We would like to thank the reviewers for taking the time to provide us with helpful feedback and will definitely incorporate the suggestions in the final version. Below are our clarifications for the questions raised.

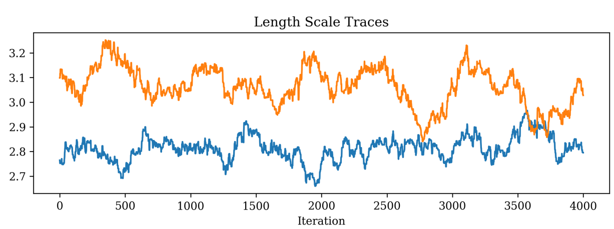
**Weak Stationarity** Since weak stationarity is a key assumption for many theorems in time series analysis, we felt it was important to explicitly show this property. In particular, weak stationarity implies that the model curves lie in a Hilbert space, permitting direct applications of Bochner's theorem, the Fourier transform, and other spectral methods. Further, the proof of weak stationarity provides an explicit formula for the covariance of $Y$, which produces the covariance structure given in Figure 2(d). In practice, the assumption of a weakly stationary covariance process can be empirically checked to determine whether the LFGP model is appropriate.

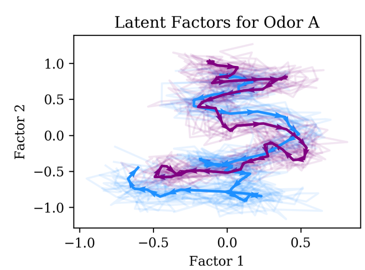
**Factor modeling** In the present work, the factors have a GP prior distribution with squared exponential kernel, which is a distribution over the functions in the RKHS defined by this kernel. With this kernel, the factorization is equivalent to a multivariate time series factor model (Lopes & West) with an AR(1) covariance, which has been extensively studied and validated. The stability of the factorization under other kernels will be important to study in future work.

**Sliding window** As our inference goals were solely regarding the covariance process, we chose to forgo specifying an explicit distribution on the observed time series. Since the LFGP model is fit across repeated trials, the artifacts introduced by sliding window estimation are unlikely to influence the modeling results, assuming that the artifacts are uncorrelated across trials. In order to conduct inference on the theta band oscillations in the DFC series, we determined an L between 50-200ms to be scientifically reasonable. L was set to 100ms to reduce some noise without oversmoothing and 15ms taper was chosen, which gives a tapered window with similar structure to that used in Lindquist 2014.

**LFP data** The local field potentials are measured by surgically implanted tetrodes and the exact tetrode locations vary across rats. Therefore, it may not make sense to compare LFP channels of different rats. This issue actually motivates the latent factor approach because we want to eventually visualize and compare the latent trajectories for all the rats. We have focused on the data from a particular rat because it has the best memory task performance so the data has the strongest signal. The six LFP channels are chosen also to help with the signal to noise ratio because they have the most neurons attached. Lastly, we used trials of odor B and C because the rats smell the first odor (A) much more often than odors in the end (D, E) due to the experimental design.

**Model fitting** In Section 3.3, the horseshoe prior is used on the loadings as an illustration of the methodology when the number of factors is unknown. A short chain of the No-U-Turn Sampler is run within each iteration of the Gibbs sampler. For analyzing the LFP data, we actually used the Gaussian-Inverse Gamma conjugate priors for the loadings and variances. As for the $\mathcal{GP}$ hyper-parameters, we used much more informative priors because of concerns about convergence and identifiability. Gamma(5, 1) prior is used for the length scales because five time points span 100ms and it still has diffused mass from 0 to 10. Based on the trace plots, we can see good convergence and sufficient shift from the prior. The effective sample size is 276 (median) for the loadings but falls short for the length scales.




**Interpretation of results** For non-stationarity and odor separation, we did not perform any statistical test because Bayesian hypothesis testing in general and the null hypothesis formulation in this case are not well-defined. However, we would argue that there are many possible ways to interpret the results. From the posterior correlation between two selected channels, we can observe that it changes from negative to positive so the covariance is varying to some extent. As for odor separation, when we fit the model to two random subsets of the 58 trials of odor A, the latent trajectories are much more intertwined. We can repeat similar analysis for more odors and for resting data.

**Model comparison** Comparing dynamic connectivity models whose approaches range from classical time series to graph based models is tricky; every paper has a different simulation study but at least the "truth" is known. On the LFP data, we know there could be dynamic connectivity and odor differentiation, but not for a certainty. Therefore, it is difficult to find a single "benchmark model" that estimates dynamic connectivity and differentiates odors. That said, the model in Andersen et al is similar to our model in the way it utilizes latent Gaussian processes. It would be very interesting to compare the two models on a publicly available data set with published results as suggested.