# Responses to reviewer comments

## Reviewer #1.

I think several parts of the paper could be written more succinctly. For example the presentation of K-fold cross-validation is fairly standard and could be omitted or at least moved in the methods section. The abstract could also be compressed into a single more compact paragraph.  Overall the main message of the paper could be delivered more briefly.

In the revised manuscript, we removed the equations defining the cross-validation loss and the relative cross-validation loss. Other changes, too numerous to list here, were also made to shorten the paper by 3 pages.

For readers unfamiliar with the relationship between partial correlations, precision matrices, and covariance matrices, it might be worth expanding the text around eqs (3-4) a bit.

In the revised manuscript, we referenced a textbook and an additional reference (ref1, ref2) that provide detailed explanations of pairwise partial correlations. Since these are commonly used concepts, we believe it is sufficient to provide these pointers.

I didn't find Fig 4G-I very interpretable or helpful.  The authors should try to improve these panels, or drop them.

We agree that drawing quantitative conclusions from the 3D connectivity plots depicted in Fig. 4G-I. The plots convey a visual impression of the connectivity between cells inferred by our method. Plot 4G is meant to convey that, even with only ~10% connectivity of 300 cells, the connectivity appears dense. These panels convey the main goals of our study: to infer the functional connectivity between pairs of cells with known positions and tuning properties. Based on these considerations, we decided to keep panels 4G-I in the paper.

The authors select the best sparsity and rank hyper parameters by cross-validation.  It would be nice to see how sharply peaked the corresponding objective function is.  For example, Fig 5 is hard to interpret if we don't know how sharply the obtained degree and latent rank are constrained by the observed data.

We have examined plots of cross-validated loss as a function of the hyperparameters. These plots are helpful in understanding the performance of the algorithm. However, they do not provide a clear representation of how well the choice of the hyperparameter values is constrained by the data. No sharp peaks appear in these plots. This does not mean, however, that the data does not constrain the choice of the hyperparameters. We did not deem these plots worthy of inclusion in the paper.

Fig 6: the definition/interpretation of "normalized connectivity" was a bit hard to understand.  Could the authors clarify this a bit?

We have added a more detailed explanation of “normalized connectivity” in the text on page 10.

The authors argue in the discussion that inference of conditional dependencies between two neurons requires the observation of all the interacting neurons. This is intuitive but I was wondering to what extent the latent + sparse model can also account for unobserved units as suggested in [49] for gaussian graphical models. This could also be interesting to test either in the simulated or the real data by deliberately leaving some neurons out.

This is an interesting idea, which applies only in idealized conditions that cannot be assumed for empirical data. Indeed, in a system of linearly interacting variables where most of them are observed and only a few are latent, given a sufficiently large sample, the spase+latent model would likely uncover the correct number of latent variables. However, in two-photon recordings, removing a few neurons from the recording does not translate into increased numbers and influence of latent units. As shown in Figure 6, larger populations produce larger numbers of latent units. This is because larger populations provide more data from which more variables can be inferred and larger populations interact with a larger number of latent units. Therefore, intuitions that apply to simple system such those described in theoretical papers break down in studies of the neocortex.

However, repeating the analysis on various subsets of the recorded neurons can be useful to demonstrate that the recovered sparse pairwise interactions are robust, or at least more robust than the partial interactions computed without regularization.

It would also be interesting to repeat the analysis but conditioning on a specific stimulus each time. In this case, I'd expect that the latent part will be smaller for each stimulus (since the common input will be restricted to a specific stimulus). But it would be interesting whether the sparse part will remain more or less unchanged, something that will be more indicative of actual anatomical connectivity.

The number of latent units and the number of pairwise interactions inferred by the method increase with the sample size: larger samples support the inference of larger number of model parameters.

In the papers where the low rank + sparse matrix decomposition method is discussed [49,50], a convex framework is proposed where the l\_1 and the nuclear norm of the matrix are penalized. The authors do cross-validation over all possible values instead of using the nuclear norm, which can reduce shrinkage but leads to a non-convex problem. Is there a specific reason for doing so? (In fact the optimization here would also be computationally more efficient since the low rank part is always a positive semidefinite matrix and thus its nuclear norm would be equal to just the trace.)

Similarly, for l\_1 regularization the authors first determine the non-zero pattern with l\_1 regularization and then do an unregularized regression restricted to that set to find the correlation values. Why is that so?

The link to the matlab code appears to be broken.

## Reviewer #2

1. The correlation models differ in the number of parameters, and therefore the merits of different models in predicting the correlation structure of the data might need to be weighted by them. However, while comparing the models presented, the number of parameters is not taken into account. I think that this is quite important, and therefore, using a method such as Akaike information criterion or similar might add some generality to the interpretation of the results of the present manuscript.

Model selection is an essential problem and many approaches have been devised. In this paper, we used cross-validation. We agree that the Akaike information criterion or other selection criteria could be used as well. We note, however, that cross-validation is more general and makes fewer assumptions about the data generating process. Indeed in the limit of large sample sizes, the two approaches have been proven to be equivalent (cite). With small sample size, cross-validation is the preferred method whenever computationally feasible. We now make this point explicitly in the discussion.

2. In the first part of the paper, data are generated through simulations using a multivariate Gaussian model. However, this model cannot fully represent the structure of the mouse visual cortex data that is modeled later. I think it would be interesting to reproduce Fig. 1 (rows 5 and 6) with a spiking generative model, such a correlated Poisson population, with different correlation structures. For instance, as a factor model, one could consider a Poisson model of independent neurons plus one or more mother spike trains that are summed up to the uncorrelated Poisson spike trains. This kind of analysis would be interesting to understand whether the reported results are robust against non-negativity of the firing rates and against large departures from Gaussianity (e.g. occurring at low firing rates).

3. Finally, although the authors make predictions about the functional connectivity that is present in mouse visual cortex, a comparison with what it is known anatomically in cortex is not present.

The distance dependence of synaptic connectivity has been studied by several investigators.