

1 We thank the reviewers for their constructive feedback, and were grateful for positive appraisals and helpful comments.  
2 However, we feel that **review #4** contained a significant misinterpretation of the existing literature and its relation to our  
3 work. We would thus humbly ask that **R4** consider adjusting their score, if they are convinced by our response below.

4 **R1: Is the length scale a user-determined constant?** No, GP length scales were inferred by maximization of the ELBO.  
5 This is an important point: we typically found longer length scales for the noise GPs than the signal GP (for both mouse  
6 and monkey data), suggesting the trial-to-trial variability has slower timescale than the stimulus-evoked component.

7 *Citation for Fourier-domain GPs.* Thanks for the reference; now cited. However, we would like to emphasize that (as  
8 the reviewer points out) this is the first time (to our knowledge) that this method has been used for BBVI, and we have  
9 found that the pruned Fourier-domain variational distribution substantially stabilizes and speeds up inference. It also  
10 preserves temporal correlations (off-diagonal elements of the time-domain covariance), allowing for smooth estimates  
11 of the time-series with few BBVI samples.

12 *Comparison to Duncker & Sahani (DS).* R1 correctly points out DS includes a trial-varying GP component. However,  
13 DS's model was designed to understand time-warping in a GP framework for a reaching task. Each trial included a  
14 time-warping function that allowed the latents to flexibly adapt to different behaviors. Here, our model has distinct  
15 dimensionalities for signal and noise subspaces to understand trial-varying activity in conditions with exact repeated  
16 stimuli. Our model is used to analyze these independently determined signal and noise subspaces. Further, DS does not  
17 adapt methods to understand trial-varying properties of multi-region data. We will clarify these points in the manuscript.

18 *Light on experimental results.* See R2 comments below and added Fig.

19 *Noise+sig subspace comparison to overall noise L2.* This is a good point and we will add a comparison to the figure.  
20 R2 is correct that higher L2 norm in the noise latents on average corresponds to anti-alignment with the signal subspace  
21 (though this doesn't always hold). Also, we disagree with R2's conclusion. When the L2 norm is high, the noise is  
22 anti-aligned with signal, and when it is low, it is slightly aligned with signal. This means even if interference is roughly  
23 constant throughout the trial, there is an overall minimization of noise interference with the signal subspace.

24 *Regularizing W.* We have actually tried this and found no meaningful difference in the learned latent dimensionality.  
25 Smaller W entries traded off with higher-variance latent trajectories.

26 **R2: "authors tried to claim too much ... that noise tends to be primarily orthogonal with no caveat that this was  
27 true in only one dataset".** We agree that our claims are too bold. We have updated the wording in the manuscript.  
28 We claim that our model suggests that in monkey V1 data, during stimulus presentation times, the signal and noise  
29 subspaces tend to be more orthogonal than expected by chance, and that in rodent V1 noise structure is shared across  
30 regions. We now make a stronger call for future directions investigating these types of claims using this model.

31 *"Concern about only selecting high rate neurons".* We agree we failed  
32 to clarify this pre-processing step. This choice was primarily motivated  
33 by the fact that many neurons fired  $< 1$ Hz, and thus contributed virtually  
34 no trial-varying information (sometimes firing only 3-4 spikes for the  
35 entire experiment.) However, the model fits are fine even using these  
36 uninformative neurons - rates are typically estimated as negligibly small.  
37 The choice simply allowed for rapid inference with similar results.

38 *"Opportunities for more rigorous analysis and reorg of fig 2 and 4".*  
39 This is very helpful and we agree. We have replaced Fig 2 and now  
40 show performance of the SNP-GPFA model on simulated data. We  
41 moved the former Fig 2 to the supplement. We instead now show BBVI has accurate recovery of signal and noise  
42 latent structure (Fig R1, A and B) and recovery of neural rates (not shown, but fits are similar to P-GPFA fits in Fig 2)  
43 on simulated data. We will also include panels showing recovered signal structure in SNP-GPFA is nearly identical  
44 standard P-GPFA latent structure.  
45

46 **R4:** The primary concern seems to be that we do not compare to Ecker et al. 2014. Indeed, Ecker et al. is similar  
47 to our model in that includes a trial-varying GP component. However, this trial-varying GP is confined to be 1D,  
48 and there is no stimulus-locked GPFA. Instead, there are trial-averaged tuning-curves which are trivially the same  
49 dimensionality as the neural population and not governed by a GP. Thus, the Ecker et al model cannot identify signal  
50 and noise subspaces and does not allow for any of the analysis that we do in the paper. Additionally, Ecker et al. does  
51 not have Poisson observations, and there is no multi-region analysis. We do find the work relevant as background and  
52 now cite it. However, the lack of reference to this paper seems unreasonable grounds for rejection.

53 *There are also other limitations ... to assume a Poisson distribution ...* We use Poisson here as we (and others) show it is  
54 a better description of neural activity than standard Gaussian noise (Fig 2D, now in supplement).

55 We thank the reviewers for identification of other small concerns, typos and figure comments. These are now addressed  
56 in the current manuscript.

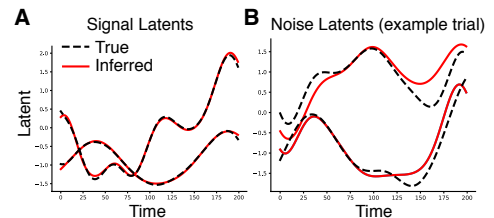


Fig R1: Inferred signal and noise (1 tr shown) latents on simulated data.