Climate of the Past Discussions

Clim. Past Discuss., referee comment RC3
https://doi.org/10.5194/cp-2020-161-RC3, 2021

Review of Ben Dor et al. - Methodological aspects

Reik Donner (Referee)

Referee 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-RC3, 2021

The authors present an analysis of annually resolved proxies of hydroclimate variability in the Levant based on sediments from the Dead Sea area. Specifically, they focus in their analysis on two about 700-year long time windows from the late Pleistocene exhibiting opposing hydroclimate trends. This specific focus on nonstationary parts of the records requires a careful choice of time series analysis methods to be employed for analyzing the corresponding variability at interannual to multidecadal time scales. The authors particularly chose a combination of singular spectrum analysis, recurrence analysis and wavelet analysis, all of which provide advanced statistical tools that appear generally suitable for the designated purposes. While focusing on three complementary proxies with regular annual resolution (thicknesses of arangonite and detrital layers and number of detrital sublaminae per year), the authors avoid possible problems with non-uniformly sampled data, which would otherwise require a particular treatment along with the chosen methods.

Since my own expertise concerns primarily the methodological aspects of the presented work, I will particularly focus in my comments on the questions that arise from this perspective. Other minor comments of more technical nature will be provided at the end of this assessment.

General comment: The referencing to the supplementary material is not self-consistent across the manuscript. Please refer to all material in the supplement as “Supplement Section A/B” instead of “Appendix A/B” and also provide a consistent numbering of the supplementary figures (e.g., Fig. A1, Fig. B2, etc., instead of Fig. S1 and Fig. S2-1, which is a bit confusing in the beginning). The same strategy would also simplify the enumeration of equations in the current supplementary material section 2.

Major comments:

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.

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.

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.

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.

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: (i) 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.

(ii) What is the reliability (power) of MWW/AB tests for such small samples of 51 (25?) values only? (iii) 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. (iv) 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.

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. (i) 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.). (ii) 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). (iii) 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. (iv) 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. (v) 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.

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).

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.

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).

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.

Minor comments:

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)?

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.

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

Technical comments:

- L.11: "relying"
- Ll.12,14: the information that two ~700 year long segments are analyzed is duplicate

- Ll.13,15: the unit “Ka” should rather read “ka”, but also later in Section 3.1
- Ll.22-23: “quasi-periodic” has a distinctive meaning in mathematics/physics which differs from what the authors attempt to express here; I suggest replacing this term by “oscillatory”
- Ll.28-29: “freshwater… stems from the interaction of… conditions” is in my opinion an awkward wording in a twofold way. Freshwater is provided by precipitation and hydrology, both of which refer to specific physical processes in the Earth system. Conditions cannot interact with each other; instead, processes can interact.
- L.30: “variability at seasonal,…”
- L.33: I think “configure” means something different than is attempted to be expressed here
- L.36: “transitional states” – do you mean “transitions between states” (my guess) or “states exhibiting frequent transitions” (where “state” would then be meant in a more colloquial way
- L.37: I suggest removing “discrete and” to avoid possible confusion
- L.38: “are harder to determine”
- L.39: “capture the full diversity of possible hydroclimatoc states”
- L.62: “opposing mean climates” does not seem to be the correct wording here, since the authors argue themselves to focus on two “transitional periods” (with opposing hydroclimatic trends; wetting/drying) rather than two “equilibrium periods” (wet/dry)
- L.65: “the periodicity of known global teleconnection patterns”
- L.77: “that are deposited”
- L.146: “where laminae are trimmed”
- L.188: “describes”
- L.205: “principal component”
- Ll.210-211: “the effect of the number of eigenvalues”? (or, rather, the number of “relevant” eigenvalues?)
- L.212: “The SSA RCs (Fig. S6 and S7)…”
- L.217: “Figs. S8-S11”
- Ll.235-236: If the underlying variables have physical units, the skewness values should have, too (consistent with Tab. 1).
- L.237: “substantially significant” is a very imprecise term, rather write “significant” here only.
- Tab. 3: Is it possible to provide actual dates (maybe along with uncertainty) of the identified periods (e.g. in years BP/B2k or others…) instead of “index years”?
- L.293: this should already be section 4.3, use “periodic” instead of “periodical” (also later)
- L.303: “do not pass the significance test”
- Fig. 6 and also several other figure captions: use the symbol \alpha for the confidence level
- L.353: “the Mediterranean”
- L.354: “coupling of observed increased thickness”
- L.367: “distinctively… implications for environmental…”
- L.379: “North Atlantic Oscillation (NAO) and the East Atlantic (EA) pattern are commonly…”
- L.432: “during the winter months”
- L.443: “by two other synoptic”
- Section 7 should be removed, Sections 8-10 not be numbered
- Fig. S2-2, caption: clarify that cumulative distributions are shown
- Fig. S2-3, caption: clarify that rank distributions are shown