Reply to the referee report

Anh Phan[1]*, Santanu Das[1,2]†, Albert Stebbins[2], Peter Timbie[1]‡, Reza Ansari[3], Shifan Zuo[4], Jixia Li[5], Trevor Oxholm[1], Fengquan Wu[5], Xuelei Chen[5,6,7], Shijie Sun[5,6], Yougang Wang[5], Jiao Zhang[8]

[1] *Department of Physics, University of Wisconsin Madison, 1150 University Ave, Madison WI 53703, USA*
[2] *Fermi National Accelerator Laboratory, P.O. Box 500, Batavia IL 60510-5011, USA*
[3] *Université Paris-Saclay, CNRS/IN2P3, IJCLab, 91405 Orsay, France*
[4] *Department of Astronomy and Tsinghua Center for Astrophysics, Tsinghua University, Beijing 100084, P.R.China*
[5] *National Astronomical Observatory, Chinese Academy of Science,20A Datun Road, Beijing 100101, P. R. China*
[6] *University of Chinese Academy of Sciences Beijing 100049, P. R. China*
[7] *Center of High Energy Physics, Peking University, Beijing 100871, P. R. China*
[8] *School of Physics and Electronics Engineering, Shanxi University, Taiyuan 030006, P. R. China*

15 December 2021

We wish to thank the editor and the referee for carefully going through the paper and for their valuable comments. Here we try to clarify the comments point by point and address all the issues. If the referee has any other doubts, we will be very happy to address those.

1 EDITOR'S COMMENTS

(i) I felt that some points are weak. For example, in the abstract you say that "The algorithm can remove most of the solar contamination": you should express this quantitatively and compare the residuals to the signal you are looking for. Similarly in Section 7 you say "We show that AlogSCR can recover the background signal pretty well" – "pretty well" is not a scientific term.

We remove the terms like "pretty well" and "most" etc. and give numerical quantities. The plots of χ^2 values for different signal strength are shown in the paper to quantify the fraction of the signal that is removed for different signal strengths.

(ii) I would be interested to see a quantitative comparison between day-time data after solar decontamination and night-time data of the same patch of sky.

For comparing between daytime data after solar de-contamination and the night time data of the same patch of the sky, we need to scan the with a 6-months interval. We don't have data which are exactly 6-month from each other. However, we do have some data which are 4 months from each other. We compare data from the common patch of the sky taken on January 2018 after sun removal with the data taken on April 2019. Visually the patches looks similar. An analytical comparison has been provided in the paper.

(iii) Please reword the acknowledgements: "This document prepared by the Tianlai Collaboration includes personnel and uses resources of ...". The document should be prepared by the listed authors. Any "personnel" who contributed significantly to the work should be listed as authors. Direct technical help (e.g., observatory staff) and indirect assistance (e.g., people who reviewed the manuscript) may be acknowledged here. If you want to mention the Tianlai Collaboration you should explain what is is, or give a reference.

We have removed the term "Tianlai Collaboration" from the Acknowledgements. The author list does indeed include all individuals who contributed to the research and the preparation of this manuscript.

(iv) Please refer to the MNRAS Instructions to Authors for formatting the reference list.

We have formatted the reference list according to the MNRAS instructions.

2 REVIEWER'S COMMENTS:

(i) The authors present their method as a narrative, describing each problem they face and their solution to overcome that problem. However, for an experiment that requires to collect and process a large volume of data, I fear their solutions would do more harm than good.

We partially agree with the referee on this point. For a large volume of data, the solar contamination removal process is a highly computational intensive project and requires lots of CPU hours.

However, all the intensity mapping experiments are at present in their pathfinder phases. We don't have sufficient amount of data of the sky. Therefore, if some algorithm allows to recover half of the observed data then it can helpful. The computational time for recovering the solar signal is not significantly high for a modern day computer cluster. In fact, the processing time is significantly smaller than many other astronomical analyses.

Apart from HI intensity mapping, the algorithm can also help many other interferometric studies that requires researchers to separate out bright external sources.

(ii) The key to their method is the smoothing of eigenvectors (appendix A), which they perform by projection onto a Stiefel manifold. That is fine, but my concern is the selection of which eigenvalue to smooth. As the EVD is performed for each t,ν pair, a permutation of the eigenvalue,eigenvector pairs can happen. One can find a unique dominant eigenvalue in most cases, but what if two eigenvalues are almost the same? This problem has not been addressed. The final subtraction (A10) pp. 22, is actually a sort of calibration of the data, so why not use a suitable fixed basis for (A6) and perform this calibration? Starting out with the eigen decomposition, the authors end up at this calibration approach, and there are many ways to perform this smooth calibration, see for example https://academic.oup.com/mnras/article/449/4/4506/1203097

The referee is correct in pointing out that the permutation can happen. Therefore we need to use a fixed basis. However, the principal assumption of the paper is that the Sun signal is far stronger than the other sources. Therefore when we decompose into eigen values, the largest eigenvector always represent the direction of the sun. The direction flipping between the other eigenvalues may happen. However, that is not of much concern as long as the eigenvalue from the Sun is separable. The method is not applicable when the solar signal is comparable to the external sources. From the χ^2 plots of Fig. 24 and Fig. 25 it is clear that if the external source strength is stronger the solar contamination will suffer significantly.

The second option of choosing fixed direction can be that as we know the direction of the sun at any given time on the sky, we can calculate the eigenvector of that particular direction for the Tianlai configuration. Then we can remove the signal from that particular orientation. Even though the process looks of mathematically, the real data the the external noise and external fields. We tried such methods. However, the such solar contamination removal process did not provide fruitful results.

We had also tried SVD decomposition based broad band RFI cleaning methods as descried in the GMRT paper. We tried SVD decomposition. However, such SVD decomposition based method was removing a significant part of the signal from the data. After removing larger SVD modes, the data that we were getting from the daytime data was significantly different from the night-time data Therefore, the SVD decomposition based methods were not giving fruitful results for the tianlai dataset.

We added some of the papers for broad band RFI cleaning for radio astronomy data, in the introduction section.

(iii) The DC offset produced by the correlator is another concern. The authors mention mutual coupling as a possible cause. It could also be a problem at the correlator, such as quantization. The simple removal of the nighttime mean might appear to work, but for a deep experiment such as HI intensity mapping, this solution needs further testing.

The reviewer pointed out that the DC offset is of another concern. We completely agree with his concern. In Tianlai dish array we are getting the DC signals which are probably not coming from the sky. Our instrumentation team is working on sorting out the issues. The exact issue is still unknown. However, our analysis with different strong radio source observation has shown that a simple mean removal fixes the problem.

(iv) The receiver noise component in the correlation matrix is bypassed by using (7) to replace the autocorrelations. But its long term effects remain uninvestigated.

The auto correlation signals are mostly noise dominated signals. Therefore we need to filter out the noise before using the signals in the matrix decomposition. This particular method, even though it is not an exact method, works well with the data. However, at this point its not possible to test whether it will have any other effect on the data or not. For doing that we need to do full map reconstruction and check whether it is adding any additional feature on the power-spectrum or not.

However, it is beyond the scope of this paper.

(v) - The introduction (section 1) gives a detailed description on HI intensity mapping, but not about the actual subject of the paper, i.e., the removal of RFI. There are plenty of related methods, I give one example below:

https://academic.oup.com/mnras/article/413/2/1174/1068136

Also worth mentioning that several different names are used to represent similar methods, such as, singular value decomposition, eigen value decomposition, subspace methods, principle component analysis etc. The authors can look into the related literature using these names.

We discuss about some of the methods in the paper and also

(vi) - Section 2 shows the extent of the problem of solar contamination, in great detail. I am not sure such length is needed, as it is a well known fact that the sun will cause interference in radio telescopes. I guess the sharp drops in amplitude in Fig. 2 can be explained by the Sun going below the edge of the dish. I am confused by the use of 'smooth part' and 'noisy part' do describe parts of the plots shown.

(vii) - Most figures (Fig 2, 7, 9, 15, 17 ...) show mostly the same information, and they can be reduced in the number of panels, and in time duration, making the paper shorter. Fig 3, shows the same plot twice and I don't see the need for this. I think the right hand plot need to be rotated by 90 degrees. Also the axes are missing. Some axes are not readable, like in Fig. 5.

We reduce the number of plots and some of the paragraphs, which were not necessary for the paper.

Fig. 3 of the previous draft has been removed. The axis were in correct direction. However, as it was a 3D plot, it was appearing that there was a 90 degree rotation.

(viii) - Notation: I would use H and V to denote the voltages and HH and VV for the correlations. It would also help if the dipole orientation is shown on Fig. 1. As I understand H is aligned north-south and V is aligned east-west, but this is not clear to me. In pp. 5, equations (1),(2),(3).. have some undefined notation such as D^s, $[]^\dagger$, $()^\star$

We address the issues.

(ix) Not using the cross-polarization correlations (HV and VH) are also not explained. Is it because the correlator does not produce them or is it that they are ignored only at this data processing step? As the Sun is highly polarized, the use of polarization data would improve the removal of the contamination, see for example https://ieeexplore.ieee.org/abstract/document/9287467

The correlator does produce the cross polarization signals. Provided we use both the polarizations together, we can expect to get 2 eigenvalues. However, right now we are in process of understandin the cross polarization signals, which are much difficult than the auto polarization signals. for example, the cross polarization signals between the two feeds of the same dish are much stronger than the other signals. This may happen from some coupling due to the proximity between the feeds. However, we need to understand how to fix the signals.

Unlike the auto-correlation signals, these signals are complex quantity and the phase matters. An equation similar to Eq.7 will not be able to act as a proxy for the actual value. This is because a slight change in phase will change all the eigen values. Therefore, we decided not to use the cross polarization signals in the present analysis.

(x) - The title (and also some parts of the text) suggest that the method presented is generalizable and can be used in other radio telescopes as well. But looking into the details and the concerns raised above, I do not see that yet. So I suggest the title should indicate that this method is especially designed to the Tianlai telescope.

We do think that AlgoSCR will work on any type of interferometer.

(xi) pp. 2, [Ali et al. 2015] is a retracted paper, and I dont see any reason for citing it as there are much better EoR results by LOFAR and MWA.

We thank the referee for pointing out the issue. We remove the paper and add other papers.