This study explores the relative contributions of the driving forces of the monsoon variability with state of the art spectral analysis tools. The authors use two time series as proxies for the monsoon intensity. (1) A mean grain size record from a loess sequence as a proxy for the East Asian winter monsoon intensity (new, high resolution data) and (2) a d18O record from speleothems as a proxy for the summer monsoon intensity. Spectral analysis and visual correlations between the filtered records and the plausible forcings lead the authors to draw conclusions about the various contributions of the forcings at different scales.
I particularly liked the review about the state of the art in that field and I found the introduction quite clear about the objectives of the paper.
The implicit model (between the forcings and the data) behind all the analysis at the basis of the manuscript is a linear one. Indeed, the authors separate the different contributions into 6 IMFs, each one being the result of filtering the time series in a given frequency band. Then, each IMF, or grouped IMFs, is thought to be related to a particular forcing because it is in the same frequency band and is visually correlated to those forcings. This is typical of a linear system in which the main frequencies of the forcing will appear in the response, possibly at a different power. The percentages of the contributions of each forcing to the total also reflects that point of view. A nonlinear point of view could imply, for example, high frequencies of the forcing having an effect on low frequencies in the observed response. This is clearly not an option considered in that study, although the climate system is certainly not linear. I then ask the authors to explicitly write that it is a linear point of view, especially in the abstract and conclusion.

Main comments:
———————

1°) In order that the readers could reproduce the results, I ask the authors to consider the following requests:
- Provide the grain size data in a file, and explain how you made the connection between the two sections (the upper 20m one and the 50m one). Details about the dating procedure must be provided as well, i.e. the details about the matching with CHILOMOS.
- To what I have seen in the reference (Wang, 2008, Nature) given by the authors, the speleothem d18O record from Sanbao/Hulu cave is not a unique time series but it is rather composed multiple fragments. Please explain the procedure to get the time series presented on figure 2. I also ask the authors to provide the data of the full record, presented on figure 2, in a supplementary file.
- Please provide in a supplementary material all the input parameters used for REDFIT, MTM and EMD. If the data were interpolated before MTM, please specify the kind of interpolation as well as the time step. Think that the reader must be able to reproduce figures 3, 4 and 5.

2°) Using both the Lomb-Scargle and MTM tools to perform spectral analysis is a very good choice, but the differences between REDFIT and MTM should be better explained:
MTM pros: trade-off between spectral leakage reduction and variance reduction is quantified and optimized.
MTM cons: the time series is interpolated (NB: « gap filling » thanks to SSA is a kind of interpolation).

REDFIT pros: the data are not interpolated, i.e. it works with raw data.
REDFIT cons: Uses WOSA method for spectral leakage reduction and variance reduction. The trade-off is not quantifiable and is thus not as good as with MTM.

3°) page 6 row 7: « significance test which is not proportional to the power of spectrum » . I don't understand. Isn't it the case with REDFIT as well?

4°) About the Empirical Mode decomposition:
- Please provide the reference for the code you used (reference « Huang et al. 1998 » is not sufficient since it does not provide any code), or provide your code as supplementary material.
- at row 19: « avoiding arbitrariness in the choices of frequency bands ». How does it work? How are determined the 6 frequency bands on figure 4?

5°) Legend for figure 2: What is OSL dating? If a correlation with the Lisiecki and Raymo stack were performed, it should be explained in the main text, together with the dating procedure. The same remark holds for « OSL » dating

6°) Figure 3: Start the frequency scale at 1/[length of the time series] , i.e. at approximately 0.004 kyr^(-1).

7°) Figures 4 and 5. Please provide details in the main text:
- Are the data interpolated before EMD analysis? If yes, how?
- How are computed the spectrum and the confidence level at the right of each IMF?
- Are the results different if the frequency range for each IMF is changed? Please, show it.

8°) Figure 6.
- Apply to the LR04 time series the same filtering procedure as for MGS and Speleothem data, in order to compare signals filtered in the same way.
- In section 4.1 of the manuscript, the authors write that d18O from speleothem « varies quite synchronously with the July insolation ». I agree with that. It's quite clear from a visual inspection on figure 6. They also claim for a good correlation between the LR04 time series and the MGS record. There, a more quantitative analysis than visual inspection is required.

9°) Figure 7.
- Are the three signals identically filtered?
- You write « YD and Heinrich events recorded in the three records ». How? Do they occur EXACTLY at the same age in the time series, as it is implicitly suggested on the figure?
- You write « gray bars indicate interglacial periods » ; interglacial???
- How do you settle the « interstadial numbers »?

10°) A more quantitative analysis is required for the last section (section 4.2 about the millennial oscillations).
- page 11, row 11/12: « … are well aligned with comparable amplitude and duration ». Please perform a quantitative analysis to compare them.
- As indicated above, the detection of the Heinrich and YD events in the records is not clear to me.

Typing comments:

ー ー ー ー ー ー ー ー

- Page 3, row 19: and -> but (?)
- Page 6, row 10: space -> sampled
- Page 8, row 20: >10kyr -> period>10kyr
- Legend of figure 2: Isiecki -> Lisiecki