5, 2959, 2008.

BGD

5, S2299–S2313, 2008

Interactive
Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

