About the origin of the Mediterranean Waters warming during the twentieth.

By García-Martínez et al.

Answer to reviewers.

First of all we would like to thank both referees for their review and comments. The manuscript has been modified according to their suggestions. Much new work has been done in order to attend these suggestions. New calculations, figures and tables are included.

In some cases some new analysis, figures and tables have been done according to the reviewer's suggestions. In other cases these analyses had already been done but not included in the previous version. Now they have been included. Some figures have been modified and supplementary material has been included for the completeness of the work. We honestly believe that the present version has followed most of the comments and suggestions made by the referees and the new manuscript is much improved.

Referee #1.

1. The first comment states that **"there are a lot of important information that is missing and some methodological issues that need to be addressed"**.

We agree with the referee and we accept that some methodological aspects should have been explained in a better and more detailed way. Section 2 is now named as Data and Data methods. Certainly the methodology used has been explained in detail. In some cases new analyses have been included. We hope that all these questions will be answered in specific comments.

2. Specific comments.

1) The paper lacks a major finding. After reading the introduction (in the way that is presented now), I was not convinced of the need of another paper in the subject. As the authors point out, there are already several published papers on the same topic, sometimes with conflicting results. So, What does this study brings up which is new? It is not clear. I asked myself: (i) does the

paper analyze new data? It seems that the answer is no, because other papers cited by the authors (e.g. Vargas-Yáñez 2005) also use the same database; (ii) do the authors employ a new methodological approach? It also does not seem to be the case; (iii) does this paper reconcile the previous conflicting results as was mentioned? I couldn't find any indication. My suggestion is that the authors need first to think deeply about two things: why a new paper is needed in this topic, and what is the specific novel result in the paper. The introduction should be written in a way that leads the readers into these two points.

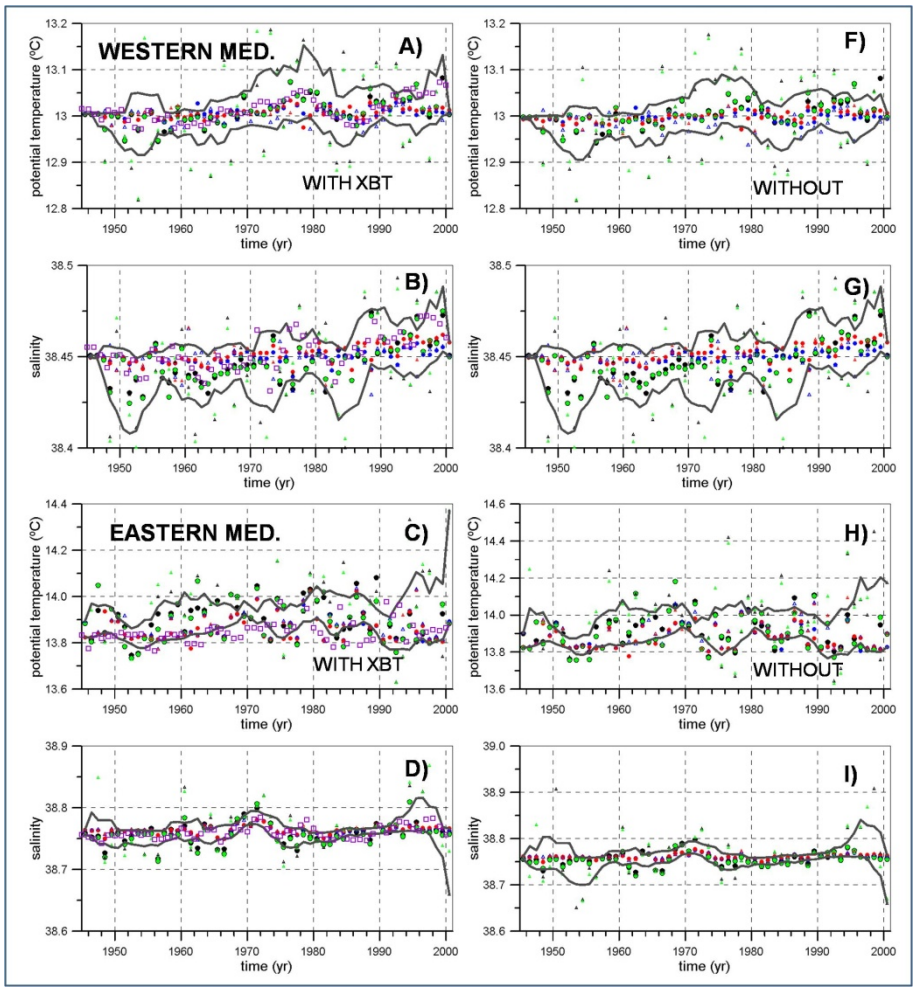
The manuscript certainly brings new findings. Nevertheless we accept that it was not clear in the way it was written in the previous version. Furthermore, following some of the suggestions made by the referee and in order to answer some of his/her comments, we have changed some methodological aspects. The new approach improves the manuscript and makes more clear which are novel results.

The introduction has been re-written to make more clear which are the objectives of the manuscript and what is really new.

a) We clearly explain that the analysis of the evolution of the Mediterranean Sea temperature and salinity during the twentieth century is very interesting because the possible effects of climate change can be masked with other anthropogenic stressors as the damming of some of the rivers draining into the Mediterranean Sea. The first step should be to establish which changes have really occurred. It is true that our research group had already published several works showing that there were some discrepancies between the results found in the literature. It is also true that we had already published some works showing that the reason for these discrepancies could be that changes in the data analysis method could alter substantially the results. So, doing this again is not anything new. What is new is that, one previous work compared the results between different methods for constructing the time series, calculating the annual values and filling data gaps (Vargas-Yáñez et al., 2009). Another work (Vargas-Yáñez et al., 2012) compared the use of time series obtained averaging in situ data with those time series obtained from the 3D objective analysis carried out by Rixen et al. (2000). A third work (Vargas-Yáñez et al., 2010) compared the results obtained using time series that included XBT data with those results obtained from time series which did not include such data. Therefore, each work dealt with an aspect of the problem.

The first thing that is new in the present work is that the annual time series for temperature and salinity have been calculated using all the possibilities already considered in previous works.
The second thing that is new is that this more complete analysis is used for providing estimates for the errors, instead of a more classical approach based on the calculation of standard deviations or similar. When a seasonal time series (one data per season of the year) is considered, some gaps are found. We first consider the possibility of averaging all the data corresponding to the same year. If, for instance, the winter data is missing, the annual mean will be artificially high. If the summer data is missing, the annual mean will be artificially low. If the gaps are randomly distributed, that is, if the sampling is not seasonally biased, some annual values would be higher than the real one and others would be lower. The variance of the time series would be increased by the gaps, but the trends would not be biased. In fact this is another criticism made by the referee in his/her point 3) **Is the number of profiles seasonally biased? Regionally biased?**
Following this suggestion we have included Chi-2 and contingency tables to check this point. These tests have been carried out for the different regions of the Mediterranean Sea and the result is that the sampling is not seasonally nor regionally biased. This result is included in the new section 2. Tables have been included in supplementary material TS1 and TS2.

Going back to the point 1). Another possibility to deal with gaps is to fill them with climatological values (zero anomalies in Gregory et al., 2004). Another possibility is to estimate the mean value of the anomalies instead of the temperature or salinity values. Let's say, if only a winter data is available, and that winter was warmer than the climatological winter, it is considered that the year is also warmer than the climatology (representative anomalies in Gregory et al., 2004). As the zero anomaly approach produced flattened time series, another possibility was to correct the climatology with the seasonal trend. Finally, time series from the Rixen et al. (2000) objective analysis were used. These time series have no gaps. Notice that for each region of the Mediterranean Sea: Alboran Sea, Balearic Sea, Ligurian Sea, Algerian basin, etc. we have constructed 5 annual time series (for each variable). Then we have to average all the time series from the WMED and all the time series for the EMED. For instance, some years we have data from all the regions in the WMED, but other years, only data from some regions are available. Once again we can fill the gaps with the regional climatology or use representative anomalies. Therefore, for each of the 5 time series we have two possibilities to accomplish the spatial mean (10 time series). Finally we repeat all the process for the data set "with" XBTs and "without" XBTs. The figure S3 in supplementary material shows the large dispersion of annual values. Notice that this figure is for the Mediterranean Waters which has a much lower natural variability than the Atlantic Waters where the problem is worse.



S3.

What is new is that we have considered for each year the minimum and the maximum values from all the values corresponding to all the time series. We have also considered the mean value for each year. In this way we construct a time series and an estimation of the error which is larger than the one obtained using the standard deviation (for instance). This is a much more conservative and novel approach.

We have to admit that this was not done in the previous version and it has been included in order to answer the reviewer comment. We think that it really improves the work.

b) We present in the introduction different hypotheses already proposed for explaining the warming and salting of the Mediterranean waters. Many works considered that it was caused by the salinity increase of the Atlantic Waters. We show, by means of heat, volume and salt conservation laws and a box model that the salting of AW in fact produces the warming of the Mediterranean waters, but is not sufficient to explain the observed trends. Therefore the freshwater

reduction in the Mediterranean Sea is an important factor, but not the only one. A reduction of the heat losses to the atmosphere is also needed for explaining the warming trends in the Mediterranean waters. We think that this work helps to clarify this question which was completely open up to the moment.

c) Referee #2 asked for the use of more recent data bases for the extension of time series to the twenty first century. This was not the main goal of the work because the main objective was to describe and explain the causes for the changes during the XX century. Nevertheless we accept that the work is much more complete including these new data. Therefore we have extended the time series in the WMED using data from our monitoring program (RADMED) and also in the WMED and EMED using the EN4 objective analysis from the Met Office Hadley centre (Good et al., 2013). The RADMED data had already been used in the previous section but the use of the EN4 data is something new.

We have re-written the introduction trying to express very clearly what is new in this manuscript. We hope that we have succeeded.

2) **My second concern is about the data. Long-term changes in thermohaline properties of deep waters are normally very small, albeit indicating important differences. In order to be able to detect these small signals (beyond reasonable doubt) we need accurate measurements, preferentially in the same locations. The things are a bit more complicated for evaluation of long-term salinity changes, because of the large uncertainties in salinity measurements when compared with the small variability typical of deep waters. Additionally, the way we measured salinity throughout the 20th century has changed a lot. These data issues have never been discussed in the MS. Indeed, there is no section describing the MEDAR/MEDATLAS dataset. Are the authors only using CTD profiles? Are the measurements (especially salinity) calibrated against in situ observations from bottles? Are the profiles Quality-Controlled? What is the accuracy for temperature and salinity measurements? Did the accuracy change throughout OSD time? Are the authors using XBT data? If yes, have the XBT data been corrected for the fall-rate changes? What is the period covered by the MEDAR/MEDATLAS data the authors are using? From the data analysis section, I understood that it is between Interactive 1943-2000. How many profiles in total? How many profiles by region? Have the num-comment ber of profiles been steady throughout the time? etc, etc .... The authors need to include a section describing this kind of thing, so we can evaluate the results. For now it is very hard, if not impossible, to understand. I am unsure if the computed trends, especially in salinity, are outside of the margin of error.**

First we have to acknowledge that the referee is right. The ideal situation for detecting long term changes which certainly are very small if compared with the inter-annual or decadal variability is to use routine measurements at fixed locations. This is one of the main points in the present work. Fixed stations were not available or very scarce in the Mediterranean Sea during the twentieth century. Therefore, the only way to accomplish the trend detection is using the available data. We accept that recent and new monitoring programs have appeared and there are new platforms and data centers which collect the oceanographic data, control the data quality and offer improved products. A mistake in the previous version is that these new platforms, monitoring programs and data centers had not been appropriately acknowledged. We have done it in the present version and the COPERNICUS, SEADATANET, Met Office Hadley data center, World Ocean Data base, etc. have been acknowledged. The Ocean State Report from COPERNICUS (Von Schuckmann et al., 2016). has also been referenced. Nevertheless, for the analysis of the XX century, MEDAR data base contains most of the existing bottles, CTD and XBT data.

Concerning the data quality control the referee is right again. No mention was made to the data quality control. MEDAR data are quality controlled and those profiles labeled as wrong or low quality have been discarded. We have explained this in the new version. In the case of RADMED monitoring program the CTDs are periodically calibrated and for each cruise the salinity data are also calibrated with salinity measurements from bottle samples analyzed in the laboratory. Beside the quality control of MEDAR data, all the time series are checked for anomalous values. First a rough inspection is carried out eliminating those temperature and salinity values out of the range 12 ºC - 28 ºC and 36 psu - 40 psu. A second checking is carried out eliminating in all the time series those outliers which depart more than 3 standard deviation from the climatological mean. This methodology has been explained in the new version of section 2.

Another important question is the use of XBT data which have been shown to be biased by errors in the fall-rate equation. We also accept that this question had not been properly addressed in the previous version. As we only have access to the MEDAR data, not to these fall-rate equations, we cannot carry out any correction. Nevertheless we have to accept that this could be introducing some errors in the trend estimations. The approach that we have followed is the following one: We construct two different data sets. A first one that does not include the XBT data, and therefore we do not have the problem of biases in the XBT. And a second one including XBT data. We check if the results are sensitive to the XBT inclusion. We had already done this analysis in a previous work and that was the reason for not repeating it in the present work. But we have to accept that in this way the present work was not clear enough.

We would like to point out that in the previous version we had estimated the standard deviation for each annual temperature and salinity value at each region and depth level, although we had not shown that result. We are grateful to referee #1 for his/her comments about the uncertainty calculations. On one hand we have included in supplementary material the standard deviations for the annual means which could be considered as a "classical" estimation of the errors. On the other hand we have included in this version a new estimation of the time series uncertainty associated to the different methods used for the construction of time series. We have already explained it in the new manuscript and in this answer letter. Certainly new calculations and results have been included and we believe that the new manuscript is much improved thanks to this suggestion. Following the referee's suggestion we have also included figures presenting the number of annual data for each basin: WMED and EMED (supplementary material).

Another point raised by the referee is the period of time covered by this work. We have to admit that it was not clear enough as some times it was referred to as 1943-2000 and other times as 1945-2000. The reason is that data from MEDAR data base had been obtained from 1943, but the Rixen et al. (2000) objective analysis covers the period 1945-2000. In the present version this has been unified and all the MEDAR time series extend from 1945 to 2000. RADMED data merged with MEDAR data extend from 1945 to 2015, and the objective analysis from the Hadley Data Center also extends from 1945 to 2015. We think that this is clearer in the present version.

3) My third concern is methodological. Appropriate statistical tools to deal with autocorrelated and inhomogeneous datasets are needed to investigate long-term trends in intermediate-deep waters from observations. This subject is not adequately discussed in the MS, although the group has authored several papers dealing with methods. I would expect to hear something about effective degrees of freedom, but there is nothing written about that. Since the authors are working with time series, I suggest them to consider more adequate methods to extract trends from short and noisy time series. Other examples of concern in the methods: Are there enough data to divide the Mediterranean into 11 different regions? What was the criterion used to divide in 11 regions? How many profiles by region/year/month? Is the number of profiles seasonally biased? Regionally biased? Data inhomogeneity can cause large biases and aliases, impacting the computed linear trends. I also have tried to understand why so many different averages (page 4) have been calculated in relation to the objectives of the work. I strongly suggest the authors not only describe the methods but also to explain the reasons for their choices. It is also not clear why the authors did not use an atmospheric reanalysis product to obtain the evaporation and heat fluxes in their box model, similar to Lozier and Sindlinger (2009). The authors could have used for example MERRA-2, NCEP or another reanalysis product. In this way they could better interpret their results. For instance, the authors state that an increase of 5-7% in C3 evaporation and reduction in oceanic heat losses of 0.4-0.5 W/m2

could explain the observed change. Evaporation depends on the latent heat flux, which is a component OSD of the heat flux budget. Therefore, the two results may not be independent.

We also admit that these questions were not clear enough in the previous version. The referee is right when he says that some references to effective degrees of freedom are missing. In fact we had already taken it into account. We did not mention it because it was explained in some of the referenced works. Nevertheless we have to admit that it was not clear in the present work. New section 2 explains how the linear trends are estimated and how the confidence intervals for the slopes are calculated. We also explain that the validity of the calculations depends on the normality of the residuals and their statistical independence. The normality is checked by means of Chi-2 and Smirnov-Kolmogorov tests. As the residuals are not statistically independent the degrees of freedom are substituted by effective degrees of freedom which are calculated using the integral time scale. All these questions are now explained in the new version.

Data are selected for different geographical areas. The referee asks for the criteria for this selection (11 areas in the WMED and 4 areas in the EMED). One criterion is to follow previous works. More important is to select areas with similar oceanographic conditions. There is an important reason for selecting temperature and salinity profiles for each region and then to calculate the average value for the WMED and for the EMED. If, for a particular year 100 profiles are available for the Alboran Sea and 10 profiles are available for the Tyrrhenian Sea, averaging all the profiles would produce a biased value. We first estimate the annual mean value for the Alboran Sea, the mean value for the Tyrrhenian Sea and then we average both values considering the volume of each region. The same is applied to temporal averages. If all the values corresponding to one year were averaged the result could be biased. We first estimate the mean values for each season and then we average the four seasonal values. Ideally we should estimate the mean value for each moth of the year, but data scarcity made us consider the use of seasonal means and then annual means.

We explained in the points above that the data sets were checked for seasonal or geographical biases (see answer to question 2).

Finally, the referee suggests the use of reanalysis products for the estimation of heat fluxes, evaporation, etc. We admit that it could be interesting to run the box model considering the temperature and salinity evolution observed from new monitoring programs and the heat and freshwater fluxes from observations or reanalysis. Nevertheless we think that this would be a different and new work. For the present work

we simply try to check how the temperature and salinity of the Mediterranean waters change when these fluxes are changed. Then, the model needs that the heat and freshwater fluxes are initially balanced and capable of reproducing the Mediterranean steady state. Actual values from different data bases and reanalysis do not necessarily reproduce the net balance of the theoretical steady state. We hope this is now clarified and we will consider this suggestion for future works.

3. Minor comments.

a. There are acronyms not defined, for example, WMED and EMED (It is comment also needed to define the geographic area the authors are considering). L#101 MEDARS/MEDATLAS, L#104 RADMED, L#146 CTD etc... Please check all throughout the text.

Ok. We have checked all the acronyms and the geographical areas are labeled in the figure 1. The geographical areas corresponding to RADMED have also been indicated in the text: Alboran Sea, Cape Palos, Balearic Sea and Gulf of Lions and these areas are indicated in the figure 1.

b. Number of decimal places in trends also needs to be checked throughout the text. Some of these seem to be well above the sensor's accuracy.

The accuracy of the trends is not directly related to the accuracy of the measurements. For instance, let's consider a thermometer which can only measure one decimal place: 13.2, 13.3, etc... The accuracy of the measurements is 0.1 ºC. Let's consider that during a period of 10 years the trend in the temperature of the sea is 0.01 ºC. Initially the temperature of the thermometer is 13.2 ºC. During the first 5 years the temperature is lower than 12.25 ºC but our thermometer will measure 12.2 as only can measure one decimal place. During the second 5 years the thermometer will measure 12.3 ºC as the temperature is higher than 12.25 ºC. The time series has the shape of a step. If a linear trend is fit by least square, the linear trend reflects the 0.1 ºC change in 10 years and therefore it is 0.01 ºC/yr. The trend is one order of magnitude lower than the accuracy of the measurements. I am not sure if this is the point raised by the referee. We hope this clarifies this question.

c. Which are the scales of salinity and temperature reported? TEOS-10? PSS-78?

PSS-78 as this is the scale used in MEDAR. psu are the units used for salinity values along the manuscript.

d. There are references with missing information. Please check.

Ok. We have checked it.

e. Figures also need improvements. For example, in Figure 5 there is no red curve, so I don't know what the red axis is about. I am also not sure why creating displaced axes in several figures.

Ok. These figures are new. Because of the changes in the work and the new estimations of the uncertainties using the new methodology, these figures are completely new.

f. I would consider putting figure 1 in the introduction to help the readers.

Ok. We have followed this suggestion.

g. In general, English also needs some improvement.

Ok. A bilingual colleague has reviewed the manuscript.

Referee #2.

1. The paper methodology is obsolete, superficial and the results not clear. None substantial contribution to scientific progress about the topic of discussion has been found in terms of concepts, ideas, tools, or data.

We accept that maybe the previous version was not clearly written. Maybe the objectives of the work were not well presented and therefore the new ideas in the previous manuscript were not clear. We have re-written the introduction and we hope that it is now better organized.

During the 1950s and 1960s, the damming of the main rivers in the Mediterranean Sea reduced the freshwater discharges. Certainly this is not a new result.

Some works had hypothesized that because of the river damming, the surface layer salinity increased. Some works considered that the Nilo River damming was the main factor and the origin of Atlantic water salting was in the Eastern Mediterranean. Others considered that the Ebro River damming was the cause of the salinity increase and therefore the origin was in the Western Mediterranean. Certainly this is not a new result.

Some works considered that saltier waters could reach the deep water density with a higher temperature. Therefore the cause of the deep water warming was the damming of the rivers. This is not a new result.

Other works considered that deep waters were warmer because of global warming and the heat absorbed by the oceans. In the case of the Mediterranean Sea, as it losses heat to the atmosphere, the cause would be a reduction in the heat losses. This is not new.

If heat, salt and volume conservation laws are considered, it can be shown that the increase of the salinity of the surface waters is able to produce the warming of the Mediterranean Waters. These conservation laws show that this warming can be understood from a different point of view. Instead of considering the winter deep convection process and the formation of dense waters with a higher temperature because of the higher salinity, it can be understood from the point of view of the exchange at Gibraltar and the heat balance of the Mediterranean Sea. Saltier waters are also denser. If the Mediterranean outflow is proportional to the density difference, the outflow would be higher. If the net evaporation and the outflow are higher, the Atlantic inflow would be higher. If the net transport at Gibraltar (inflow minus outflow) compensates for a higher net evaporation, the net transport also increases. If the Mediterranean water temperature remained constant the result is a net heat gain. The final result is the temperature increase of the Mediterranean water until a new heat balance is reached. This provides a new explanation for this hypothesis.

Although the AW salting produces the warming of Mediterranean waters, this warming is lower than the observed one. Therefore, this hypothesis cannot explain the observed temperature trends completely. This is a new result.

A combination of a reduction of the heat losses to the atmosphere with the damming of the Nilo River could explain the warming and salting of the Mediterranean waters. This can be shown in a simple way just using the conservation laws for heat, salt and volume. This is a new result.

It had been proposed that the salinity trends of waters flowing into the Mediterranean Sea through the Strait of Gibraltar could be the cause of the Mediterranean water warming. The box model used in this work shows that this AW salting could produce a certain warming of the Mediterranean waters, but once again it is not enough to explain the observed trends. Therefore the heat loss reduction has to be taken into account. This is a new result.

Many of the hypotheses proposed for the Mediterranean Water warming and salting tried to explain contradictory results. We have shown that the in situ data available for the second half of the twentieth century are scarce and make the construction of time series and the trend estimations very sensitive to the data processing methods. We have

to admit that this is not a new result. Nevertheless, thanks to the referee's comments we have included some new calculations. We have used several methodologies to deal with data gaps and the construction of annual time series. We have estimated the uncertainty of the time series using the uncertainty associated to these methodological aspects. We have compared this uncertainty with other classical approaches as the estimation of the standard deviation. This is a new approach.

The main objective of this work is to analyze the changes occurred along the twentieth century. The reason is that this is the time when climate change and river damming acted together on Mediterranean waters. MEDAR data base is quite complete if in situ data are considered. On the other hand, objective analysis from Rixen et al. (2000) has also been included in the new version. If the time series are extended until 2015 certainly MEDAR data are not useful.

We have to admit that other data sets and platforms such as EN4, Copernicus, SeaDataNet, etc. had not been referenced in the previous version. We have referenced these data sets, data centers and some works published using these data in the new version. Furthermore, we have included new analyses using EN4 data in the WMED and EMED. We have compared the results obtained using EN4 data in the WMED with data from the RADMED monitoring program in the Spanish Mediterranean.

2. The in situ observations used for the analysis are from MEDAR/MEDATLAS dataset (2002!) which actually does not present observations in the bozes considered for the analysis in figure 1 until 1948. I would recommend to start the analysis from 1950-1955 if you not find additional observations. Data sparseness has been frequently mentioned to motivate the lack of robust results but none effort has been done to overcome it. It is not acceptable when dealing with climate change issues and long term estimation.

In the previous version we had found some profiles from 1943, nevertheless they were very few. In the present version, as we have already explained, we have used different methods for constructing annual time series. One of them is the use of 3D temperature and salinity fields from an objective analysis carried out by Rixen et al. (2000). These 3D fields are available from 1945. Therefore this is the initial date in the new version. The in situ data from MEDAR data base include most if not all the existing bottles, CTDs and XBTs from the twentieth century. Obviously the referee is right if the twenty first century is considered. In that case new data bases are needed. Then, the conclusions have to be obtained from the available data sets. If the data sets are not appropriate for obtaining robust results we honestly believe that it has to be stated in that way. This is the case for the upper Atlantic layer. On the other hand, robust

conclusions can be obtained for the Mediterranean water. This is shown in the present manuscript and we believe that this is a new result. On the other hand, the box model helps to explain the processes occurred in the Mediterranean Sea during the twentieth century when the in situ data are not enough to conclude.

3. The extension of the analysis till 2015 only considering the Spanish waters monitoring observations is not scientifically sound when the Copernicus Marine Service provides both Near Real Time and REPprocessed data starting from 1950 till 2017 and 2016 respectively. It provides also global OA maps in recent publications to estimate the Ocean Heat Content anomaly and trend (see Ocean State Report 2016).

Ok. The referee is right. The present work is not complete if the extension until 2015 is only based on the RADMED data from the Spanish Mediterranean waters. We have made a comparison between the data from the RADMED area including RADMED data merged with MEDAR data with the data from the whole WMED. In this way we show that the RADMED monitoring program is suitable for studying, at least, the long term changes and the decadal variability in the WMED. Figures 7C and D and the correlation estimated between both data sets support this conclusion.

Following the referee's suggestion we have used data from the EN4 objective analysis from Met Office Hadley Data Center. In this way we have analyzed the changes in the temperature and salinity of both the WMED and EMED until 2015.

In this new version we have tried to state that the lack of significant trends in the surface layer could be caused by the data scarcity. We show that routine measurements from satellite have positive linear trends. This could be caused because of the systematic sampling which make it possible to detect long term changes or because of an acceleration of the warming and salting trends during the end of the twentieth century and beginning of the twenty first century. We have used results from the Ocean State Report from the Copernicus data service as suggested by the referee.

4. The statistical analysis is very weak without any detail about the data distribution, the number of data per box.

We have to admit that the referee is right. We had already done most of these statistical analyses but we did not include them for trying to keep the manuscript not too long. We have included some of this information in the new manuscript and in the supplementary material; therefore the length of the manuscript is not excessive. Several new points have been included in the new manuscript:

a. Figures S1 and S2 provide the number of annual data for the WMED and EMED for several depth levels. Notice that the data availability is highly dependent on the depth level.

b. We have also shown the number of seasonal data for the surface level for several regions of the WMED and EMED. We have shown by means of Chi-2 tests that all the seasons of the year are similarly sampled and that this situation stands for all the regions. We have also shown that this situation does not change along the second half of the twentieth century. That is, the sampling is not seasonally biased and this is true for the first part of the data set and for the second one. This is explained in section 2 and table TS1 and TS2.

c. We have included the standard deviation estimated for the annual mean values for the WMED and EMED in figures S4, S5, S6 and S7.

d. We have explained that MEDAR data are quality controlled. The flags provided by MEDAR are taken into account, but furthermore, we have carried out our own quality controls neglecting suspicious data. A first inspection discarded data out of a certain range. Then, all data more than 3 standard deviations from the climatology were discarded. This had already been done for the first version, but we had not explained it. Now it has been included in the new version.

e. Following the suggestions made by both referees to strengthen the statistical analysis, we have explained the estimation of linear trends and the confidence intervals for such trends, taking into account the effective degrees of freedom caused by the auto-correlation of the time series (new section 2). As in the previous point, we had already done this for the first version, but we had not explained it. Now it has been included in the new version.

5. The implementation of a box model is not able to answer to the posed questions and in many cases the author recommends the use of satellite data or 3D models. Once again, the Copernicus Marine Service provides many satellite and reanalysis products, both global and regional, covering the analysis period and it is advisable to consider these products to sustain the results.

Ok. In our answer to some of the previous points we have also admitted that these new data services such as Copernicus, Sea Data Net, etc. had not been properly acknowledged in the previous version. First we have now referenced the Copernicus Ocean State Report (Von Schuckmann et al., 2016) which supports our conclusion about a possible acceleration of the Mediterranean warming and salting trends. We have also included the analysis of other data bases such as the EN4 (Good et al., 2013).

We agree with the referee and we consider that the new products provided by Copernicus and other data centers and services will help to clarify the temperature and salinity evolution of the oceans in the near future. This has been explicitly acknowledged in the conclusions section. One of the main points of the present work is to distinguish between changes in the Mediterranean caused by the damming of main rivers during the XX century and a possible warming caused by climate change (reduction of heat losses). We believe that the MEDAR data base and the box model used has been able to answer some of these questions: The salinity increase of the Atlantic Waters, caused by damming of the river Nilo, or river Ebro or the AW trends in the nearby Atlantic Ocean, are not able to explain the warming of Mediterranean Waters. A decrease of the heat losses to the atmosphere has to be also considered.

We hope that this is better explained in the new results, discussion and conclusions sections which have been completely re-written.

6. The manuscript is not well organized, ...

The section 2: Data and Methods have been partially re-written. New subsections have been included. Sections 3 and 4 have also been re-written. Section 1 has also been partially re-written in order to express the objectives in a more clear way. We hope that everything is now clearer and the objectives and main results can be more easily understood.

Specific comments.

- It shuld better specify the period considered in the analysis.

Ok. We have already answered in point 2 in main concerns. The analysis is from 1945 to 2000 for the study of the changes during the second half of the XX century and then from 1945 to 2015. We hope it is now better explained in section 2.

- " A new analysis of MEDAR data" what is the novelty of your approach?

Ok. We think this is answered in the answer to main concern 1. The analysis of different methodologies and the use of all of them for constructing a mean time series and its uncertainty is really new. The analysis of possible causes for these changes using the heat, salt and volume conservation laws is also new.

• There are some errors in the English, thus I recommend to take care about the language before a new submission

Ok. A bilingual colleague has reviewed the manuscript.

• The use of conditional in many sentences makes the paper weak, especially when mentioning the results in the last phrase. The verb is wrong.

Ok. We have tried to review this. Nevertheless we use the conditional for not making very strong statements.

• The Western Mediterranean Transition, its causes and effects, have been already studied in recent papers (not listed here), thus I recommend to read and cite them.

The introduction deals with changes along the XX century. Nevertheless the results and the discussion deal with changes until 2015. New works such as Pinardi et al., 2015, Schroeder et al., 2017, 20166, von Schuckmann et al. 2016 have been cited. This last work is also cited now in the introduction.

Introduction

• Line 60à Line 65: A lot of literature has been dedicated to this and your superficiality is not allowed. Please consider more recent papers (i.e. Pinardi et al., 2015; Schroeder et al. 2017, von Schuckmann et al. 2016)

Von Schuckmann et al., 2016 has been cited in the introduction. Pinardi et al., 2015; Schroeder et al., 2017 has been cited in the discussion.

• Line 95: What is the relationship between the Ebro damming and the AW salinization?

This relationship was proposed by Krahmann and Schott (1998). According to these authors, the river Ebro damming is responsible for the AW salting presented in the figures in that work.

• Line 102-103: actually, you are considering not the twentieth century but the 1943-2000 time period and since there are no data until 1948 I would change the analysis time period in the second half of the century.

Ok. We have corrected this and explained in previous answers. Now the study period starts in 1945.

• Line 103: which transients? EMT and WMT? It is not clear.

Introduction has been re-written. The transients are mentioned in the discussion. Both transients are clearly referenced now.

• Line 105-107: The intent to use the box model is very limited, moreover you should at least state clearly which one are the hypothesis you are considering.

Ok. Following the referee's suggestion this section has been re-written and we have tried to write very clearly which are the hypotheses that we check and the main objectives of the work.

Data and Methods

• Line 114-115: Almost all regions do not present data in the forties, what is the statistical significance of your temperature and salinity monthly averages in these years?

Ok- We have started our time series in 1945. Notice that one of the time series used are the objective analysis from Rixen et al. (2000; 2005). For all the years the standard deviation is included in suppl. material S1 and S2. Nevertheless if annual means for these years have a large error, the important thing is that they contribute to a large dispersion of all the time series used (figures 4 to 7 and S3) and they reduce the significance of the trend estimations. Therefore we adopt a very conservative approach. Those trends which are significant considering the large uncertainty and despite the different methods used can be considered as robust ones.

• What is the reason to compute monthly than seasonal and then yearly averages, is it a robust approach? Then you average spatially in the EMed and WMed and then again to get the Med average. I consider this approach very superficial, especially when trying to compute long term trends. None consideration about the different instrument types of observations and consistency issues. Did you consider XBT data, which correction did you apply?

This has been explained extensively in previous points. XBTs are considered constructing two data sets, one including and another one excluding these data. We search for differences between the results obtained from both data sets. We use seasonal means because monthly data are very scarce for intermediate and deep layers. Seasonal time series (or monthly if more data were available) have to be computed because otherwise, computing directly annual means would result in biased annual values. We have already explained it in this answer letter and we believe that the methodology is better explained in the new version. In the same way, annual mean values are calculated for each region previous to the WMED or EMED calculations to avoid biased induced by the data gaps in some regions. We have also explained this in previous answers.

• How did you propagate the error from your monthly averages per each region to your annual Med profiles or to the AW and MW annual properties?

The errors are calculated as the standard deviation. Then when averaging different seasons or different geographical areas, the pooled standard deviation is considered (Zar, 1984). Nevertheless, take into account that finally the error used is the dispersion associated to the different methodologies used. The reason is that this uncertainty is larger than the one associated to the standard deviation. Once again we select the most conservative approach. We have included these new calculations thanks to the referee's comments.

• Line145: a data distribution map is advisable.

Certainly in would be interesting, but we think that there are many new figures and the new version is clear enough now.

• Line 149-150: this is not clear at all.

These lines do not exist anymore.

• Line165-178: sensitivity tests and results should go here to motivate your model

implementation.

• Section 2.4 is very superficial, there is not a detailed description of the experimental design.

We hope that the work is now better structured. Most of sections 1,2 3 and 4 have been re-written.

Results

• Section 3.2, Lines till 285: they should go in the methodology description. Only here ee know that you are running the model for 1000yrs with a time step of 1 year. You are presenting the sensitivity tests results here and they should go in 2.5. You should motivate your chose at line 175.

This section has been re-written. These lines do not exist anymore.

• For the three sensitivity test the spin up time of the model is 200 years, why did you run it for 1000 years before testing your hypothesis?

Certainly the spin up time could be shorter. Nevertheless it does not alter at all the results. Considering a period of time shorter than 200 years would be wrong, but considering a period of time for the spin up has not any importance. On the other hand the figures result clearer and the reader can see clearly that the steady state is reached.

• Again you are presenting results without explaining your experimental design.

Ok. Re-written, we think that the objectives of the box model experiment and the different experiments carried out are now clearer.

Discussion and Conclusions

• Lines till 330 should be part of the introduction, what's their meaning here?

This section has been re-written and these lines do not exit.

• Lines 340 and Lines 355: if you think that these studies should be made considering 2D circulation models why did not you use them. Copernicus Marine

service provides many reanalysis products for the Med region and you did not even mention them).

As we have already answered, this was a mistake. These services and some of the publications have been referenced.

• Lines 385-395 is a repetition of the introduction

• Lines 410-415: you are describing only here figure 7c and d, why? They should go in the results.

Ok. This has been re-written.