Climate of the Past Discussions

Clim. Past Discuss., author comment AC4
https://doi.org/10.5194/cp-2020-161-AC4, 2021

Reply on RC3

Yoav Ben Dor et al.

Author comment on "Hydroclimatic variability of opposing Late Pleistocene climates in the Levant revealed by deep Dead Sea sediments" by Yoav Ben Dor et al., Clim. Past Discuss., https://doi.org/10.5194/cp-2020-161-AC4, 2021

Response to Report #3

By Reik Donner

Response to general comments:

We appreciate Prof. Donner's comments and very detailed suggestions, as we highly appreciate his proven experience and rich publication background with time series analyses in geoscience. We are therefore grateful for the opportunity to improve the manuscript according to his comments and ideas. We will adjust, remove and recalculate any necessary analyses to make sure they comply with his suggestions.

Additionally, we agree with the notion that the current supplementary is too long, as was pointed out by the other reviewers, and that some parts of it are indeed redundant. We will revise and reduce it, and also make sure that its referencing is clear and consistent throughout the manuscript.

Response to major comments:

Comment 1: The presentation of the employed time series analysis methods (recurrence analysis, wavelet analysis, singular spectrum analyses) is very short and hardly assessable to non-specialists, which will form the vast majority of the readership of Climate of the Past. I think that a more detailed introduction (possibly as part of the supplement) would be justified.

Response 1: We will improve and expand the presentation of the methods where necessary in accordance with the journal's formatting policy.

Comment 2: It is pretty unlikely that paleoclimate variability in the study area has undergone (exactly) periodic oscillations at interannual to multidecadal scales. Therefore, I do not agree with using the term "periodic components" (e.g., in l.65), yet would expect something like "narrow-banded oscillations" or similar. What can actually be characterized in the presented study is the relevance of certain "spectral bands" (interannual, decadal, multi-decadal) and how their respective spectral power may differ among the two study periods. I will further comment on this point below when addressing the performed wavelet analyses and the interpretation of the corresponding results.

Response 2: We agree with that comment. We will rephrase that accordingly and replace the current terminology with the one hereby proposed by Prof. Donner.

Comment 3: According to ll.146-147, missing values have been imputed by the median – of the whole time series (segment)? It needs to be noted that this strategy may have quite different effects on the different time series analysis methods used in this work. For SSA, it might actually be better to just ignore missing values. What is more, missing value imputation based on the results of the singular value decomposition of the lagged trajectory matrix (e.g. Kondrashov & Ghill, Nonlin. Proc. Geophys., 2006) would allow for a more reasonable (e.g., consistent with the records' power spectra) gap filling and, hence, likely more reliable results of the other analysis methods.

Response 3: We appreciate this very useful suggestion, and we will impute the missing values accordingly.

Comment 4: I appreciate that the authors use nonparametric statistical tests for the homogeneity of distributional (location and variability, respectively) properties of their three proxies between the two study periods (e.g., using the MWW test instead of classical parametric ANOVA). However, when reporting the corresponding results (Tab. 2), what is presented turns awkward. Notably, according to its contents Tab. 2 apparently relates the results of MWW and AB tests to either the fall (MWW) or the rise (AB) period (which would be meaningless), while both tests actually compare both periods. I suppose the authors accidentally copied the first column of the Table from Tab. 1, yet it must be removed from Tab. 2 to make any sense. In Figs. S2 and S3, I don't quite get the statistical reasoning beyond multiplying the p-values of both tests (which are not independent of each other); this should be better motivated.

Response 4: Thank you! Yes, this is indeed an unfortunate mistake and column 1 should be removed. The table compares the properties of the two periods. The general idea of multiplying the p-values was to try and emphasize places in which the two values drop substantially together. However, we agree that it has no clear statistical reasoning and we will also remove that figure from the supplementary.

Comment 5: For detecting regime shifts within the study period, the authors use running MWW and AB tests for 51-year sliding windows in time. Here, I am wondering about several things:

- What are the two samples which are compared by the two tests? The 25-year sub-periods before and after the reference point? I don't find this clearly explained in the text.
- What is the reliability (power) of MWW/AB tests for such small samples of 51 (25?) values only?
- Sliding windows mean that the samples considered in subsequent tests largely overlap,

potentially causing multiple testing problems when assessing the significance of pointwise MWW or AB tests. The same applies to performing the same tests for different window sizes. I don't see this aspect being addressed, so it would be good if the authors could comment on this and why they might think it could be ignored in the context of the present work (which I personally believe could, but has to be justified). In any case, it can be expected that results for neighboring windows and different window widths will mutually depend on each other.

- For assessing statistical significance, the authors use resampling of the full time series. If I understand correctly, this is being done by permuting individual values. However, this procedure does not only destroy any differences between sub-periods, but also any serial dependencies of proxy values within such periods, thereby making the obtained confidence bounds over-confident (i.e., potentially too narrow). Figure S4 indicates the absence of such serial dependencies (to a great extent) and thereby could justify the employed procedure (i.e., not using block bootstrapping instead of point-wise resampling), but this should be mentioned explicitly in the text.

Response 5:

- Yes, we will clearly state that in the revised manuscript.
- We chose a window size of 51 years after trying several window sizes, and examining their performance. On one hand, we wanted to avoid over-detection of "noise" by choosing a too narrow window, and on the other hand we wanted to avoid looking into too long time intervals, which would depreciate the implications of the analyses. Because each series spans ~700 years, we consider the semi-centennial timescale as a reasonable window length that serves as a compromise between the two abovementioned aspects. We did not conduct a test to determine the power of this approach, because we think this is not strictly necessary for the discussion and is beyond the scope of this paper. In our view, this analysis is only carried out as a complementary method to refine the results of cluster detection based on the Monte-Carlo approach elaborated later in the manuscript (currently SM2).
- This is again a very good comment. Because we don't rely solely on this approach and consider it together with the other approaches elaborated troughout the manuscript, we did not account for the multiple-tests issue related to the overlap. However, because the p-values of the sliding windows are likely to be dependent, we do not consider the p-values of the tests themselves (e.g., vs. a specific alpha level etc.), but instead we look at the patterns they form along the series, and more specifically whether substantial minima can be identified, which signify substantial differences between the two halves of the window. This is why we think this issue can be ignored in this context, and we can assume that this is what Prof. Donner is referring to.
- This will be explicitly mentioned in the revised manuscript.

Comment 6: Still in Section 3.2, the authors describe their rationale for using recurrence analysis, which is probably rather unfamiliar to the vast part of the readership of Climate of the Past. Yet, also the authors appear not to be specialists in employing this technique, which is suggested by a couple of observations.

- The authors claim that they have also performed cross-recurrence analyses (l.176) but do not report any such results (which might also not be very useful since they would compare two proxies with different meanings, physical units, etc.).
- Recurrence analysis can indeed be used to infer short-term periodic and quasi-periodic dynamics (in the proper meaning of both terms), as claimed in l.177, but neither of the corresponding approaches is used in this manuscript (e.g., studying the properties of the tau-recurrence rate a.k.a. generalized auto-correlation function; cf. Zou et al., Phys. Rev. E, 2007).

- Comparing the use of recurrence analysis with that of harmonic and wavelet functions (l. 182) and their respective "robustness" is somewhat odd since those methods serve completely different purposes.
- A vast body of recent work has detailed the problems of using a fixed recurrence threshold and taking the recurrence rate RR as a parameter, as other quantitative characteristics of recurrence plots intimately depend on the value of RR. Fixing the threshold at a multiple of the standard deviation of the data only partially solves the problem, since the distribution of distances between state vectors (values) evaluated is commonly non-Gaussian and may crucially differ between different settings studied. Most notably, the values of epsilon (sigma or even 1.5sigma) are far larger than those commonly recommended in the literature (e.g. Schinkel et al., Eur. Phys. J. Special Topics, 2008). The resulting RR values (Figs. 4 and 5) approach values between 0.1 up to even 1.0, which are far too large to allow for any meaningful interpretation of the transitivity values (Zou et al., Phys. Rep., 2017). This also explains why RR and transitivity show the same type of time dependence, while the transitivity should actually be independent of RR for a reasonable range of epsilon values.
- Using time delay embedding and reporting/justifying the corresponding embedding parameters is key for interpreting the results of recurrence analyses and making them reproducible. This is poorly described in this manuscript, although all necessary results are found in the supplement. It is particularly interesting to observe that both proxies more or less instantaneously de-correlate (Fig. S4c,d), which is a behavior common for white noise. On the other hand, using embedding dimensions of 4 or 5 for 700 data points might already exceed what might be required for a reliable statistical inference of the key recurrence structures. Here, m=3 might be a more pragmatic choice. In general, using the same embedding parameters for all recurrence (and joint recurrence) plots would help making the obtained structures, as well as their quantitative characteristics, better comparable.

Response 6:

- We acknowledge the fact that we are not experts in employing recurrence analyses. However, we have done our best efforts to apply these methods based on the available literature and the software package distributed by PIK (Marwan et al., 2007). The mentioning of cross-recurrence is indeed redundant, as in the final manuscript we eventually report the results of recurrence and joint-recurrence, rather than the cross-recurrence, and will be corrected accordingly.
- We appreciate this comment as well. We will perform the abovementioned analyses and contact Prof. Donner in case we need additional specific guiding on the proper way of applying them. We will make sure that the methodological aspects related to these methods are resolved, and we are willing to consider removing them from the manuscript in case Prof. Donner finds it necessary.
- This will be rephrased and clarified.
- Again, we appreciate this comment. We have indeed tried multiple approaches and values for the epsilon parameter before the selection of these parameters that we found suitable. Because of our appreciation of Prof. Donner's background with these methods, we will revise the calculations based on his (and the references) suggestions and contact him if necessary, to make sure that the analyses are carried out properly.
- The calculations will be modified in accordance with his suggestions, and this will be elaborated in the revised manuscript. We note that this information is clearly available in the current supplementary material.

Comment 7: Regarding the wavelet analysis, I again appreciate that the authors provide the results of significance testing with an AR(1) red noise null model. However, what is crucial to remark is that they perform this test in a point-wise manner (which is unfortunately still the standard in the applied geosciences literature), thereby overemphasizing possible false positive results. Due to the serial dependence of point-wise values of the wavelet coefficients at neighboring times and scales, false positives can only be ruled out using areawise tests (Maraun & Kurths, Nonlin. Proc. Geophys., 2004; Maraun et al., Phys. Rev. E, 2007). I don't argue here that it is necessary to employ such tests as part of the present study, but recommend to evaluate and interpret the results of point-wise tests with more caution. Quite a few of the high-frequency episodic significant patches in the wavelet spectrograms shown in this manuscript (e.g. Figs. 6 and 7) could potentially be associated with such false positives. (Referring to "non-persistent periodic[al] components of 2-6 years" in ll.371-372 appears more like wishful thinking than proper interpretation of the obtained results.) Along with my former comments on the use of the terms "periodic" and "quasi-periodic", I recommend to focus on spectral power in different frequency bands instead of seeking for true periodicities which are unlikely to exist at the timescales of interest (due to an absence of obvious mechanisms except for maybe solar activity variations).

Response 7: We will recalculate the wavelet analyses using an area-wise false-positive test to make sure that the results can be accordingly analyzed. We also agree with the notion that these results should be interpreted cautiously and that the overall discussion should follow the identification of "different frequency bands instead of seeking for true periodicities".

Comment 8: Singular spectrum analysis (SSA) appears to be primarily used here for detrending and "denoising" (i.e., reconstructing the underlying signal based on a few modes, which is probably related to what the authors refer to as "overfitting" in l.209 – this should be clarified for non-specialist readers). As already outlined above, I would recommend using this method also for gap filling and, hence, as a first analysis step. It might also be worth mentioning that SSA is more flexible than wavelet analysis (or at least classical spectral analysis based on Fourier transform or harmonic regression models) in that it allows for an arbitrary shape of possible oscillatory components along with time-dependent amplitudes (like in wavelet of classical EOF analysis/PCA), but not for time-dependent frequencies. The latter restriction might be alleviated by using other even more data-adaptive time scale decomposition techniques like empirical mode decomposition, which the authors decided not to consider in their present work (which is fine, since possible advantages of other methods also come along with additional caveats). Notably, the authors also use the multivariate extension of SSA, yet only show the corresponding results as plots in the supplement without further discussion and interpretation, so one may argue that this material might not be relevant. (If relevant, it should also be discussed in the text.) In a similar spirit, it is notable that Figs. S6 and S7 are currently not (respectively, wrongly) referenced in the text, but should be referred to in l.212.

Response 8: We will clarify that as part of the revision process, and we will also apply this approach for gap-filling as well. As for the application of additional methods, such as empirical mode decomposition, we tried to limit the amount of applied methods to a reasonable extent, which would suffice for addressing the goals of the research. We agree that more analyses can be done, but as Prof. Donner suggests, every method has its advantages and disadvantages, and our general notion is that enough analyses are presented in the current version, so we would refrain from conducting additional analyses. We will remove the multivariate SSA from the supplementary, as it shows similar results that do not contribute substantially over the univariate SSA. We will reduce the amount of supplementary material and will additionally review all references and make sure they are properly referenced.

Comment 9: A bit worrying is the application of cross-wavelet analysis, which is not well described in the manuscript (ll.212-214). My understanding is that the authors first use SSA for detrending and denoising the time series under study and then estimate the cross-wavelet spectrograms for pairs of time series. It is notable that this type of analysis is commonly not recommended, since it can provide large spectral power even if only one of the two signals actually exhibits a "periodic" component. For the purpose of seeking for joint oscillatory components, the normalized wavelet coherency should be the method of choice instead (Maraun & Kurths, Nonlin. Proc. Geophys., 2004).

Response 9: Following the provided reference, we agree that the cross-wavelet analysis is not robust enough on its own to support the identification of mutual periodicities, so this will be removed and replaced with the wavelet coherence.

Comment 10: Section 4.3, 2nd paragraph: It should be clarified that substantial spectral power in the low-frequency part is a common feature of climate time series. Hence, the fact that the wavelet spectrogram does not indicate statistical significance in this range of frequencies indicates that any low-frequency (inter-decadal) oscillations embedded in the signals do not follow a strictly periodic pattern. Note that at the mentioned time scales, the cone of influence becomes so narrow here that the number of oscillations may not be sufficient to identify properly any periodic structure.

Response 10: We agree with this comment. This will be clarified.

Response to minor comments:

Comment 11: The second paragraph of the introduction briefly discusses key drivers of hydroclimate variability in the Levant. In this regard, I am somewhat missing any brief statements on possible teleconnections from the Indian Ocean. In modern times, there exist anomalous circulation patterns linking the Arabian Sea branch of the Indian summer monsoon with the climate of the Eastern Mediterranean region. The active Red Sea troughs (ARST) are a manifestation of associated episodic events providing heavy precipitation to the study area, as also mentioned by the authors in the last paragraph of Section 5.3. It might be interesting to explore, or at least speculate about a possible link between elevated flood frequency and Indian monsoon failures as documented in historical heavy precipitation events of the recent past. More specifically, I am wondering if there is a way to (indirectly) link the inferred flood frequency to late Pleistocene Indian monsoon variability. Or can we expect the corresponding teleconnection not to play an important role during that period (e.g., due to a suppressed monsoon-desert mechanism)?

Response 11: This is an interesting suggestion that we will look into. Nevertheless, to our knowledge, the Indian summer monsoon is currently not a trigger for precipitation at the eastern Mediterranean. In contrast, the monsoon-desert connection indicates that the monsoon causes air subsidence in the Levant and does not deliver precipitation (Rodwell and Hoskins, 1996; Dayan et al., 2017). Furthermore, the ARST and the RST are seasonally unrelated to the summer Indian monsoon. they are late fall and winter phenomena, when the monsoon is long gone from the Arabian Sea into the southern hemisphere.

Comment 12: L.386: I don't quite get what an "NAO-like periodic component" should be, since the NAO does not have any clear periodicity. In a similar spirit, ll.467-468 claim "quasi-periodic ~3-4 years components, possibly related to the North Atlantic Oscillation", which I am not aware of to exist.

Response 12: It is true that NAO does not have a strong periodic component, however, there are different opinions on that aspect. We will revisit the literature and consider rephrasing accordingly.

Comment 13: L.409: Please check if the reduced recurrence rate is not just due to an increased variance within the considered time window.

Response 13: We will check this following the previous comments made with respect to the recurrence analyses.

Response to technical corrections:

All technical corrections will be corrected accordingly.

Cited references:

Dayan, U., Ricaud, P., Zbinden, R., and Dulac, F.: Atmospheric pollution over the eastern Mediterranean during summer–a review, Atmospheric Chemistry and Physics, 17, 13233-13263, 2017.

Marwan, N., Romano, M.C., Thiel, M., Kurths, J., 2007. Recurrence plots for the analysis of complex systems. Physics reports 438, 237-329.

Rodwell, M. J., & Hoskins, B. J., 1996. Monsoons and the dynamics of deserts. Quarterly Journal of the Royal Meteorological Society, 122(534), 1385-1404.