March 13, 2020

To the Editors of *Cerebral Cortex*:

In February 2019 we submitted a prior version of the enclosed manuscript (CerCor-2019-00027), entitled *A Gaussian process model of human electrocorticographic data*. We have carried out an extensive revision of our manuscript in response to the reviewers' insightful comments (detailed point-by-point responses are on the following pages; reviewer comments are bolded and our responses are in italics).

In our prior submission, we presented a model for reconstructing full-brain activity patterns from a limited number of ECoG electrodes. While both reviewers indicated general enthusiasm for the work, Reviewer 2 also raised some important questions about the ability of our approach to specifically recover activity patterns within different frequency bands. In our revised manuscript, in addition to addressing both reviewers' other concerns, we have carried out a series of detailed analyses evaluating reconstruction accuracy for activity patterns within different frequency bands (as well as broadband activity). We show that we are able to recover activity accurately at a wide range of frequencies. One of the most interesting new findings to come from these new analyses concerns the interpretation of different types of neural patterns. Prior work has largely treated region-specific narrowband and broadband activity as an indicator that activity at those frequency ranges reflects that the given region is representing or supporting a particular function. Our work suggests an alternative interpretation that when we observe a particular neural pattern in a particular brain region, it may instead reflect how that region is transmitting information to the rest of the brain via signalling at the given frequency range.

We appreciate the opportunity to submit a revised version of our manuscript. In addressing the reviewers' concerns, we believe that our manuscript has been substantially improved. We hope that you now find it suitable for publication.

Sincerely,

Jeremy R. Manning, Ph.D.

Reviewer 1 Comments

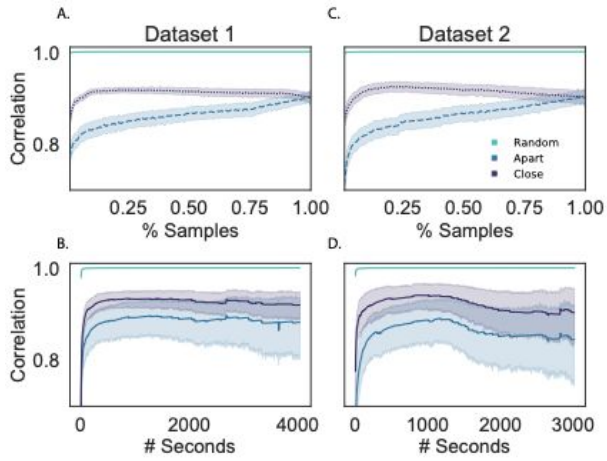
In this submission the authors propose a method that aggregates iEEG recordings from multiple individuals (limited to select brain areas) and uses a statistical model to infer activity at unobserved locations and, more interestingly, the entirety of the brain. This approach overcomes issues with iEEG including sparse brain coverage and interindividual differences in electrode location while preserving its unique combination of high temporal/spatial specificity. My overall impression of this article is that its topic is timely, it is extremely well-written and cogent, and makes important methodological contributions to computational and clinical neuroscience that open up multiple avenues for future research.

In general, my feeling is that this submission needs very little work to be considered publication-ready. I have a few comments and clarifications and, provided that the authors can address this adequately, I'm happy to recommend this article for publication.

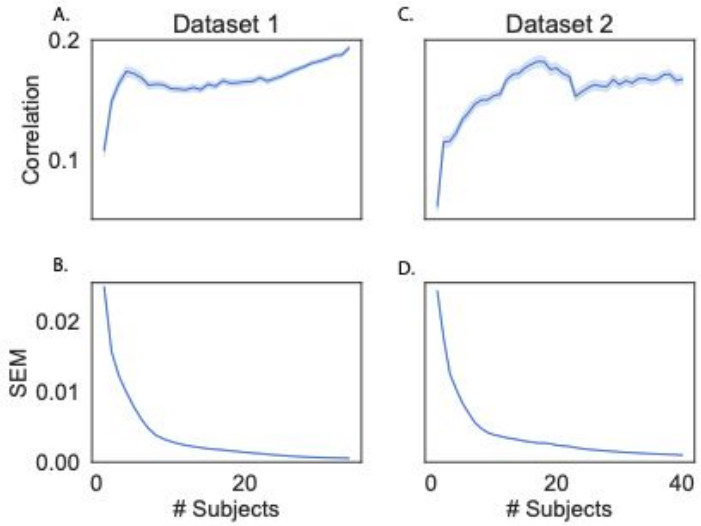
Thank you for the positive feedback, and we hope that our changes have adequately addressed your remaining concerns.

Comments:
p. 4 – the authors leverage a fairly unique datasets – ECoG data from many subjects recorded over a long period of time. I'm curious about how much data is necessary for "SuperEEG" to perform as well as it does in the paper. It would be useful to see a figure analogous to Fig. 4 in Laumann et al (2015) in which some measure of reconstruction error is plotted as a function of number of subjects/duration of recording.

The reviewer raises an important question. We explored this in two ways. In our new Supplemental Figure S7, we examined the stability of the patient-specific correlation matrices we estimated, as a function of the proportion of each patient's data (top row) or as a function of the total duration of the recordings (bottom row). In summary, we found that, with roughly 5--10 minutes of data (sampled at 250 Hz), our approach produces estimates that were correlated at above r = 0.9:

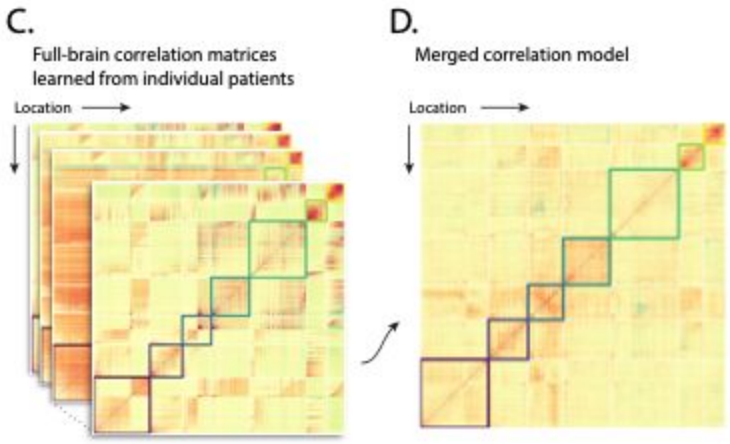


In a second analysis, displayed in Supplemental Figure S8, we compared our algorithm's estimates of full-brain correlation matrices using different numbers of (randomly selected) patients. We examined both the similarities of these matrices (measured as the correlation between the upper triangle of the full-brain correlation matrices estimated from non-overlapping subsets of patients) and the stability of the estimates (standard error of the mean of the correlations between the matrices estimated for different subsets of patients). The similarities between the matrices plateaued at around 10 patients for Dataset 1 and 20 patients for Dataset 2:



p. 6 – the authors use a monotonically-decaying function (that decays to 0) to describe the influence of electrodes on the activity at some point. I'm wondering if this assumption is sufficiently general to capture some known features of observed brain activity correlation structure. Specifically, I'm concerned that the function, which decays to zero and is always positive, limits the frequency with which we observe anticorrelations which are known to exist for certain brain systems, e.g. DMN with dorsal attention. Can the authors provide some clarification on this?

The reviewer is correct that the activity patterns recorded from different electrodes can sometimes be meaningfully anticorrelated. Indeed, while less common than positive correlations in our data, we frequently observed anticorrelated activity patterns in both of our ECoG datasets (e.g., see the blue-green colored parts of the correlation matrices in Fig. 1, Panels C and D; in those matrices, reddish colors denote positive correlations and blue-green colors denote negative correlations):



The key question the reviewer is getting at is about how these anticorrelations observed in the data from individual patients are treated when we estimate full-brain correlation matrices.

The weighting function the reviewer references is used to "blur out" the patient-specific correlation matrices (capturing correlations in activity patterns at the locations of that patient's recordings) to the full set of locations we are modeling. Although the weighting function is always non-negative, the values of the observed patient-specific correlation matrices can be (and often are) negative. Therefore our estimated full-brain correlation matrices contain both positive and negative values. Further, the balance of positive versus negative correlations in the estimated full-brain correlation matrices matches (within the constraints of the patients' electrode placements) the spatial profile of the original data.

p. 6 – If I understand correctly, the same function also presupposes that brain activity varies smoothly over the cortex, so that the influence of nearby brain areas is greater than distant. I'm wondering whether this feature fails to account for other well-known features of correlated brain activity, namely the abrupt variation in regional/voxel correlation patterns near the boundaries of systems, e.g. the 7 networks defined in Yeo et al. Near those boundaries, small variations in spatial location can result in large differences in voxel-level correlation patterns. Are reconstruction errors greater near these boundaries? Can the authors provide some additional clarification?

The reviewer's point that brain anatomy can vary abruptly is well-taken. The smoothly varying spatial function we use to estimate full-brain correlation patterns cannot (in and of itself) account for those spatially abrupt changes in connectivity (or, in our case, correlations). However, it is important to clarify how the smooth spatial function affects the estimates of activity that we are generating.

The spatial blurring process (governed by the smooth function that the reviewer references) is used to estimate correlations at locations where we have no data from the given patient. In other locations, where the patient's brain is sampled, we can (and do) observe spatially abrupt changes in correlation profiles. Our approach in the current manuscript is to treat spatial smoothness as the "prior" (all else being equal). However, when the data indicate sharper transitions, the full-brain correlation matrices we estimate can retain them (also see pages 6--7 of our revised manuscript).

In future work, our approach might be further refined to take into account known anatomical properties of the human connectome (e.g. learned from DTI data). We suspect that this could enhance the predictive accuracy of SuperEEG, although we have not yet tested this claim.

p. 7 – I disagree with the claim that correlation structure is a sufficient stand-in for anatomical connectivity. The paper that the authors cite are only able to obtain high structure-function correspondence by choosing a transformation (a rotation, actually) of matrices that maximizes a correspondence of eigenmodes. Many studies using human and non-human imaging data have found a broad correspondence between anatomical connectivity and correlation structure (see, for example, Adachi et al 2011, Honey et al 2009, Goñi et al 2014) but that there are many instances of strong correlations in the absence of connections and vice versa. It might be useful if the authors revised this statement so that correlation structure and anatomical connectivity are not equated.

This is a fair point. We have added the following statement (p. 7): "Although anatomical and functional correlations can differ markedly (e.g. Adachi et al., 2011; Honey et al., 2009; Goñi et al., 2014), we use temporal correlations in the data to estimate these anatomical connections (Becker et al., 2018)."

p. 10 – for whole-brain correlation matrices, it would be useful to show some of the stronger positive/negative correlations in anatomical space. An additional supplementary figure would help.

We have added two additional figures (Figs. 6 and S5) that display brain maps of the regions with the highest information scores (i.e. strongest correlations; p. 17). We repeated this analysis for activity patterns in different frequency bands and found that the maps vary depending on which aspects of activity are being considered (i.e. which range of frequencies). We have pasted below an example map displaying the regions with the highest information scores (when considering the raw voltages; see Figs. 6 and S5 for additional maps and figure legend):

p. 10-11 – surely it must be true on average, but how similar are activity traces over long time periods (Fig2A shows 5 s only). Are the errors in reconstruction more or less uniformly distributed across time? Is error density greater at certain points? If so, why?

How accurately we can reconstruct activity patterns depends on our ability to accurately (and stably) estimate the underlying correlation matrices. We therefore examined how well we could estimate these correlation matrices under different conditions (see analyses reported in Supp. Figs. S7 and S8; also pasted above in response to a previous comment).

In summary, we found that, over timescales of (roughly) hours, there are reliable differences in the structure of the full-brain correlation matrices that drive our approach (e.g., see blue curves in Panels B and D of Supp. Fig. S7). However, the magnitudes of these differences seem to be relatively small-- we're seeing an average change in correlation from roughly r = 0.8 (for data sampled roughly an hour apart) to r = 0.9 (for nearby data).

Whether particular timepoints in the recordings are associated with higher (or lower) errors is an interesting question. We explored this informally by spot-checking data from several patients and did not notice any obvious pattern. However, we would not be surprised if reconstruction accuracy varied meaningfully with behavior. For example, if the patient's brain patterns were to change abruptly (e.g., during an event boundary or another sharply defined moment in the experiment), our use of a single (stable) full-brain correlation matrix to reconstruct their activity patterns both before and after that change point may be sub-optimal. One could imagine (in future work) extending our approach to explicitly detect event boundaries using Hidden Markov Models (e.g. following Baldassano et al., 2017), and then separately fit the full-brain correlation matrix within each event.

Fig 2. – the r = 0.91 correlation is strong, but given the implication that the method proposed here might have some clinical utility, is that correlation good enough? Similarly, Figure 3 suggests that out of sample correlations are good but not exceptional. It would be interesting to hear the authors' perspectives on this.

First, we note that we have removed Fig. 2 following a suggestion from Reviewer 2. Here the reviewer is raising an important point: how good is "good enough?" One evaluation criteria might be to compare our approach to the current "gold standard" approaches employed by clinicians. The current gold standard is to simply ignore brain regions that have not been recorded from. By this criteria, "some information" may be better than "no information" about what those unrecorded areas are doing. Another evaluation criteria might be to ask whether the reconstructions we are generating have some (clinical) predictive value. Although this question is beyond the scope of our current manuscript (where we are presenting SuperEEG as a "proof of concept" approach to introduce and evaluate the method), in other work we are beginning to explore clinical applications. For example, in a recent preprint we use the SuperEEG reconstructions to predict depression symptoms (Scangos et al., 2020, bioRxiv; https://doi.org/10.1101/2020.02.14.943118).

p. 15 – typo. “covererage” should read “coverage.”

Fixed.

General comment – All of the data analyzed here came from subjects with pharmacologically resistant epilepsy. This limitation isn’t discussed but probably should be. It might also be interesting to discuss what this implies for our understanding of how brains of otherwise healthy individuals are organized (I recognize that this requires the authors to speculate a bit).

This is an important point for all ECoG studies (including ours). We have added the following paragraph to our discussion section (p. 26):

> *A second potential limitation of our approach is that it relies on ECoG data from epilepsy patients. Recent work comparing functional correlations in epilepsy patients (measured using ECoG) and healthy individuals (measured using fMRI) suggests that there are gross similarities between these populations (e.g., Kucyi et al. 2018; Reddy et al. 2018). Nevertheless, because all of the patients we examined have drug-resistant epilepsy, it remains uncertain how generally the findings reported here might apply more broadly to the population at large (e.g., non-clinical populations).*

General comment – I think that the discussion section needs at least one (maybe two) additional paragraphs. First, I know that this is a methodological paper, but I’m left wondering what’s the new bit of information about brain organization or function that I’ve learned. Is it that brains are pretty similar across individuals (they must be, otherwise you couldn’t accurately infer missing observations from one subject using data from other subjects). I feel like it would be useful to add a paragraph that clearly explains what is the new neuroscience that we glean from the reported findings.

We have added two paragraphs (one to the results section on p. 23 and another to the discussion section on p. 25). The new paragraph in the results section highlights a potential interpretation of our new frequency-specific findings and the new paragraph in the discussion section highlights implications for how we consider ECoG data more broadly:

> *[FROM RESULTS] Taken together, the frequency-specific information maps suggest a potential new interpretation of many of the above previously reported findings. Prior work has largely treated region-specific narrowband and broadband activity as an indicator that activity at those frequency ranges reflects that the given region is representing or supporting a particular function. Our work suggests an alternative interpretation that when we observe a particular neural pattern in a particular brain region, it may instead reflect how that region is transmitting information to the rest of the brain via signalling at the given*
> *frequency range.*

[FROM DISCUSSION] By providing a means of reconstructing full-brain activity patterns, the SuperEEG approach maps ECoG recordings from different patients into a common neural space, despite that different patients' electrodes were implanted in different locations. This feature of our approach enables across-patient ECoG studies, analogous to across-subject fMRI studies (e.g., Haxby et al. 2001, 2011; Norman et al. 2006). Whereas the focus of this manuscript is to specifically evaluate which aspects of neural activity patterns SuperEEG recovers well (or poorly), in parallel work we are training across-patient classifiers by leveraging the common neural spaces obtained by applying SuperEEG to multi-patient ECoG data. For example, we have shown that SuperEEG-derived activity patterns may be used to accurately predict psychiatric conditions such as depression (Scangos et al. 2020). Analogous approaches could in principle be used to develop improved brain-computer interfaces and/or to carry out other analyses that would benefit from high spatiotemporal resolution full-brain data in individuals, projected into a common ECoG space across people.

A second paragraph that might be useful to add would be a set of future experiments that could be carried out using this new technique that otherwise wouldn't be possible to carry out. For instance, comparing whole-brain correlations patterns during free recall tasks (this is a boring example, but you know what I mean). My sense is that the paper would benefit from a sort of "future directions" section that, again, shifts the story back to the neuroscience.

The discussion paragraph we referenced in response to the previous comment is also intended to address this one. In summary, we see our approach as facilitating a range of across-subject analyses in the ECoG domain. This could benefit ECoG-based neural decoding applications, clinical applications, and others.

Reviewer 2 Comments

In the present manuscript, Owen et al. introduce a method that infers the whole-brain activity profile of data that has been spatially sparsely sampled by means of intracranial EEG in epilepsy patients. While the algorithm sounds exciting and the methods are state-of-the-art and are well implemented, this manuscript has also the potential to lead the field of intracranial EEG astray. In their abstract, the authors state: 'Our approach assumes that different people's brains exhibit similar correlational structure, and that activity and correlation patterns vary smoothly over space.'

This statement is simply not correct and heavily influenced by MRI network analyses (where spatial smoothing is commonly applied). Yes, at the macro level (and at low timescales) it might be true, but the power of intracranial recordings is the exceptional spatio-temporal resolution. E.g. the Chang lab showed that there is independent information that can be decoded when 3 mm grids are used instead of the typical 1 cm grids. Hence, there is independent information even in immediate proximity and spatial smoothing is not the most appropriate method to analyze intracranial EEG data. Figure 1A nicely illustrates this issue. While fMRI suggested that the brain is organized into a handful of stable networks, intracranial EEG without smoothing simply shows a very different pattern. The authors might want to consult some of the Parvizi lab papers (e.g. Foster or Kucyi et al.), which clearly show that adjacent electrodes can have very different response profiles. Similar findings have been reported by the Knight Lab (e.g. Szczepanski et al. or Johnson et al.).

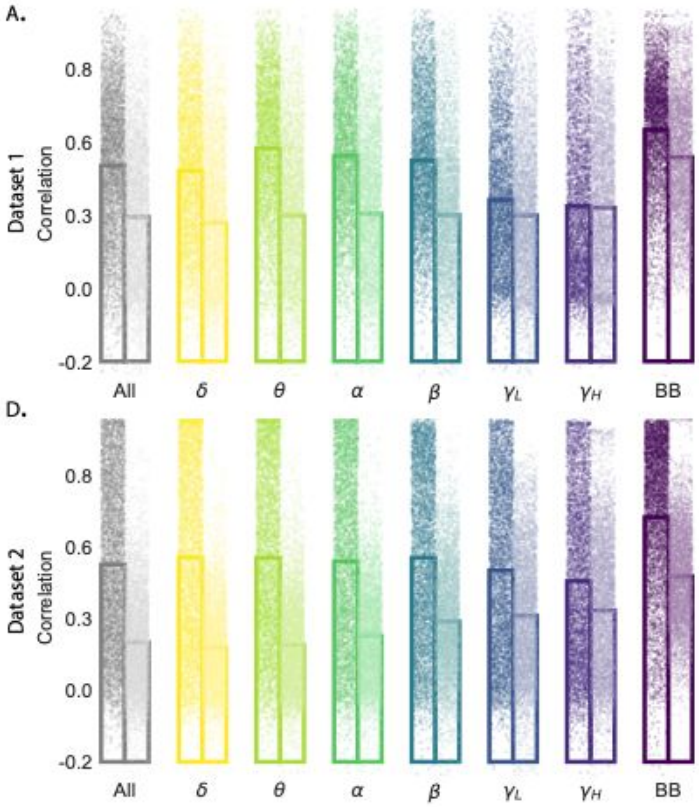
The reviewer raises some important points. First, we have modified the sentence in the abstract to "Our approach ***makes the simplifying assumptions*** *that different people's brains exhibit similar correlational structure, and that activity and correlation patterns vary smoothly over space." We hope that this clarifies that we do not believe these assumptions to be true in their strictest sense; rather we are testing whether (in spite of their inaccuracies) these assumptions can be leveraged to predict held-out brain data.*

The reviewer's point about sharp changes in response profiles (or connectivity) across nearby electrodes is also well-taken, and one also raised by Reviewer 1. Our approach in the current manuscript is to treat spatial smoothness as the "prior" (all else being equal). However, when the data indicate sharper transitions, the full-brain correlation matrices we estimate can retain them (also see pages 6--7 of our revised manuscript).

In future work, our approach might be further refined to take into account known anatomical properties of the human connectome (e.g. learned from DTI data). We suspect that this could enhance the predictive accuracy of SuperEEG, although we have not yet tested this claim.

How well does this algorithm perform as a function of frequency-band? I.e. can the author infer motor cortex high gamma activity in a subject with limited spatial sampling of the MTL that is also predictive of behavior on a trial-by-trial basis?

This is a great suggestion. We carried out a series of analyses to examine these questions (Figs. 4, 6, S4, and S5). In summary, our approach recovers activity well across the power spectrum, although we recover low-frequency and broadband patterns better than high-frequency patterns:



We also carried out analyses to examine which brain regions were most informative about full-brain activity patterns in different frequency bands (Figs. 6 and S5).

How does the reconstruction scale as a function of distance to the sampled region? I.e. is it easier to reconstruct activity the frontal lobe than in the visual cortex when the patient is implanted with a typical 64 channel temporal grid? Does electrode coverage matter for classification accuracy? In other words, are there certain key nodes that need to be sampled to accurately reconstruct activity in a given region (a region-by-region plot would help)? Figure 5 seems to address this issue, but the effect seems smeared and difficult to localize. In particular, the correspondence of the ventral view (plots on the lower right) in Figure 4B and 5B seem to indicate that sampling does in fact matter, i.e. densely sampled regions are easier to reconstruct (Figure S5 highlights the hippocampus, where probably most contacts are located already). Any differences between sensory and higher-order cortex? Are there differences between depth and grid electrodes?

These are important questions. We explored them in Figures 2, 3, 5, 6, and S5 in our revised manuscript. Figure 2 (Panels C and D) displays reconstruction accuracy by location. As shown in the figure, reconstruction accuracy varies across the brain:

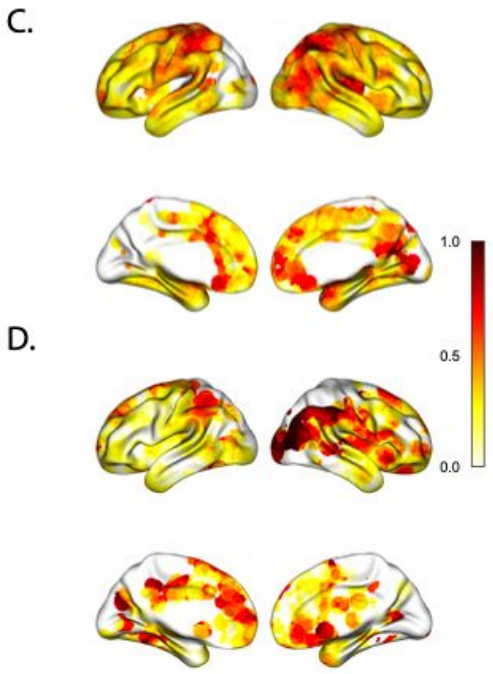


Figure 3 summarizes the electrode sampling densities in both datasets (i.e., how many electrodes are contained within 20 MNI units of the given voxel). We found no reliable positive correlations between sampling density and reconstruction accuracy (p. 16; Dataset 1: r = 0.05, p = 0.70; Dataset 2: r = -0.21, p = 0.05).

We also examined which regions appeared to be particularly informative with respect to reconstructing activity patterns throughout the rest of the brain. In other words, were there some locations that (if implanted) yielded particularly good reconstruction accuracy at other locations? We developed an ***information score*** *(p. 17) to characterize how informative a given recording location was about activity in the rest of the brain. In Figures 5, 6, and S5 we display maps of the most informative recording locations about different types of neural activity patterns. In summary, we found that brain regions that are highly (anatomically) interconnected with the rest of the brain appear to be particularly informative about full-brain activity patterns.*

Furthermore, it is unclear why the authors believe that exclusion of epileptic electrodes is sufficient. Epilepsy (just like cognition) operates in networks, hence, IEDs can observed in epileptic tissue (e.g. the MTL), but are often accompanied by IEDs in inter-connected regions, such as the OFC or the ACC. How is that reflected in the algorithm? In other words, how would an excluded channel look like after reconstruction? Can this tool be used to infer epileptic sources?

We agree in principle that epileptiform activity is not confined solely to epileptic tissue, and we also agree that our approach (i.e., excluding epileptiform tissue identified by our clinical partners and kurtosis thresholding) may fail to capture some epileptic activity. While our approach does enable us to predict activity patterns at these excluded locations, we cannot utilize the same comparisons to "ground truth" recordings at those locations that we employed elsewhere in the manuscript. For example, the signal quality for epileptiform activity is often poor due to amplifier ceiling effects.

We think the best way to approach the question of how well SuperEEG might be used to infer epileptic foci is to reconstruct full-brain activity patterns and then to apply standard neurological tools for seizure identification to those reconstructed patterns. This allows us to track the onset and spread of seizures in 3D through the brain as it develops. We can then use these reconstructions to predict surgical outcomes (and compare those predictions to ground truth observations from the same patients). These approaches are beyond the scope of the current manuscript, which is intended to specifically focus on reconstruction quality. However, we are currently exploring seizure activity patterns in follow-up work.

The figures seem sub-optimal: I.e. Figure 2A looks like a cherry-picked example and is not informative, while e.g. recons as shown in Fig 4/5 look worse than a beamformer from 64 scalp electrodes. Here, it'd be crucial to demonstrate that the algorithm can reconstruct high frequency activity, or activity in deep brain structures, which are simply not accessible by scalp M/EEG.

We agree that Figure 2 was ineffective and we have removed it from our revised manuscript. We wish to clarify that the map in Figure 4 in our previous manuscript (Fig. 3 in the revised manuscript) displays the electrode sampling density-- it is not analogous to a beamforming source reconstruction (it is simply a summary of our data).

We explored the question of how well we are able to reconstruct high frequency activity patterns (e.g. gamma activity) that are filtered out by the skull in a series of new analyses (Figs. 4 and S4). True versus reconstructed gamma patterns are correlated at around r = 0.4 on average (the relevant panels from Fig. 4 are also pasted above in response to a prior comment).

Taken together, the proposed method seems to be technical sound, but it is unclear if the reconstructed patterns hold any behaviorally-relevant information and critical information is missing. For now, it seems to be more of a toy than a tool and it is unclear if it actually outperforms a beamformed high-density scalp EEG. The most crucial aspect is whether any behaviorally-relevant or clinically-relevant information can be reconstructed. I.e. can ERPs, low frequency power or even high gamma be reconstructed.

The reviewer is raising two important points here. First, we explored the question of how well we are reconstructing activity patterns in different frequency bands (Figs. 4 and S4). The relevant

panels from Figure 4 are pasted above in response to a prior comment. In summary, we examined raw voltage patterns, bandpass-filtered signals at different frequency bands, power within different frequency bands, and broadband power. Our algorithm reconstructs all of these activity patterns well, although it performs best in the (roughly) 4--12 Hz range.

Second, the question of clinical relevance requires some further exploration, although our preliminary results have been promising. For example, in parallel work (now cited in our manuscript) we used the SuperEEG approach to predict depressive symptoms in a similar population of patients with drug-resistent epilepsy (Scangos et al., 2020, bioRxiv; https://doi.org/10.1101/2020.02.14.943118). In ongoing work we are also using SuperEEG to predict surgical outcomes. Ultimately, as with any new approach, the clinical utility of SuperEEG will need to be tested empirically for each prospective application. Certainly our hope is that our work will have widespread clinical relevance, but our focus in the current manuscript is on specifically evaluating the method.