Reviewer comments are in black, answers from the authors are in blue and corrections added to the manuscript are in green

Anonymous reviewer 1

Summary: The manuscript investigates several aspects of heat flux dynamics in the Mediterranean Sea. First the manuscript compares the long-term mean heat flux from two reanalysis products with different spatial resolutions (ERA5 and ECMWF) and the higher resolution reanalysis (ECMWF) is found to provide a heat flux consistent with the 'closure hypothesis'. The authors attributed this difference to spatial resolution. Then the authors look at the PDF of the turbulent heat fluxes and also look at the impact of extreme events on the heat budget. The authors find reasonable PDFs that capture the statistical patterns in the heat fluxes terms and that fall/winter cooling events are criteria to achieving the negative long-term mean of ECMWF that is consistent with the 'closure' hypothesis.

The work is interesting and will likely be of interest to a broad range of scientists. Couple of questions that I think should be address in some fashion.

Authors: we thank the reviewer for the interest in our work, and we will try to answer all the questions posed.

Major comments

Seems like there's an issue with using the closure hypothesis as evidence for which result is 'correct'. Finding data/results that fit the expectation assumes the hypothesis is truth which seems a bit problematic to me. It seems like the authors should be framing the results differently

We base the closure hypothesis on the heat and mass conservation equations for the Mediterranean Sea. These equations are fundamental for all the earth system modelling; they are not a specific dynamical balance. We argue that, since Gibraltar unequivocally brings heat and water in, the surface fluxes should in principle balance this input. How much is the balance it is not known, and we have hypothesized it is perfect; this is clearly an approximation. As in the seminal work of Bryden and Kinder (1991) and the recent work by Cessi et al. (2014) and Jorda' et al. (2017), the balance between volume integrated heat and mass content helps to understand the basin dynamics. Specifically, being heat entering laterally the Mediterranean Sea, then we search for a negative net surface heat flux. How negative we do not know but searching for a negative net heat flux is a conservative assumption aligned with current scientific understanding.

We have now modified the introduction at line 60:

Furthermore, the estimate of the Mediterranean Sea heat budget from ECMWF meteorological analysis data sets has not been done before.

After line 76 we discuss the hypothesis behind the "closure of the heat budget":

We realise that assuming perfect balance between lateral and vertical heat fluxes, even in the Mediterranean Sea, is an approximation. Being heat clearly entering the Mediterranean Sea through Gibraltar, we search for a negative net heat flux, which we call the closure hypothesis.

How negative such net heat flux is, we do not know but searching for a negative value is a conservative assumption aligned with current scientific understanding.

References

- Bryden, H. L., & Kinder, T. H. (1991). Steady two-layer exchange through the Strait of Gibraltar. Deep Sea Research Part A. Oceanographic Research Papers, 38, S445-S463.
- Cessi, P., Pinardi, N. and Lyubartsev, V., 2014. Energetics of semi enclosed basins with two-layer flows at the
591 strait. Journal of physical oceanography, 44(3), pp.967-979, https://doi.org/10.1175/JPO-D-13-0129.1
- Jorda', G., Von Schuckmann, K., Josey, S. A., Caniaux, G., Garc.a-Lafuente, J., Sammartino, S., ... & Mac.as, D. 624 (2017). The Mediterranean Sea heat and mass budgets: Estimates, uncertainties and perspectives Progress in Oceanography, 156, 174-208 https://doi.org/10.1016/j.pocean.2017.07.001

It's not clear that the spatial difference in the reanalysis products is the only potential cause of the difference. Could there be other possible causes? For example, could the cloud physics be the issue. Looking at Figure 1, the long wave and short-wave radiation are the terms that look most different. Since SST is the same in both cases, the differences in long wave radiation could point to the cloud cover parameters, maybe? Is cloud representation done the same way in these two reanalysis products. It is not clear to me the only difference in this reanalysis products is the spatial resolution. Maybe I missed that though.

We agree that spatial resolution may not be the only factor causing the difference. Our method allows to eliminate the SST as a possible cause, as noted by the reviewer. We agree with the reviewer that the distribution of cloud cover is also an important difference (reported in Fig A1 below). In fact, the difference is a complex function of different quality of the atmospheric variables. To be noted is that ERA5 and ECMWF analyses use approximately the same model and data assimilation systems. However, we have added also the consideration of cloud cover among the potential differences.
We would like to point out that we had already a sentence at line 223: “Furthermore, ECMWF and ERA5 different values are connected to different cloud cover.”

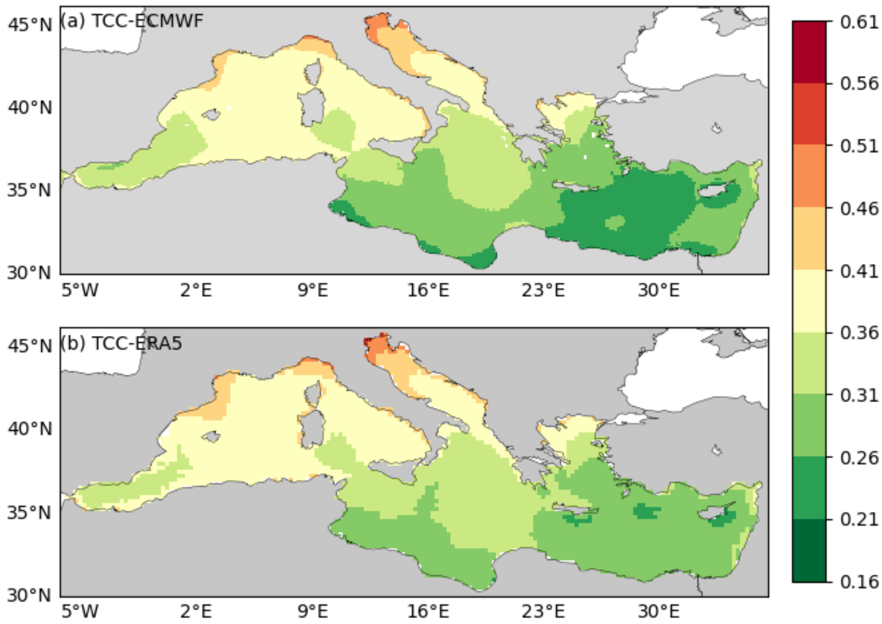


Figure A1: Total cloud coverage (%) mean computed for the period 2006-2020, a) ECMWF and b) ERA-5 input datasets.

To make a more balanced statement we have eliminated in the abstract the sentence at line 35 “This highlights the importance of high-resolution atmospheric data for accurately capturing air-sea interactions and ensuring physically consistent climate modelling over the Mediterranean Sea.”

replacing it with:

“Only ECMWF fields are consistent with the heat budget closure hypothesis.”

Why are the statistical distributions just the turbulent heat fluxes explored. Its fine, but it seems like the author should comment on why these are the target and why the distributions of the longwave and shortwave radiation are not explored.

We thank the reviewer for pointing out this missing justification. We wanted to compare our turbulent flux distributions with similar work in the literature which uses only turbulent heat flux components like Gulev and Belyaev (2012) and Korolev et al (2015), also listed below.

- Gulev, S. K., & Belyaev, K. (2012). Probability distribution characteristics for surface air–sea turbulent heat fluxes over the global ocean. Journal of Climate, 25(1), 184-206, https://doi.org/10.1175/2011JCLI4211.1
- Korolev, V., Gorshenin, A., Gulev, S., & Belyaev, K. (2015). Statistical modeling of air-sea turbulent heat fluxes by finite mixtures of Gaussian distributions. In International Conference on Information Technologies and Mathematical Modelling (pp. 152-162). Springer, Cham, https://doi.org/10.1007/978-3-319-25861-4_13

Furthermore, if we look at the time series of the fluxes (Fig. A2 below), it is clear that the turbulent components are the one that exhibit larger anomalies with respect to the seasonal cycle, hinting to the presence of skewness and kurtosis in their distribution.

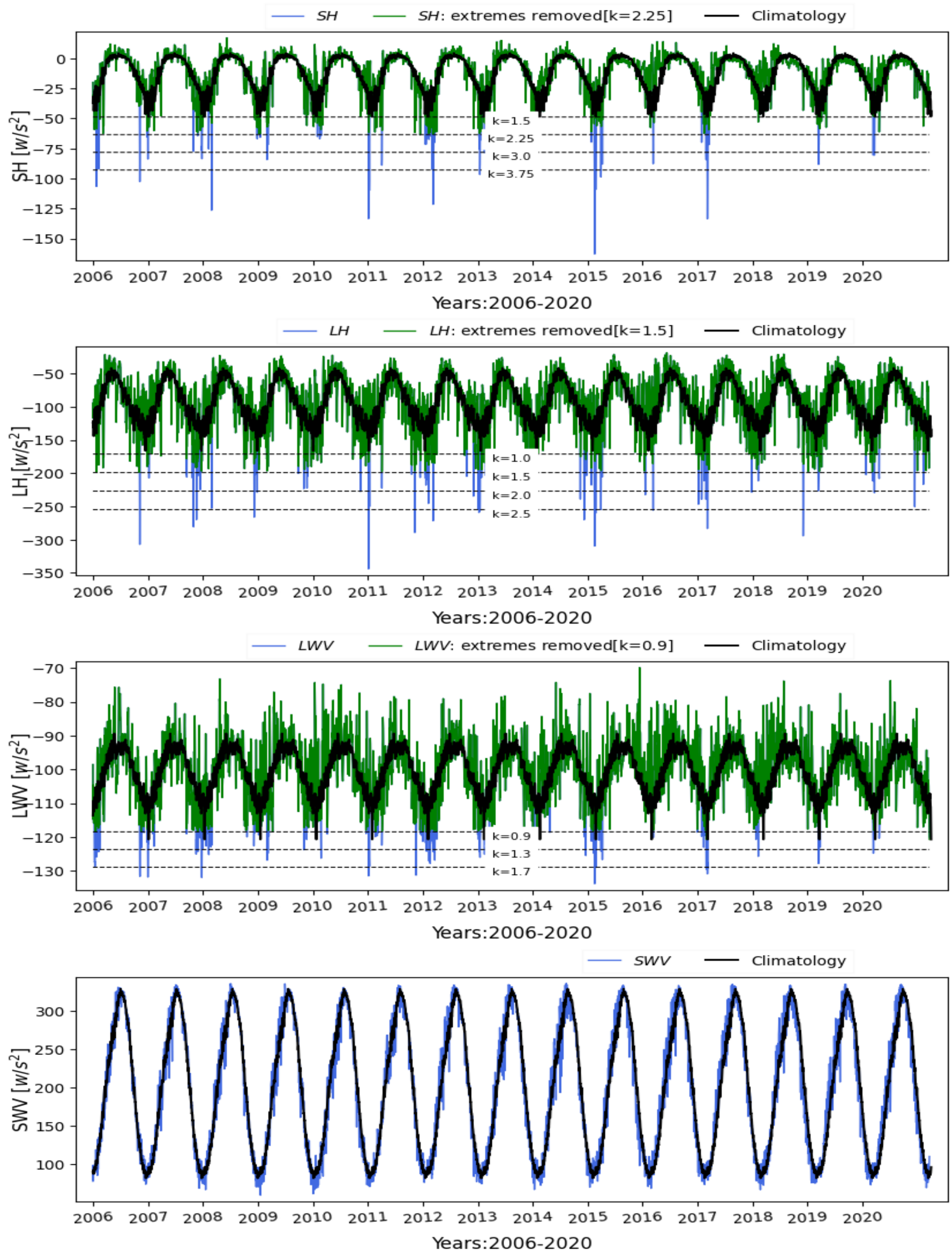


Figure A2: Time series of the basin averaged SH, LH, LWV an SWV and their respective extremes removed, overlapped with long-term yearly climatology. Lower quantile boundary lines are marked with dashed lines with different k values used to remove and replace with climatology values

We have now changed the beginning of the section 4 adding, after line 295, the following statement and Fig. A2 in the supplementary material:

Recent studies by Gulev and Belyaev (2012) and Korolev et al. (2015) have analysed the statistical distributions of turbulent heat fluxes, and their findings are used here for comparison. Radiative flux components are excluded from this analysis, as they do not exhibit extremes of comparable magnitude to those of turbulent fluxes (Supplementary Material, Fig. ??). This suggests low skewness and kurtosis in their distributions, reducing the relevance of a detailed probability density function analysis for these components.

With regard to the extreme events. I guess it should not be surprising that if you remove the most extreme negative values and then average the heat flux the mean will get warmer. However, potential feedback may not be accounted for that should be mentioned. For example, if the extreme heat flux events are removed but there impacts on SST is not, then the subsequent fluxes may be lower than should be if the extreme event in the heat flux never occurred. So, the relative gain associated with the extreme event may not be as significant indicated by just removing the extreme heat flux. There may be other potential feedback that the removal methodology does not full consider that should be mentioned.

Thanks for raising your concern here. We wanted to show the impact of extreme values on the net heat flux time series because it has never been done before, and it highlights where the major uncertainties in the heat budget closure reside. We understand that in a fully coupled atmosphere-ocean system the feedback can be important but our is a diagnostic study and we cannot change arbitrarily the SST. We do not think we need to comment this in the text but if requested, we can add a comment saying that this is not like to do a simulation changing the atmospheric forcing fields extremes which in turn will generate a different SST.

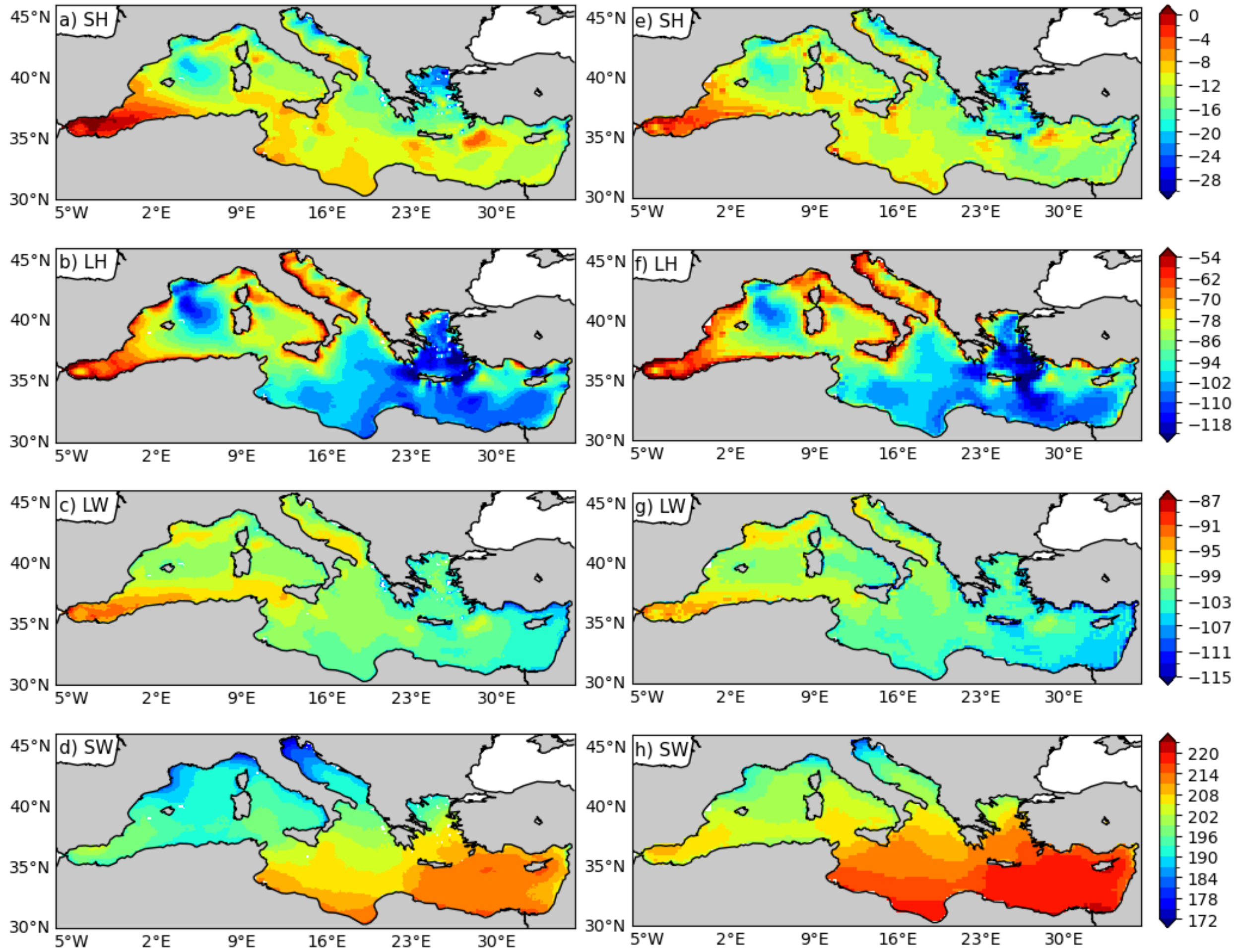
Can more details or discussion be provided on why spatial resolution limits the ability of ERA5 to represent extreme heat loss in fall and winter?

Thanks for the question. A higher spatial resolution is important for capturing many small-scale atmospheric and oceanographic features. In contrast, ERA5 comes with a horizontal resolution of ~31 km, which can smooth many small-scale features and underestimate frequency of extreme fluxes, such extremes are often associated with local wind flows, sharp air-sea temperature contrast and coastal orographic effects.

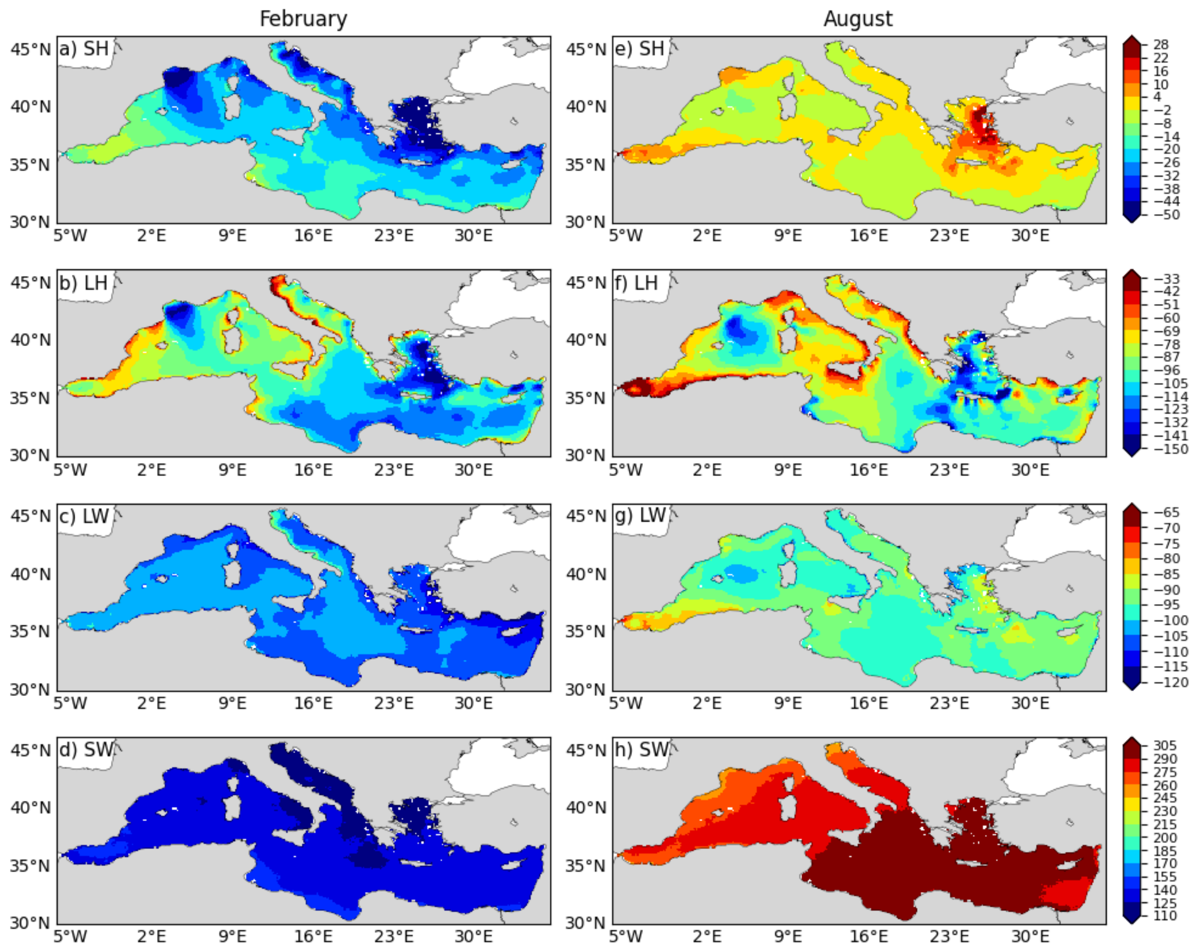
We apologize for the mistake; we displayed a picture of the fluxes (Fig. 1 and 2) with plot smoothing which did not help to see the noisiness of ERA5 with respect to ECMWF. Here we show an updated Figure 1 and 2 where the differences in resolution are evident in addition to other issues.

We have now replaced Fig. 1 and 2 with the new ones, unsmoothed pictures, and commented in the text after line 229:

Fig. 1 shows the noisiness of the fluxes due to the ERA5 low resolution with respect to ECMWF while retaining an overall consistency.



New Figure 1: Mean annual heat flux components for the period of 2006-2020 computed from ECMWF (left) and ERA5 (right) daily time series



New Figure 2: Seasonal variations of heat flux components: Left column is the monthly average values for February and right column is the average for August for the period 2006-2020 (ECMWF data).

Similarly related – Figure 7 shows the Qnet, but what is driving the extreme Qnet – typically I assume latent heat flux (and reduced shortwave) is the main driver of fall/winter cooling events, but Figure 1 suggests that these are quite similar between ERA5 and ECMWF at least in terms of the mean… Is that the case in the extreme events?

Thanks for your comment. Yes, we agree that latent heat flux is a major component for the fall/winter extreme events. We show now in Fig. A2 that also sensible heat flux provides considerable fall/winter extremes.
The location of the heat flux anomalies (Fig.1) is the same due to the geometry of the basin and the atmospheric forcing structure, but the extremes are different. We now show the time series of Qnet for ERA5 in Fig. A3 which in our opinion makes the point together with Fig. 8.

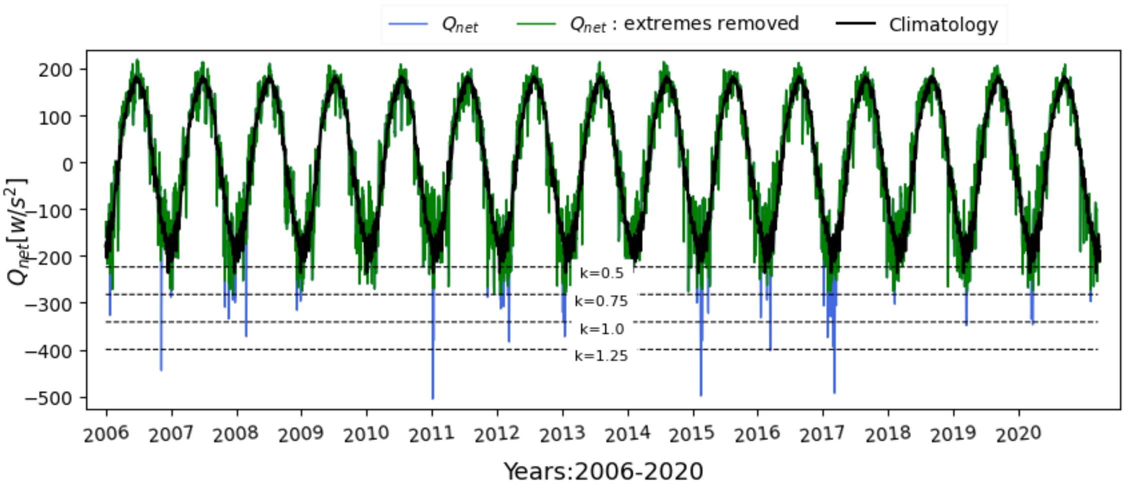


Figure A3: Time series of the basin averaged $\boldsymbol{Q_{net}}$ (ERA5) means, $\boldsymbol{Q_{net}}$ extremes removed, and long-term yearly climatology and four lower quantile boundary line marked with dashed lines using different k values [k=1.25, 1,0.75, 0.5].

Minor comments

Line 50-51 Awkward phasing

Thanks, sorry for the unclear phrasing. Now it is:: Moreover, the Mediterranean net heat budget comprises of several terms that show a considerable range of uncertainties (Jorda' et al., 2017).

Line 83 'have shown large deviation' - revise phrasing

Thanks, sorry for the unclear phrasing. Now it is: Using downscaled NCEP/NCAR global reanalysis of ½° x ½° resolution, Ruiz et al. (2008) computed a heat budget of -1 Wm-2. However, their heat flux components values are not close to most of the literature values (for instance, the major difference was in the value for net short wave with 84 Wm⁻2).

Line 162 Was ρ (rho) defined?

Added " ρ =1.22 kg/m3".

Line 400 ' that this is the reason why…'

Corrected and added, here we refer the Figure A3 also

Line 404-405 I don't really understand this sentence.

Thanks for suggesting an explanation, here is the new text:

The *Q*net could become an impact indicator of the Mediterranean for sea level trends in the basin. The net heat budget in fact relates to the sea level tendency (Pinardi et al., 2014) in the Mediterranean Sea and could be considered as a key indicator of climate impacts in the Mediterranean Sea.

Pinardi, N., A. Bonaduce, A. Navarra, S. Dobricic, P. Oddo, 2014. The mean sea level equation and its application to the mediterranean sea. J. Climate, 27, 442–447, doi: 10.1175/JCLI-D-13-00139.1

Line 453 'Differences appear…' The main difference were in the long and short wave radiation. Figure 1 showed that latent and sensible heat fluxes were not that different.

Thanks for suggesting the clarification, the text is now: "Differences appear in the structure of the fluxes, especially the SW and LW, when different atmospheric data sets are used,…"

Line 485-487 These sentences are confusing to me. I suggest revising them in some way.

Thanks for pointing out the missing minus sign. The phrase:

"The threshold value that produces a positive basin mean heat loss is 231 W m^{-2}. Thus, if the basin mean heat loss does not exceed this value, the basin is not in steady state. This might be a good indicator of Mediterranean Sea heat content trends to be exploited in the future."

Is substituted by:

"The anomaly threshold value of -231 W m^{-2} (Table 2) results in a long-term positive net heat flux, which is inconsistent with the basin's energy closure assumption, thereby indicating the presence of long-term changes within the basin due to atmospheric forcing"

Line 493 poor phrasing – this sentence should be revised.

Revised "Furthermore, the PDF analysis of turbulent heat fluxes will allow us to have a better understanding of the extreme events and their contributions in the net negative heat budget. "