We'd like to thank all the reviewers for your time and your constructive comments. They are valuable for our future research.

we thank Reviewer 1 (**R1**) for pointing out the clarity of notation definitions. We will provide more explanations for the quantity definitions in the camera-ready version. As in p.98-99, C denotes a subset of input X, whose elements' corresponding observations are the relevance vectors (derived from ARD in Eq. (2)) with the maximum non-zero component of the weight vector θ_k.

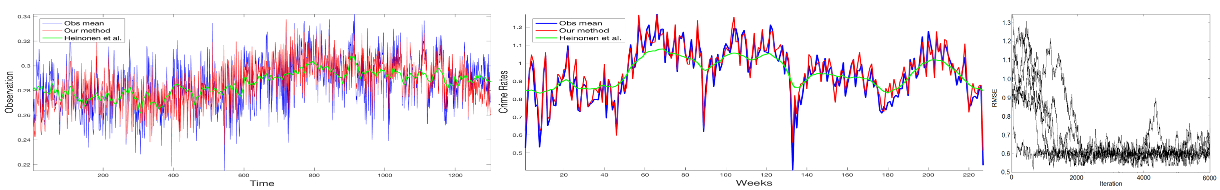
As **R2** and **R3** suggested, we implement two nonstationary covariance function methods in the two applications for empirical comparisons. One is a classical nonstationary Gaussian process regression: Paciorek et al. "Nonstationary covariance functions for Gaussian process regression", which is cited as [14] in our paper. The other is Heinonen et al. "NonStationary Gaussian Process Regression with Hamiltonian Monte Carlo" coupling nonstationary GP with priors on the hyperparameters of the squared exponential covariance functions. We implement these methods with the same experimental settings as ours. In particular, since Heinonen et al. can only model univariate observations, we have to apply it to model each observation variable independently and evaluate the means for the multivariate cases.

Table 1: Performance evaluation of the SOZ channel detection

Methods	AUROC	Precision	Recall (Sensitivity)	F1 score
Our method	0.81 ± 0.05	0.45 ± 0.07	0.77 ± 0.06	0.51 ± 0.07
Nonstationary covariance function models				
Paciorek et al.	0.63 ± 0.07	0.41 ± 0.05	0.43 ± 0.09	0.39 ± 0.04
Heinonen et al.	0.67 ± 0.05	0.43 ± 0.03	0.58 ± 0.03	0.42 ± 0.07

Table 2: Monthly average RMSE of one-week-ahead predictions of the crime rates in 2019.

$RMSE \pm error$	Jan. 2019	Feb. 2019	Mar. 2019	April 2019
Our method	0.638 ± 0.025	0.707 ± 0.023	0.815 ± 0.029	0.817 ± 0.027
Paciorek et al.	0.949 ± 0.034	1.122 ± 0.055	1.176 ± 0.209	1.462 ± 0.147
Heinonen et al.	0.704 ± 0.031	0.875 ± 0.118	0.931 ± 0.763	1.069 ± 0.014



(a) (b) (c)

Figure 1: (a) Empirical mean of the 8 PIB features of a true positive iEEG observation's heldout segment (blue), the predictive mean of our method (red), and the predictive mean of Heinonen et al. method (green). (b) Plots of observation mean (blue) of the crime occurrence rates of 179 CTs in 2015-2019, our method's posterior and predictive mean (red), and Heinonen et al. mean (green). (c) Plot of RMSE vs. 6 Gibbs sampling chains for the convergence.

As in Tables 1 and 2, both methods are inferior to ours, but Heinonen et al. method is comparable to the partition-based nonstationary methods. For Heinonen et al. method, we run 3 chains of 5000 samples of HMC-NUTS sampling to infer the three sets of hyperparameters (noise variance, signal variance, and lengthscale), and initialize the method as suggested in the paper. As in Figure 1 (a) and (b), one key to the success of the method is the balance between the signal variance and the nonstationary lengthscale, which is intrinsically related to the partition-based idea. For Paciorek et al. method, we use the Matern covariance function described in the paper. The Matern kernel leads to less smooth functions, but it still assumes a stationary process in that the covariance structure is the same throughout the entire input space. One major challenge to implement Paciorek et al. method is that the number of hyperparameters increases fast in multivariate cases. In particular, computation of the kernel matrices at each input location is slow because of the matrix decomposition ($O(n^3)$). In contrast, our method is more computationally efficient by introducing the conditional independence given the hyper-GP as in Eq. (10). As suggested by **R3**, we plot the RMSE vs. Gibbs samples in Figure 1 (c) to demonstrate the convergence for DC crime case. We'll add these results to the camera-ready version. To **R2**, if partitions are not local, the method will reduce to a stationary one. Chinese restaurant process will be our future work.