Review of: "Artifact Subspace Reconstruction (ASR) for electroencephalography artifact removal must be optimized for each unique dataset"

Arefeh Sherafati[1]

1 Washington University, Saint Louis

Potential competing interests: The author(s) declared that no potential competing interests exist.

Summary: This study explores the effect of changing the SD parameter in the Artifact Subspace Reconstruction (ASR) approach for motion removal of EEG data on a previously published dataset. The study explores values of 20 SD, 40 SD, 60 SD, 80 SD, 100 SD, 120 SD for ASR as opposed to the recommended value of SD ~= 20-30 by the authors of the ASR method. The effectiveness of ASR was quantified on the following metrics: the number of dipolar independent components (ICs), model order for multivariate autoregressive modeling, and the number of preserved trials.

Findings:

Contrary to previous literature, the present study shows that the optimal ASR parameter could be substantially higher than 20 to 30 and could be as high as 120, depending on experimenter decisions for what to preserve. The study finds that more ‘aggressive’ parameters for ASR produces a lower model order than less aggressive parameters, which is an important characteristic of rPDC and other information-theoretic techniques. The study also found that there was no clear prediction on the relationship between ASR parameter choice and the number of ICs produced. The author thus concludes that the ASR parameter choice should be justified in each study using quantitative preliminary analysis.

Major Comments:

- This is an interesting and potentially valuable study for better understanding the tradeoff between removing noisy data and reducing the true signal and its effect on the statistical power. However, I think this paper can be significantly improved by investigating the reproducibility of the actual brain responses to the cognitive tasks rather than spending most of its time performing different statistical analysis on the ‘number” remained trials, the “number” of ICs, and the “model order” gained after trial rejection using each of the SD values without investigating the underlying reasons and the resulting brain signals. The study will be a lot more valuable if the main figures would present the resulting brain maps for each of the SD values and presenting statistical analysis and metrics of reproducibility (such as voxel-wise similarity or Dice coefficient) between the actual brain maps and moving most of the current figures to supplementary analysis.

- This general point aside, the discussions are very helpful reflecting an important conclusion that in denoising methods that reject the noisy trials, such as ASR, increasing the SD corresponds to more lenient thresholds which results in keeping more data and hence more noise. The study shows that more lenient thresholds (corresponding to SD values around 120) do not necessarily result in lower quality brain activity.

- The author did acknowledge that the relationship between the number of ICs and the SD threshold is not clear. I think it is worth exploring those remaining components rather than the "number" of them. The quality of the remaining ICs in capturing the important features of the brain activity could also be very important.

- Lastly, for improving the interpretability of the effect of raising the SD threshold in rejecting the trials a simulated dataset could be useful. Specially to identify the reason behind the fact that very high SD numbers did not necessarily result in larger amounts of data.

Minor Comments:

- There are wrong or corrupted links for some of the references. Examples:
  - The following link points to the wrong article: Harvard Automated Processing Pipeline for Electroencephalography (HAPPE) (Gabard-Durnam et al., 2018).

- The following link is not working: data correspond to experiment one inNyhus (2010).

- In the last paragraph of the introduction, there is a type "the lower the lower"
- Check for typos "trails" instead of "trials"