Thesis proposal

Information flow in networks based on nonstationary multivariate neural recordings

January 17, 2019

Natalie Klein

Advisor: Robert E. Kass

Committee: Valérie Ventura, Max G'Sell, Tobias Teichert

Abstract

Neural recordings, such as local field potentials (LFPs), reflect the activity of populations of neurons in time-varying voltage traces and, due to high temporal resolution, they are well-suited for identifying networks of interacting brain regions. Typical analyses are performed on the average across many repetitions of the same task, which eliminates the variation needed to quantify statistical associations between nonstationary signals. In this thesis, I extend statistical and machine learning methodology from graphical models, time series and spatiotemporal models, and Bayesian hierarchical models to develop new tools appropriate for identifying networks of interacting brain regions from multivariate neural recordings. I discuss three different methods, each designed to focus on different a characterization of association. First, we developed dynamic kernel canonical correlation analysis (DKCCA) to identify time-varying lagged correlations between multi-electrode LFPs from two brain regions (Rodu, Klein, *et al*, *J. Neurophys.*, 2018). Second, in work submitted for publication, we explored a novel undirected graphical model suitable for identifying lagged synchronization of neural oscillations via phase coupling and provided inference methods for graphical structure learning. Finally, my current work seeks to infer neural circuitry during stimulus processing on a finer spatial scale using LFP recordings in one cortical area. In particular, I propose using a biophysically-motivated spatiotemporal Gaussian process model to solve an ill-posed inverse problem and recover the current source density (CSD) generating the observed LFPs. In addition, I propose a Bayesian hierarchical model for variation in nonstationary stimulus responses, where correlated current source variation across cortical layers may indicate information flow. I plan to demonstrate these methods in laminar LFP recordings from primary auditory cortex.

Contents

1 Introduction

Electrical signals recorded from electrodes placed in brain tissue reflect time-varying neural activity. The high-frequency content of the recorded signal indicates spikes of individual neurons in the vicinity of the electrode, while the lower-frequency content (consisting of timescales slower than about 500Hz) is termed the *local field potential* (LFP) and reflects synaptic processes across a population of nearby neurons [Buzsáki et al. 2012, Einevoll et al. 2013]. However, unlike spiking activity that can typically be attributed to a small collection of individual cells very close to the electrode, the LFP is an indirect and aggregated measure of the activity of many neurons and does not only reflect activity in the vicinity of the electrode [Lindén et al. 2011, Kajikawa and Schroeder 2011, Herreras 2016].

While my previous projects used LFPs directly to investigate information flow between distant brain regions, in my proposed work, I seek to recover measures of the neural activity of specific cell populations near each recording electrode. That is, the neural activity of interest is the aggregate current flow in and out of specific populations of neurons, called the *current source density* (CSD). Biophysical models relate the CSD to the measured LFPs through a *forward model*; attempting to invert the forward model to infer the CSD from measured LFPs is the goal of CSD estimation methods [Pitts 1952, Nicholson and Llinas 1971]. The CSD provides a measure of neural activity that may better represent the activity of specific neural populations near the electrodes, and thereby help to understand information flow within a neural circuit. In particular, trial-to-trial variation in responses to a stimulus [Arieli et al. 1996] are of great interest, and previous work in the auditory system suggests that trial-to-trial variation in stimulus-evoked responses is better understood using the CSD than the LFPs [Szymanski et al. 2011].

In my proposed work, I consider the common situation where LFPs are recorded using electrodes along a linear probe that is oriented perpendicular to the cortical layers to capture activity in all cortical layers. I develop a Gaussian process spatiotemporal model for the latent CSD (GPCSD) that gives rise to the measured LFPs; in addition, I model each trial as the sum of a trial-specific stimulus-evoked response and ongoing activity to make inferences about trial-to-trial variation. The portion of the GPCSD model that governs how LFPs arise from the CSD is derived from a biophysical forward model, while the statistical model provides the inverse solution (that is, predicts the CSD from measured LFPs).

To illustrate the basic idea of GPCSD, I use an example trial of real LFP data recorded in primary auditory cortex (see details in Section 3.4). The real data is shown in Figure 1.A with LFP voltage traces overlaid on a heat map depicting the spatiotemporal LFP. In the model, the latent noiseless LFP (panel B) decomposes into the sum of an evoked response (middle) and ongoing activity (right). Likewise, the latent CSD (panel C) also decomposes into a sum of evoked and ongoing activity, and the forward model describes how a given CSD generates an LFP. As Figure 1 demonstrates, the forward model dictates that LFPs are a spatially smoothed version of the CSD, so the CSD better reflects neural activity in specific spatial locations. In particular, the CSD evoked response, consisting of several separate bumps, more likely reflects stimulus-related activity of individual cell populations which are difficult to disentangle from the spatially blurred LFP evoked response. As a result, trial-to-trial variation of the CSD evoked responses should be more informative about relationships between cell populations than trial-to-trial variation of the LFP evoked responses.

While most of this proposal document focuses on my proposed methodology for recovering

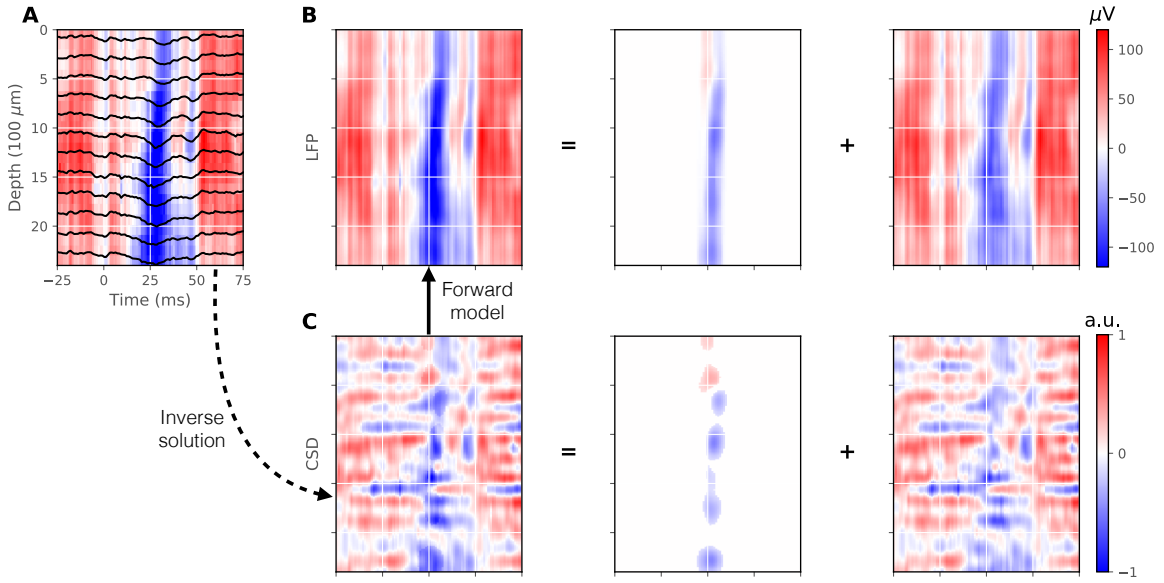

Figure 1: A) A single trial of real LFP data (detailed in Section 3.4), with LFP voltage traces overlaid on a heat map describing the variation in LFPs across electrode depth and time. B) Illustration of the model for the LFP (top row) and the CSD (bottom row), where each is decomposed into an evoked response (middle column) and ongoing activity (right column). A biophysical forward model describes how a given CSD generates an LFP, while the proposed GPCSD statistical model provides the inverse solution (recovers CSDs from LFPs).

the CSD from the LFPs, I give a brief overview of my previous projects using LFPs in Section 2. In Section 3, I outline relevant notation and definitions, describe the biophysical forward model relating the CSD to LFPs, review existing methods for estimating the CSD from measured LFPs, and describe the experiment and data used for preliminary results. Section 4 outlines the GPCSD model and compares it qualitatively to existing CSD methods. Preliminary results are shown in Section 5 for both simulations and real data, demonstrating the utility of GPCSD in recovering the CSD and inferring trial-to-trial variation. Finally, in Section 6, I outline the future work to be completed and give an estimated timeline for completing the work.

2 Previous projects

My previous projects focused on inferring information flow between separate brain regions based on LFPs from each region. Section 2.1 briefly describes the first project, dynamic kernel canonical correlation analysis (DKCCA), which resulted in an estimator of the cross-correlation between two regions based on multi-electrode LFPs within each area. Section 2.2 outlines the second project, torus graphs for multivariate phase coupling analysis, which provided a method for estimating undirected graphical models describing interactions between regions based on trial-to-trial phase relationships in oscillatory LFPs.

2.1 Dynamic kernel canonical correlation analysis (DKCCA)

Canonical correlation analysis (CCA) considers linear combinations of each of two sets of variables in which the weights are chosen to maximize the correlation between the linear combinations of each set of variables; kernel CCA generalizes CCA by mapping the sets of variables to higher-dimensional feature spaces prior to optimizing over the weights [Hardoon et al. 2004]. Even if nonlinear feature maps are not used, kernel CCA provides a framework for computationally-efficient regularized CCA when the number of variables is large relative to the number of observations. To understand cross-correlation between LFPs in two brain regions based on multi-electrode recordings within each region, we developed *dynamic kernel CCA* (DKCCA) as an extension of kernel CCA to multivariate time series [Rodu et al. 2018]. By estimating the kernel CCA weights in a sliding-window fashion, the weights varied smoothly over time and provided coherent time-varying linear combinations for each set of variables. The resulting one-dimensional projections for each region provide an informative cross-correlation matrix that may indicate lagged correlations between brain regions. We compared DKCCA to simpler methods based on assuming equal weights for all electrodes over time (equivalent to averaging across all signals within a brain region before computing cross-correlation) or on computing all-pairwise cross-correlation matrices and then averaging the matrices across all pairs. In simulations, we found that DKCCA recovered the correct lagged structure while the other methods did not. In real LFP data from hippocampus and prefrontal cortex in an associative learning task [Brincat and Miller 2015], DKCCA suggested task-relevant lagged correlations during memory retrieval that the other methods did not.

2.2 Torus graphs for multivariate phase coupling analysis

The concept of phase coupling is frequently applied to neural recordings with substantial oscillatory activity, where phase coupling indicates across-trial phase-locked relationships between oscillations in two different brain regions that are a potential marker of long-range neural integration [Lachaux et al. 1999, Marek et al. 2018]. However, existing methods for assessing phase coupling are inherently bivariate in nature, meaning they construct graphs representing interactions between brain regions based on evaluating phase coupling for each pair of regions individually. In contrast to bivariate methods, undirected graphical models provide the very nice interpretation that an edge between a pair of nodes is absent if and only if the corresponding pair of nodal random variables are independent after conditioning on all the other nodal variables; in this sense, an edge represents a unique association of two nodal variables that can not be explained by the other variables. To estimate undirected graphical models based on phase relationships, we developed a model suitable for multivariate phase angles, the torus graph model, which generalized previous work in multivariate distributions for circular random variables [Klein et al. 2018].

In particular, for a vector of multivariate angles $\mathbf{y}$, the torus graph is an exponential family distribution $p(\mathbf{y}|\boldsymbol{\theta}) \propto \exp\left(\mathbf{T}(\mathbf{y})^T\boldsymbol{\theta}\right)$ characterized by sufficient statistics with second-order interactions between angles y_i and y_j:

$$\mathbf{T}(y_i, y_j) = [\cos(y_i), \sin(y_i), \cos(y_i - y_j), \sin(y_i - y_j), \cos(y_i + y_j), \sin(y_i + y_j)]^T. \quad (1)$$

We show that this model generalizes previous work in multivariate circular statistics and that the full model is necessary for modeling phase angles extracted from neural data. Furthermore, we propose a computationally efficient and statistically consistent estimation method

based on score matching [Hyvärinen 2005], as an intractable normalization constant complicates the use of more standard techniques such as maximum likelihood. Finally, we detail asymptotic inference or regularization approaches to determining an undirected graph structure using the model.

In the context of multivariate phase coupling of neural oscillations, we illustrated using simulations how torus graphs can overcome drawbacks of bivariate phase coupling measures; specifically, bivariate measures are sensitive to the marginal distributions of the variables and are influenced by both direct dependencies and indirect dependencies (mediated by other nodes). In neural oscillation phase angle data from 24-dimensional LFPs, previously studied in Brincat and Miller [2015], we demonstrated that torus graphs were able to recover meaningful phase-based functional connectivity structures between prefrontal cortex and three hippocampal subregions.

3 Background

My proposed work is based on biophysical models of LFPs recorded on a linear probe, though I plan to extend it to data from other recording devices (Section 6). Notation and definitions are outlined in Section 3.1, while biophysical models for LFPs are discussed in Section 3.2. Section 3.3 reviews existing methods for estimating the CSD from LFP recordings. Finally, the experiment and LFP data set used in the preliminary results is described in Section 3.4.

3.1 Notation and definitions

The local field potentials (LFPs) and current source densities (CSDs) may be conceptualized as spatiotemporal functions of a single spatial coordinate and a time coordinate, so I will write the LFPs as a function $\phi : \mathbb{R}^2 \to \mathbb{R}$ and the CSDs as a function $c : \mathbb{R}^2 \to \mathbb{R}$. The portion of the activity time-locked to a stimulus is called the *evoked response*, while all other activity is called the *ongoing activity*. A single *trial* captures the neural activity in a fixed time epoch around a single stimulus presentation. The average across trials, aligned in time to the stimulus onset, provides an estimate of the evoked response common to all trials and is called the *average evoked response*.

To denote a realization of a function or parameter on the nth trial, I will use the notation $f^{(n)}$. The scalar output of a function $f^{(n)} : \mathbb{R}^2 \to \mathbb{R}$ at input point (z, t) will be denoted $f^{(n)}(z, t)$, where z is the spatial location (depth along the probe) and t is the temporal location. If the function is evaluated at a grid of space-time points represented by vectors $\mathbf{z} \in \mathbb{R}^D$ and $\mathbf{t} \in \mathbb{R}^T$, then $f^{(n)}(\mathbf{z}, \mathbf{t})$ will be a $D \times T$ matrix of function values. By default, $\mathbf{z} \in \mathbb{R}^D$ and $\mathbf{t} \in \mathbb{R}^T$ to refer to the measured spatial locations and time points of the observed LFPs. Often, such matrices will be vectorized; I will use $\mathbf{f}^{(n)}_{\mathbf{z},\mathbf{t}} \equiv \text{vec}\left[f^{(n)}(\mathbf{z}, \mathbf{t})\right]$ to represent this DT-vector. I will use $\tilde{\boldsymbol{\phi}}_{\mathbf{z},\mathbf{t}}$ to refer to the observed LFPs to distinguish them from latent function values representing noiseless LFPs.

When considering Gaussian process models, I will use a trial-specific spatiotemporal mean function $\mu^{(n)} : \mathbb{R}^2 \to \mathbb{R}$ and spatiotemporal covariance functions $k : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$. The scalar value of a covariance function at a pair of input points $\{(z, t), (z', t')\}$ is denoted $k(z, t; z', t')$. *Separable covariance functions* take the form $k(z, t; z', t') = k^z(z, z')k^t(t, t')$; evaluating such a covariance function at vectors of inputs $\{\mathbf{z} \in \mathbb{R}^D, \mathbf{t} \in \mathbb{R}^T, \mathbf{z}' \in \mathbb{R}^{D'}, \mathbf{t}' \in \mathbb{R}^{T'}\}$ then yields a

covariance matrix $\mathbf{K}_{\mathbf{z},\mathbf{t};\mathbf{z}',\mathbf{t}'} \in \mathbb{R}^{DT \times D'T'}$ that decomposes as a Kroencker product,

$$\mathbf{K}_{\mathbf{z},\mathbf{t};\mathbf{z}',\mathbf{t}'} = \mathbf{K}^z_{\mathbf{z},\mathbf{z}'} \otimes \mathbf{K}^t_{\mathbf{t},\mathbf{t}'}, \tag{2}$$

where $\mathbf{K}^z_{\mathbf{z},\mathbf{z}'} \in \mathbb{R}^{D \times D'}$ and $\mathbf{K}^t_{\mathbf{t},\mathbf{t}'} \in \mathbb{R}^{T \times T'}$ are matrices formed by evaluating the covariance functions on pairs of input vectors. The expression $\mathbf{K}_\mathbf{t}$ will be used as shorthand for $\mathbf{K}_{\mathbf{t},\mathbf{t}}$.

Linear operators will be denoted $\mathcal{A}_z$ where z is the input acted upon by the operator; that is, $\mathcal{A}_z f(z) = h(z)$. When applying operators to functions with more than one input, I will use the following conventions. By $\mathcal{A}_z f(z,t)$, I mean to apply the operator to f as a function of z with the other argument, t, held fixed. To apply a linear operator to both arguments of a function, I will use the following notation: $\mathcal{A}_z f(z,z')\mathcal{A}^T_{z'}$.

3.2 Biophysical models of LFPs

Biophysical models give the relationship between the LFPs and the underlying CSD at any instant in time. In particular, current flow across a single cell membrane creates a current source or sink and the resulting field potential can be derived using volume conductor theory. Including the contributions of all such current sources or sinks in space results in a biophysical *forward model* relating LFPs to the CSD. Because it is not possible to estimate the contribution of individual transmembrane currents from measured LFPs, the CSD may be conceptualized as a continuous function in three-dimensional space that reflects average transmembrane current in a small area [Einevoll et al. 2013]. In the following discussion, I first give an overview of the three-dimensional forward model, then detail the use of additional assumptions, which I call *a priori physical models*, to adapt the forward model to one-dimensional LFP measurements; other cases, such as two-dimensional LFP measurements, can also be treated given appropriate *a priori* physical models.

Three-dimensional biophysical models Using the quasi-static assumption and assuming an isotropic, homogeneous medium with scalar conductivity σ, the relationship between the CSD c and the LFP ϕ is governed by the Poisson equation [Pitts 1952]:

$$\sigma \nabla \cdot (\nabla \phi(x,y,z)) = \sigma \left(\frac{\partial^2 \phi(x,y,z)}{\partial x^2} + \frac{\partial^2 \phi(x,y,z)}{\partial y^2} + \frac{\partial^2 \phi(x,y,z)}{\partial z^2} \right) = -c(x,y,z) \tag{3}$$

While this appears to give a formula for computing the CSD from the LFP, it requires detailed knowledge of the LFP in three dimensions, without which it fails to accurately recover the CSD [Nicholson and Freeman 1975].

Instead, the differential equation in Equation 3 may be inverted to an integral equation which gives ϕ in terms of an integral operator on c. Assuming an infinite volume conductor with negligible boundary conditions leads to the following integral operator [Nicholson and Llinas 1971]:

$$\phi(x,y,z) = -\frac{1}{4\pi\sigma} \int\int\int \frac{c(x',y',z')}{\sqrt{(x-x')^2 + (y-y')^2 + (z-z')^2}} \, dx' \, dy' \, dz'. \tag{4}$$

One advantage of this formulation is that it is easy to incorporate prior beliefs about the three-dimensional behavior of the CSD when adequate three-dimensional measurements are not available.

One-dimensional biophysical models When observations are only available in one dimension, we must provide some prior beliefs about the behavior of the CSD in the unmeasured dimensions through an *a priori* physical model. As in previous works, I use an *a priori* physical model in which the CSD is assumed constant in the dimensions perpendicular to the linear probe on a cylinder of radius R around the probe and zero elsewhere; previous work has shown deviations from this shape do not have a large impact on the results [Nicholson 1973, Potworowski et al. 2012]. Additionally, as linear probes are typically inserted to cover all layers of cortex, it is reasonable to assume the CSD is nonzero only on an interval $a \leq z \leq b$, leading to the following *a priori* physical model that describes the variation of the CSD in the z direction through $g(z)$:

$$c(x, y, z) = g(z)\mathbb{1}(x^2 + y^2 \leq R)\mathbb{1}(a \leq z \leq b). \tag{5}$$

Under this *a priori* physical model, as shown in Appendix A.1, Equation 4 reduces to

$$\phi(z) = \mathcal{A}_z g(z) \equiv -\frac{R}{2\sigma}\int_a^b g(z')\underbrace{\left[\sqrt{\left(\frac{r}{R}\right)^2 + 1} - \sqrt{\left(\frac{r}{R}\right)^2}\right]}_{b(r;R),\ \text{where } r=z-z'} dz'. \tag{6}$$

Thus, under this *a priori* physical model, the one-dimensional LFP is the result of applying a linear operator $\mathcal{A}_z$ to $g(z)$, where the weighting function $b(r; R)$ decreases with distance r but also depends on the radius R; see Figure 2 for an illustration.

3.3 Existing CSD estimation methods

As discussed in Section 3.2, in principle, Equation 3 may be used to infer the CSD if the LFPs are observed densely in three-dimensional space; in practice, this is generally not the case, and even if such recordings were available, numerical second derivatives are sensitive to noise. Nevertheless, the *traditional CSD* (tCSD) method is based on a direct application of Equation 3 to one-dimensional recordings under the assumption that neural activity is constant in the directions perpendicular to the probe. In contrast, a second class of techniques, which I will call *inverse CSD* methods, uses Equation 4 along with *a priori* physical models to estimate the CSD by inverting the resulting linear operator (e.g. inverting $\mathcal{A}_z$ in Equation 6).

Traditional CSD (tCSD) method Ignoring the x, y directions in Equation 3 yields

$$\sigma\left(\frac{\partial^2\phi}{\partial z^2}\right) = -c(z). \tag{7}$$

This suggests estimating the CSD by the second spatial derivative of the recorded LFPs, assuming equal spacing Δz between electrodes [Nicholson and Llinas 1971]:

$$\hat{c}(z_i) = \frac{\tilde{\phi}(z_{i+1}) - 2\tilde{\phi}(z_i) + \tilde{\phi}(z_{i-1})}{(\Delta z)^2},\ \forall\, i \,|\, i \in \{2, ..., D-1\}. \tag{8}$$

However, implicit in this interpretation of Equation 3 is the assumption that the LFPs have zero curvature in the x and y directions, which has been shown to correspond to assuming

$R \to \infty$ in the *a priori* physical model of Equation 5 [Pettersen et al. 2006]. As a result, tCSD has been shown to work poorly when the CSD is actually confined in a cylinder of radius R around the recording probe [Nicholson and Freeman 1975, Einevoll et al. 2013].

In addition, the utility of tCSD on single-trial recordings is questionable as tCSD does not account for measurement noise (and the numerical calculation of second derivatives is sensitive to noise). While spatial and/or temporal smoothing may be used before or after tCSD, it is unclear how to choose appropriate smoothing parameters and how to propagate uncertainty about the smooth function when further analysis is performed on the smoothed estimate. Perhaps for this reason, tCSD is typically applied to many trials, then averaged across trials and further smoothed for visualization. Furthermore, tCSD can only provide estimates of the CSD at the same locations where the LFP is measured (excluding the edge electrodes), and tCSD must be applied separately on each trial and at each time point.

Inverse CSD methods While ideally Equation 4 can be inverted analytically (resulting in Equation 3), once an *a priori* physical model is incorporated to describe variation in unmeasured dimensions, this is no longer the case. Additionally, viewing this problem through the lens of inverse theory suggests that it is an ill-posed inverse problem, meaning inverse solutions are highly sensitive to noise and may not be unique [Kropf and Shmuel 2016]. For the one-dimensional case with the *a priori* physical model of Equation 5, Potworowski et al. [2012] developed *kernel CSD* (kCSD) which models the CSD $g(z)$ as a sum of finitely many basis functions (and includes so-called iCSD by Pettersen et al. [2006] as a special case). In particular, the CSD is modeled using M known basis functions $\tilde{b}_j(z)$:

$$g(z) = \sum_{j=1}^{M} a_j \tilde{b}_j(z). \tag{9}$$

Then, applying the forward model, one may obtain basis functions for the LFPs, denoted $b_j(z)$. Then minimum-norm inverse solution (see Appendix A.2), in terms of kernel functions $k(z, z') \equiv \sum_{j=1}^{M} b_j(z) b_j(z')$ and $\tilde{k}(z, z') \equiv \sum_{j=1}^{M} \tilde{b}_j(z) b_j(z')$, is

$$\hat{\mathbf{c}}_{\mathbf{z}'} = \tilde{\mathbf{K}}_{\mathbf{z}',\mathbf{z}} \mathbf{K}_{\mathbf{z},\mathbf{z}}^{-1} \tilde{\boldsymbol{\phi}}_{\mathbf{z}}, \tag{10}$$

where, unlike tCSD, kCSD can provide predictions at new spatial locations $\mathbf{z}'$. Additionally, the inverse solution may be regularized by replacing $\mathbf{K}_{\mathbf{z},\mathbf{z}}^{-1}$ with $[\mathbf{K}_{\mathbf{z},\mathbf{z}} + \lambda \mathbf{I}]^{-1}$.

Inverse CSD methods are preferable to tCSD because the forward model parameter R is not implicitly assumed to be infinite and because regularization reduces sensitivity to noise and ensures a unique inverse solution. However, like tCSD, kCSD must be applied separately to each trial and each time point. In addition, kCSD requires choosing the functional form of the basis functions in addition to the placement and number of basis functions, for which Potworowski et al. [2012] provide only heuristics. While cross-validation is suggested to choose the regularization parameter λ and the width of the basis functions, no suggestion is made for selecting R.

3.4 Experiment and LFP data

The data used for preliminary results are comprised of LFP recordings (1 KHz sampling rate) from a 24-electrode linear probe inserted in primary auditory cortex of a macaque monkey

with stimuli consistning of short tones at different frequencies (Walter, 5/12/2016). I defined the duration of a trial as 100 ms before the tone to 450 ms after the stimulus. Trials with more than one stimulus within this time epoch were excluded, as were trials with LFP ranges (from minimum to maximum voltage value) over 900 μV; this resulted in 3010 trials. The experimental manipulation that most clearly impacts the measured neural activity is the inter-stimulus-interval (ISI), or the time since the last tone was played; trials with longer ISIs tend to contain larger responses with somewhat different timecourses. The ISIs are distributed in a right-skewed fashion to attempt to cover the space of different ISIs while maintaining a reasonable recording session length; as a result, the natural logarithm of the ISI is used as a covariate. For numerical reasons, the amplitude of all trials was rescaled by a common factor so that single-trial LFPs have maximum absolute values near 1 (units: $10^{-4}V$). For the preliminary results, I focus on the time period of the trial that captures the earliest evoked responses to the tone by clipping the time period to 25 ms before the stimulus to 75 ms after the stimulus. Whie in principle the conductivity scalar σ of Equation 3 should be measured experimentally, in this work, I focus on recovering the pattern of CSDs and am not overly concerned about the exact amplitude values, so for now I use $\sigma = 1$ and rescale all CSD estimates to the arbitrary units $[-1, 1]$.

4 Proposed methodology

In Section 4.1, I give a high-level overview of the proposed Gaussian process CSD (GPCSD) model for one-dimensional LFP recordings which models the evoked CSD using a trial-specific mean function and the ongoing CSD using a spatiotemporal covariance function (shared across trials). To borrow strength across trials, the trial-specific mean functions are encapsulated in a hierarchical model with covariate fixed effects and per-trial random effects. I show that under the *a priori* physical model of Equation 5, the CSD and LFP are jointly a Gaussian process, allowing prediction of the latent CSD at arbitrary space-time points (Section 4.2). While GPCSD shares some similarities with the kCSD method discussed in Section 3.3, I outline specific advantages of my approach in Section 4.3.

4.1 Gaussian process CSD (GPCSD) model

Under the *a priori* physical model of Equation 5, the time-varying CSD on trial n can be described as a spatiotemporal function $g^{(n)}(z, t)$. To understand information flow following the stimulus presentation, I model g as a combination of a trial-specific mean function plus Gaussian process ongoing activity; as illustrated in Figure 1, application of the forward model to the CSD model yields the LFP model, while the inverse solution will be obtained through the statistical model. In particular, let $g^{(n)}(z, t)$ be decomposed as

$$g^{(n)}(z,t) = \mu^{(n)}(z,t) + \eta^{(n)}(z,t), \qquad \eta^{(n)}(z,t) \sim \mathrm{GP}\left(0, k(z,t;z',t')\right). \tag{11}$$

Combining this model with Equation 6 and putting iid additive noise on the observed LFPs then yields a joint Gaussian process model for the CSD and LFPs (Appendix A.4). The forward model operator $\mathcal{A}_z$ influences the mean and covariance functions of the joint model, but as the integral is generally not available in closed form, I use Gauss-Legendre quadrature.

Covariance functions Because we expect the CSD to reflect the temporal evolution of neural activity in spatially fixed neural populations, it appears to be reasonable to model the covariance as separable in space and time:

$$k(z,t;z',t') = k^z(z,z')k^t(t,t'). \tag{12}$$

This allows simpler specification of the spatiotemporal covariance function and also yields considerable computational advantages (Appendix A.5).

I consider stationary covariance functions under the assumption that ongoing neural activity within a single trial can be treated as approximately stationary in time and space. In general, the spatial and temporal stationary covariance functions may be sums of multiple covariance functions, each with its own functional form and hyperparameters, allowing processes with multiple temporal or spatial scales. In addition, predictions can be made at each time or spatial scale separately [Duvenaud 2014].

To apply the forward model to the Gaussian process $\eta^{(n)}(z,t)$, note that the integral operator is linear and applies only to the spatial part of the covariance function, so that

$$\mathcal{A}_z \mathbf{K}_{\mathbf{z},\mathbf{t};\mathbf{z}',\mathbf{t}'} \mathcal{A}_{z'}^T \equiv \mathcal{A}_z \mathbf{K}^z_{\mathbf{z},\mathbf{z}'} \mathcal{A}_{z'}^T \otimes \mathbf{K}^t_{\mathbf{t},\mathbf{t}'}. \tag{13}$$

Trial-specific mean models To help disentangle stimulus responses of individual neural populations, the trial-specific CSD mean model should be able to capture evoked responses of spatially separated current sources and sinks which are nonstationary in time; I will call these responses *evoked current events* (see Figure 5.C for illustration). That is, if $a_m(z,t)$ represents the spatiotemporal profile of one of M evoked current events, the overall mean is $\mu^{(n)}(z,t) = \sum_{m=1}^M a_m^{(n)}(z,t)$. Then, I propose a hierarchical model for variation in the timing and strength of the evoked current events across trials:

$$\mu^{(n)}(z,t) = \sum_{m=1}^{M} w_m^{(n)}(\mathbf{x}) \cdot a_m\left(z, t + \tau_m^{(n)}(\mathbf{x})\right) \tag{14}$$

The latent variables $w_m^{(n)}(\mathbf{x})$ and $\tau_m^{(n)}(\mathbf{x})$ represent trial-specific amplitude and timing deviations the mth evoked current event, which in general may depend on covariates $\mathbf{x}$ in addition to trial-specific random effects. To infer across-trial relationships between stimulus responses of different neural populations, I propose multivariate priors of the form

$$\begin{aligned} \mathbf{w}^{(n)} &\sim p(\mathbf{w}|\mathbf{x},\boldsymbol{\theta}_w),\ \boldsymbol{\theta}_w \sim p(\boldsymbol{\theta}_w|\boldsymbol{\gamma}_w) \\ \boldsymbol{\tau}^{(n)} &\sim p(\boldsymbol{\tau}|\mathbf{x},\boldsymbol{\theta}_\tau),\ \boldsymbol{\theta}_\tau \sim p(\boldsymbol{\theta}_\tau|\boldsymbol{\gamma}_\tau) \end{aligned} \tag{15}$$

where $\mathbf{w}^{(n)} \in \mathbb{R}^M$ and $\boldsymbol{\tau}^{(n)} \in \mathbb{R}^M$ are M-vectors for trial n.

4.2 Predicting the CSD using GPCSD

One of the key goals of the GPCSD model is to provide predictions of the CSD on the single-trial level and to decompose it into the evoked response and ongoing activity. Suppose that the trial-specific mean function and the covariance and forward model hyperparameters are

known. As shown in Appendix A.4, the distribution of the CSD at inputs $\mathbf{z}'$, $\mathbf{t}'$ conditional on the observed LFP at inputs $\mathbf{z}$, $\mathbf{t}$ is given by

$$\mathbf{g}^{(n)}_{\mathbf{z}',\mathbf{t}'} \Big| \tilde{\boldsymbol{\phi}}^{(n)}_{\mathbf{z},\mathbf{t}} \sim \mathcal{N}(\boldsymbol{\mu}^*, \mathbf{K}^*). \tag{16}$$

In particular, the conditional mean may be used to predict the CSD:

$$\boldsymbol{\mu}^* = \boldsymbol{\mu}^{(n)}_{\mathbf{z}',\mathbf{t}'} + \mathbf{K}_{\mathbf{z}',\mathbf{t}';\mathbf{z},\mathbf{t}} \mathcal{A}_z^T \left[\mathcal{A}_z \mathbf{K}_{\mathbf{z},\mathbf{t}} \mathcal{A}_z^T + \sigma^2 \mathbf{I} \right]^{-1} \left(\tilde{\boldsymbol{\phi}}^{(n)}_{\mathbf{z},\mathbf{t}} - \mathrm{vec}\left[\mathcal{A}_z \mu^{(n)}(\mathbf{z},\mathbf{t}) \right] \right). \tag{17}$$

Similarly, we can use the distribution of the latent LFP $\boldsymbol{\phi}^{(n)}_{\mathbf{z}',\mathbf{t}'}$ conditional on the observed LFP $\tilde{\boldsymbol{\phi}}^{(n)}_{\mathbf{z},\mathbf{t}}$ to predict the LFP evoked response and smoothed ongoing activity.

4.3 Advantages of GPCSD over existing CSD methods

The drawbacks of tCSD were already discussed in Section 3.3, so here I will mostly focus on comparing GPCSD to the existing inverse CSD methods such as kCSD. Both GPCSD and the inverse CSD methods begin with the forward model of Equation 6, propose models for g, then invert the integral operator $\mathcal{A}_z$ to estimate the CSD. In addition, Gaussian process regression and RKHS regression result in similar-looking estimators (see Appendix A.3), so I expect that under some conditions, GPCSD and kCSD would perform similarly. However, the specification of the prior Gaussian process through mean and covariance functions appears more natural than the specification of some number of user-selected basis functions as in kCSD. In addition, unlike kCSD, GPCSD explicitly models multi-trial data with both evoked responses and ongoing activity (which may be decomposed into multiple spatial or temporal scales). Furthermore, unlike existing methods, GPCSD incorporates temporal variation into the model, which not only provides for smoothing and separation of processes into multiple time scales, but also improves estimation by sharing information across nearby correlated time points (sometimes called *multi-task* or *transfer learning* [Alvarez et al. 2012]). Also, unlike existing inverse CSD methods, GPCSD provides principled data-driven inference of the forward model hyperparameter R.

5 Preliminary results

Preliminary results are provided to demonstrate the utility of the GPCSD method. Section 5.1 involves a simulation study in which a realistic ground-truth CSD is generated from a zero-mean Gaussian process; Figure 2 demonstrates the impact of the forward model parameter R on the resulting LFPs, while Figure 3 shows that GPCSD recovers the ground truth CSD from the generated LFPs across a range of R values while tCSD does not. In Section 5.2, I show results of applying the GPCSD model to real LFP data. Figures 4 and 5 demonstrate the predicted CSD ongoing activity (separated into two distinct timescales) and the predicted CSD evoked response, respectively. Posterior inferences on a version of the hierarchical model for across-trial variability are shown in Figure 6, indicating reasonable estimates of trial-to-trial variability along with a covariate fixed effect that appears to be confirmed by exploratory data analysis (Figure 7).

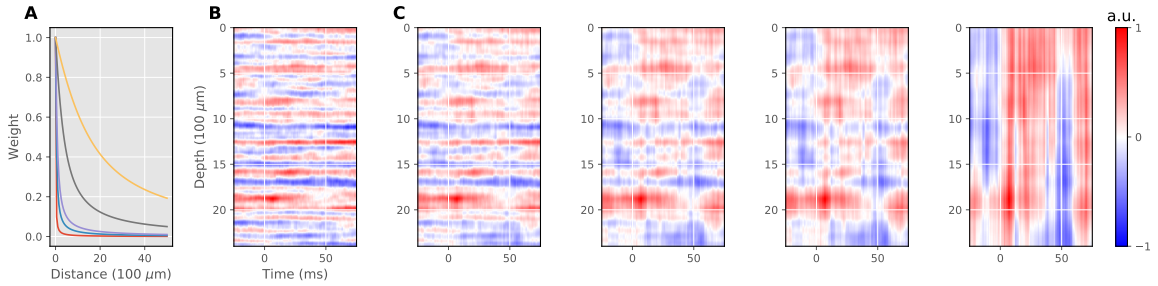

Figure 2: A) Plot of the weight function $b(r; R)$ of Equation 6 as a function of distance r between current source and measured LFP; the different lines represent different R values, where smaller R values lead to faster decay of the weight function with distance. B) Ground truth CSD generated from a zero-mean Gaussian process. C) LFPs generated from the ground truth CSD with four different R values (increasing from left to right); larger R leads to increasingly spatially smooth LFPs. (LFP values in arbitrary units since R affects amplitude and smoothness.)

5.1 Simulation study

The ground truth CSD, generated from a zero-mean Gaussian process, is shown in Figure 2.B. The forward model was applied for $R \in \{10, 50, 100, 500\}$ (μm) to obtain the LFPs shown in Figure 2.C (with R increasing from left to right) which show that the LFP appears more spatially blurred as R increases; even for small R (leftmost plot of Figure 2.C), some of the spatial detail of the CSD is lost in the LFP; the LFPs are rescaled to arbitrary units to empahsize how R impacts the spatial smoothness (though in general it also affects the amplitude). The weight functions $b(r; R)$ of Equation 6 for various R are shown in Figure 2.A as a function of distance, r, from current source to measured LFP location, where the rate of decay with distance is larger for smaller R, demonstrating how larger R results in a spatially smooth LFP.

As discussed in Section 3.3, one of the major drawbacks of the tCSD method is its implicit assumption that $R \to \infty$. By applying tCSD to noiseless simulated LFPs with high spatial resolution, we can obtain much higher resolution tCSD estimates than would be possible on real LFPs, enabling a more fair comparison between tCSD and GPCSD estimates. Figure 3.A shows the ground truth CSD, while Figure 3.B shows the results of applying tCSD to the resulting LFPs (generated with increasing R from left to right). For small R, tCSD does not recover the ground truth CSD and seems to infer spurious spatial variation, coinciding with earlier results of Nicholson and Freeman [1975]. In contrast, Figure 3.C shows the estimated CSD by GPCSD for each R (using the true hyperparameters); GPCSD recovers the ground truth CSD pattern for each R. Note that the CSDs are shown in arbitrary units to compare the patterns of activity, as tCSD and GPCSD infer CSD activity on different scales (and the pattern is the main goal of CSD analysis).

5.2 GPCSD applied to real LFP data

GPCSD with the same mean function for all trials First, I used a single mean function shared across all trials and a covariance function made up of a single spatial covariance and a composite temporal covariance, allowing separation of the predicted CSDs into slow- and fast-timescale components. Figure 4 shows the predicted ongoing activity for one

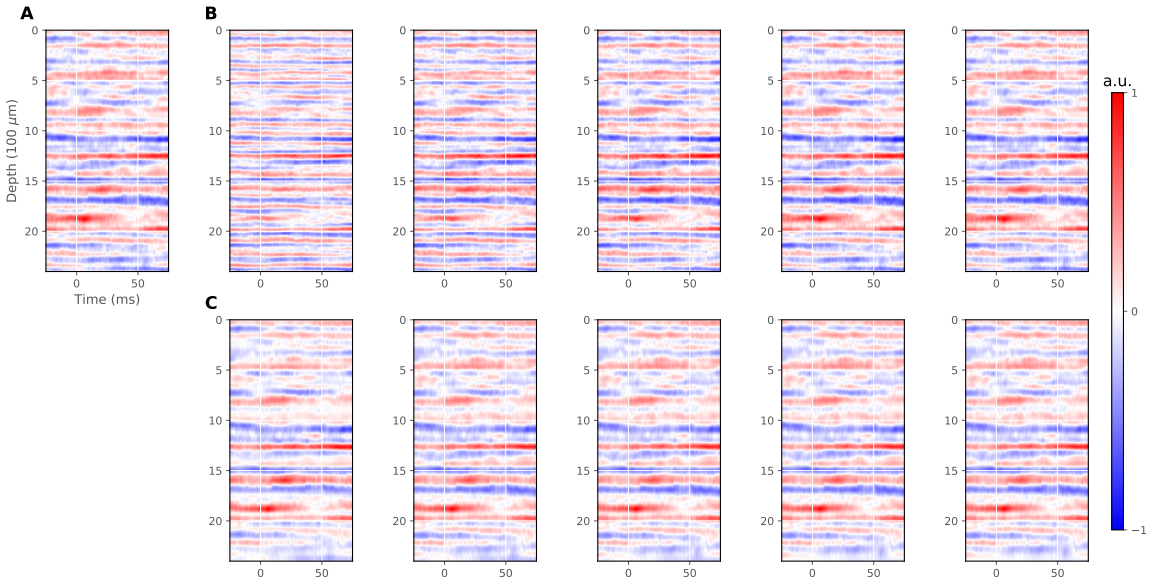

Figure 3: A) Ground truth CSD generated from a zero-mean Gaussian process. B) Across: tCSD estimates based on LFPs generated with different R (increasing from left to right), showing that for small R, the tCSD estimates do not match the ground truth pattern, while for larger R, they do. C) Across: GPCSD estimates for the same LFPs, showing that when the proper R is used as part of the GPCSD model, the ground truth CSD pattern can be recovered from each LFP. (All values in arbitrary units for comparison between tCSD and GPCSD.)

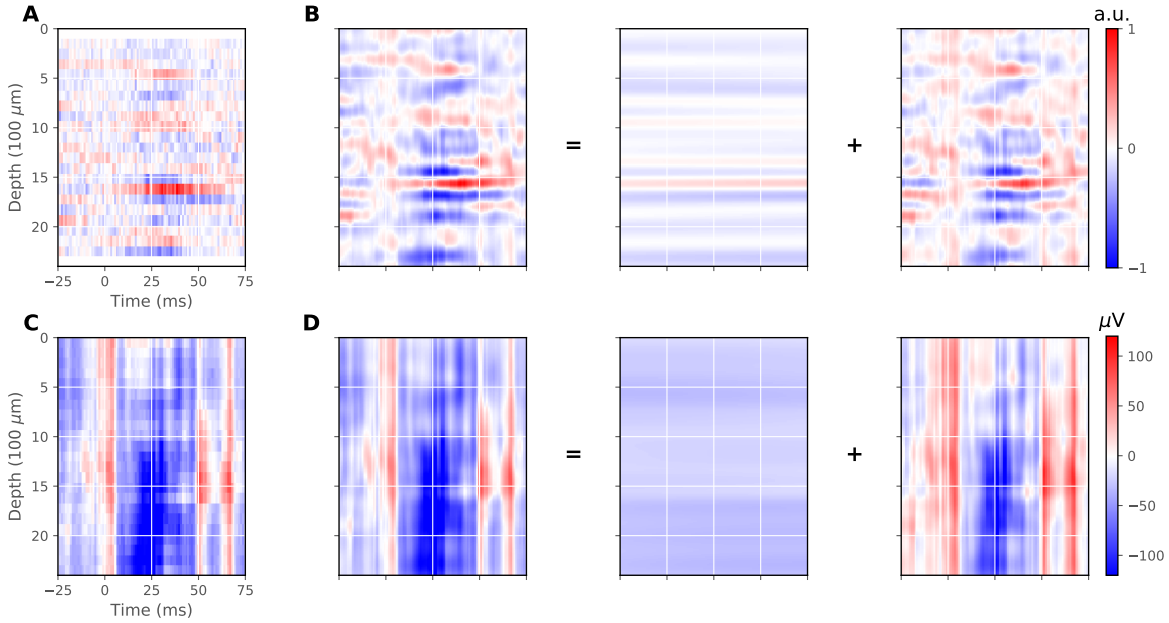

Figure 4: A) The tCSD prediction of the CSD for the ongoing activity in a single trial of real LFP data. B) The GPCSD prediction of the CSD for the same trial, which decomposes into slow- and fast-timescale components (middle and right). Notice the GPCSD prediction bears some resemblance to the tCSD prediction, but the latter has poorer spatial resolution and a lack of temporal smoothness. (CSDs are in arbitrary units for comparison between tCSD and GPCSD.) C) The observed ongoing LFP for this trial. D) The GPCSD prediction of the LFP for this trial, which again decomposes into slow- and fast-timescale components (middle and right). The GPCSD model LFP prediction resembles the real data, and the slow-timescale component appears to capture a baseline shift effect.

example trial. Specifically, Figure 4.B shows the predicted ongoing CSD for a single trial, which is split into fast- and slow-timescale components (middle and right columns). For comparison, the tCSD estimator (computed on the data minus the average evoked response) is shown in Figure 4.A; arbitrary units are used for the CSD to allow comparison between tCSD and GPCSD patterns. While there appear to be some similarities between the tCSD and GPCSD estimates, the tCSD spatial resolution is much lower and the lack of temporal smoothness is apparent. Figure 4.D shows the GPCSD prediction of the noiseless LFP, again including separation into two different time scales in panels E and F, while Figure 4.C shows the observed LFP from this trial for comparison. The GPCSD prediction for the LFP appears to match the observed LFP, and the slow-timescale contribution appears to capture a baseline effect as it is nearly constant over time.

The fitted CSD mean function representing the evoked response common to all trials is shown in Figure 5.C; the left plot is the overall fitted mean surface, while the right plot is a thresholded version that separates the evoked response into six distinct evoked current events which are more interpretable as current flow related to specific neural populations (see Appendix A.6 for discussion of thresholding). The spatially smoothed LFP evoked responses corresponding to these CSD mean functions are shown in panel B, and they are similar for both the original and thresholded CSD mean. For comparison, the average evoked response

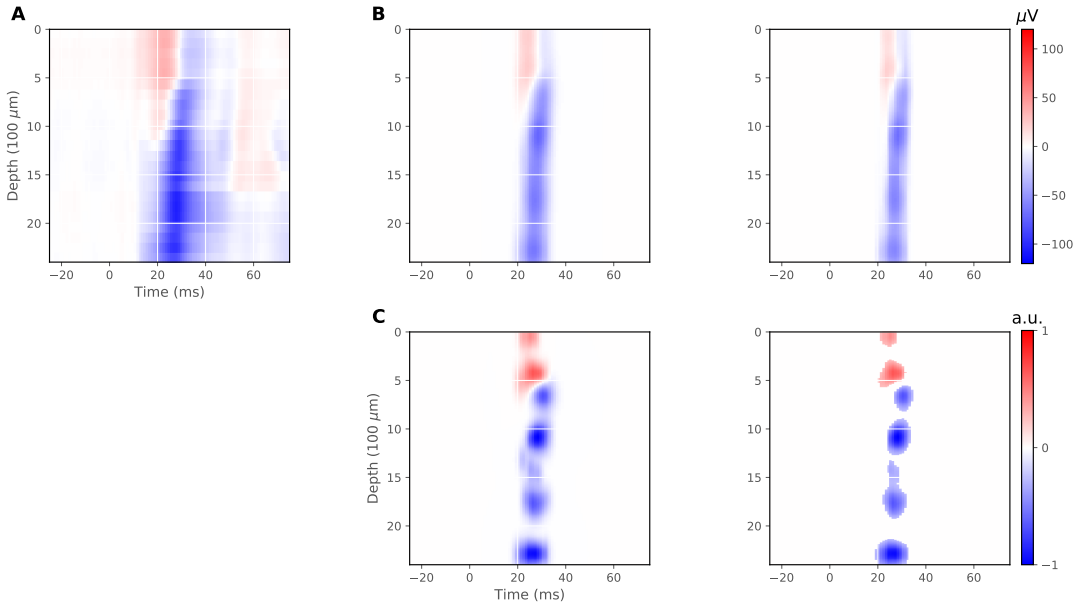

Figure 5: A) The observed average evoked LFP across all trials from the real LFP data. B) The fitted GPCSD model LFP mean function (middle) is similar to a mean function where the CSD mean has been thresholded to obtain separate evoked current events (right). C) The fitted GPCSD model CSD mean function (middle) and thresholded version, showing roughly six separate evoked current events.

is shown in panel A.

GPCSD with trial-to-trial mean function variation Holding the covariance hyperparameters, forward model hyperparameter, and evoked current event shapes from the previous model fixed, I applied a simplified version of the hierarchical model for trial-to-trial variation of evoked responses. In this model, described in detail in Appendix A.6, the entire evoked response surface varies by a single time shift and scale parameter per trial.

Figure 6.A shows a scatter plot of the posterior mean scale parameters $w^{(n)}$ for each trial against a rescaled version of the log ISI. Overlaid is a line showing the posterior mean estimated linear fixed effect of log ISI on the scale parameter; the histograms of posterior samples for the intercept β_0 and slope β_1 of the line are shown in Figure 6.B, suggesting that the amplitude of the evoked response increases as ISI increases. In contrast, Figure 6.C suggests no relationship between the posterior mean time shift $\tau^{(n)}$ and the log ISI. Figure 6.D shows a histogram of the posterior samples of the time shift SD, σ_τ, which concentrates cleanly around 1 ms, a reasonable value for early auditory responses. Finally, Figure 6.E shows a histogram of the posterior samples of the scale parameter SD, σ_w, which is somewhat difficult to interpret; however, as shown in Figure 6.A, there is substantial trial-to-trial variation in the scale of the evoked response.

These results appear to be confirmed by exploratory data analysis. The histogram of the log ISI covariate is shown in Figure 7.A. As described in Appendix A.6, the possible relationship between evoked response amplitude and log ISI was explored by binning trials

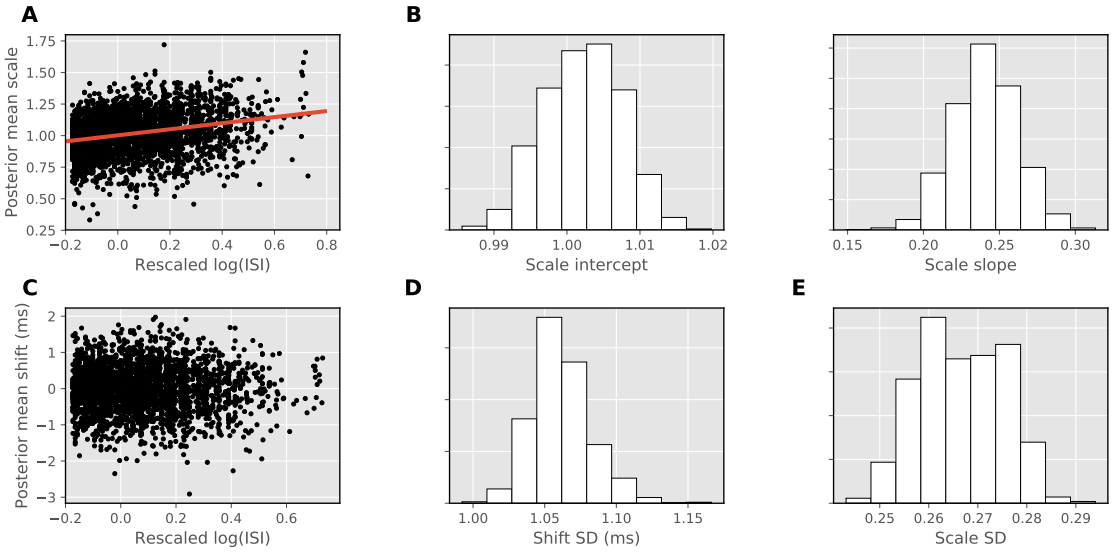

Figure 6: A) Scatter plot of the posterior means of the scale parameters $w^{(n)}$ for each trial against the rescaled log ISI, with a line overlaid representing the posterior mean prediction of the fixed effect of the rescaled log ISI on the scale. B) Across: histograms of posterior samples for the intercept β_0 and slope β_1 of the line relating rescaled log ISI to the scale, which show a non-zero positive slope. C) Scatter plot of the posterior means of the time shifts $\tau^{(n)}$ for each trial against the rescaled log ISI; no fixed effect was included in the model and the inferred time shifts do not appear to vary systematically with ISI. D) Histogram of posterior samples for the shift standard deviation σ_τ, which concentrates around 1 ms. E) Histogram of posterior samples for the scale standard deviation σ_w.

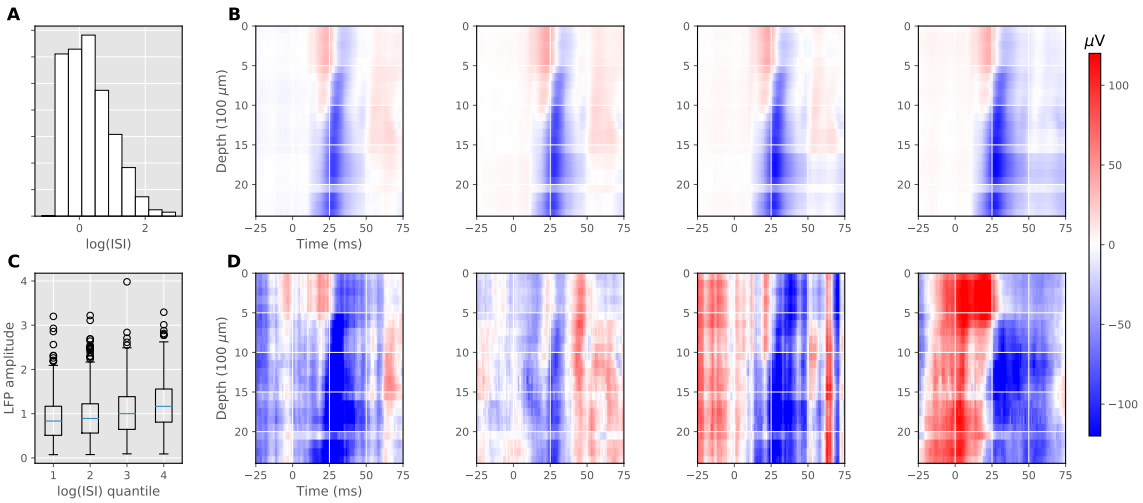

Figure 7: Exploratory data analysis on LFPs. A) Histogram of log(ISI). B) Across: average evoked responses for trials split by log(ISI) quantile (increasing ISI from left to right). The evoked response occurs at about the same time for each quantile, but increases in magnitude with quantile. C) Boxplot of single-trial evoked response amplitude estimates split by log(ISI) quantile. Amplitude appears to increase with log(ISI), but with large, right-skewed trial-to-trial variation. D) Across: four example trials, demonstrating that the evoked response patterns (panel B) are likely highly variable on the single-trial level.

by the quantiles of the log ISI and examining average evoked responses within each quantile; in addition, single-trial estimates of evoked response amplitude were extracted. As shown in Figure 7.B, the average evoked responses appear to increase in strength with log ISI (which increases from left to right). Boxplots of the single-trial evoked response amplitude estimates, split by log ISI quantile, are shown in Figure 7.C. Again, there appears to be an increase in amplitude with log ISI, but also notable is the large variation in amplitudes across trials and the right-skewed distribution of single-trial amplitudes. Four example trials of real LFP data are shown in Figure 7.D to demonstrate the large amount of trial-to-trial variation.

6 Future work and timeline

In this section, I discuss some aspects of the proposed work roughly in order of priority; the action items with a rough timeline are summarized in Table 1.

Hierarchical model for M evoked current events The preliminary results using the hierarchical model to shift and scale the entire CSD mean on single trials appears to yield promising results regarding the amount of trial-to-trial variation and the fixed effect of ISI on the amplitude on a per-trial level. However, I am interested in understanding trial-to-trial covariation of individual evoked current events in the context of information flow. This requires (i) some suitable definition of a mean function based on M evoked current events, which I propose to obtain by thresholding a mean function fitted to all trials (Appendix A.6) and (ii) the implementation of a larger hierarchical model to capture covariation among each of M components.

Tobias suggested that two of the evoked current events (the source/sink pair located near depth 5 in Figure 5.C) should be highly correlated both in time and amplitude. As a preliminary step, I will implement the hierarchical model with $M = 2$ to model covariation between those two evoked current events. I will test in simulations then on a subset of the auditory LFP as a proof of concept. Then, I will extend to larger M (likely on the order of $M = 6$ for the early evoked response). Possible difficulties include increased computation time as the hierarchical model grows in size and difficulties with MCMC convergence. However, as I have a simple $M = 1$ version of the model working, the extension is conceptually easy.

If I encounter computational speed issues or MCMC convergence issues, I will focus only on $M = 2$ and subset the trials/spatiotemporal domain first to simplify the problem. If $M > 1$ fails to yield interesting or proof-of-concept results, I think even $M = 1$ is already potentially of interest for describing trial-to-trial variation. As an alternative, I could also extend the model by fitting both the early (~ 30 ms) and later (~ 100 ms) evoked responses, then letting there be separate variation in the early and late parts (instead of across evoked current events). The potential relation between early and late and the fixed effects of covariates could itself be of interest. Another idea is to try fitting a space-dependent time shift function without thresholding, and to model these hierarchically across trials. The across-trial variation in the time shift functions (for instance, examined by functional PCA) could reveal patterns of information flow.

Develop forward model for Neuropixel and include in GPCSD I have focused so far on LFPs recorded using a linear probe, which prompted the choice of the *a priori* physical model of Equation 5. However, if other types of recordings in two- or three- dimensions were available, the *a priori* physical model could be changed and the GPCSD model could easily incorporate such model. One promising option is Neuropixel data, which will feature about 300 electrodes with two-dimensional coverage. While some previous works have proposed *a priori* physical models for two-dimensional Utah array recordings [Potworowski et al. 2012, Hindriks et al. 2017], I think a different formulation may be needed for Neuropixel data.

I anticipate getting access to some Neuropixel LFP data soon, so I will work to extend the GPCSD model to this case. This will require a reasonable *a priori* physical model to replace Equation 5, then an assessment of whether Equation 4 may be simplified analytically under this physical model to obtain a forward model with an integral that can be easily approximated. The Gaussian process would then have two-dimensional spatial coordinates, which I would start by modeling using a separable covariance function for simplicity, and the new forward model would apply in the same way as in the one-dimensional case. As the evoked response model is still up in the air and I am not sure about the task details of the Neuropixel data, I would focus first on a zero-mean GPCSD implementation which I would use in a an exploratory manner to examine the Neuropixel data. It would be interesting to see how variable the inferred CSD is in the horizontal direction (given that our one-dimensional model assumes constant CSD in the horizontal direction, and I am not aware of any other work that has tried to validate this experimentally). I will also validate this zero-mean two-dimensional GPCSD in simulated data.

User-friendly software for zero-mean GPCSD I plan to release user-friendly software for zero-mean GPCSD in Python and possibly Matlab. This would come with reasonable

default settings for the Gaussian process hyperparameters, potentially also allowing the user to specify some simple constraints on spatial/temporal scales and R. Then it would take multi-trial LFPs and compute a CSD prediction at desired space/time points for each trial, after optimizing the marginal likelihood function to set the hyperparameters. This would allow practitioners to estimate spatiotemporal single-trial CSD profiles and would have some advantages over existing methods (as described in Section 4.3).

Simulations to compare GPCSD to other CSD methods While GPCSD has several stated advantages over existing CSD methods (Section 4.3), I want to do a more thorough comparison on simulated data to demonstrate that GPCSD actually works better in a predictive sense. While the brief visual comparison of tCSD and GPCSD shown in Figure 3 is a start, more investigation is needed.

In particular, I will start with a ground truth CSD from a zero-mean GPCSD model, generate noisy LFPs from the CSD, then compare methods in how well they recover the CSD and/or predict the LFP. I will cross-validate the prediction error both by leaving out single electrodes and by leaving out single trials; the parameters of GPCSD, kCSD, and tCSD (with some smoothing to attempt to remedy some of its drawbacks) would be tuned on the training set and prediction error computed on the held out set.

It also seems in practice that the average of tCSD estimates for a large number of trials is needed to obtain a reasonable estimate of the average evoked CSD, while GPCSD can fit the across-trial shared mean function with a small number of trials, so I want to design a simulation with ground truth across-trial shared mean function and demonstrate that GPCSD can recover it with fewer trials that tCSD.

More thorough data analysis of all data sources So far, I have been focusing on a narrow time window around the early evoked response in the auditory LFPs and I have been testing my methods on a single recording session from a single animal. Once the other items are completed, I will apply the hierarchical model to the other auditory LFPs to determine and interpret results with respect to information flow. Another direction that could be pursued with the auditory LFPs is to look at the learned evoked responses in three dimensions (from non-simultaneous recordings with the probe at different locations).

In addition, I will look at results from Neuropixel data with the two-dimensional model and, though I do not have the data yet, I anticipate two key results. First, the behavior of the predicted CSD in the horizontal direction could be informative on whether the one-dimensional CSD model (which assumes constant CSD on a cylinder around the probe) is reasonable. I am not aware of any existing data that would allow this kind of exploration. Modifications to the cylinder model for 1D could be made if it seemed warranted (such as tapering away to zero instead of constant then zero). Second, hopefully the Neuropixel data will have some other interesting results, but I am not sure what they might be yet; at the least, GPCSD will provide a method for CSD on this type of array for future analyses.

Action item	Target completion date
Finalize proposal document and get committee approval	January 25
Hierarchical model for M evoked current events - Use $M = 2$ and test in simulations - Test $M = 2$ as proof-of-concept on subset of auditory LFPs - Extend to $M > 2$ and test in simulations - Consider space-dependent time shift function (if time)	February 22
Develop forward model for Neuropixel and include in GPCSD - Devise *a priori* physical model and compute 2D forward model - Implement zero-mean GPCSD with 2D forward model - Test on simulated CSD passed through 2D forward model - Test on subset of Neuropixel data and explore possible results	March 15
User-friendly software for zero-mean GPCSD	March 29
Simulations to compare GPCSD to other CSD methods - Leave-one-out LFP predictions: GPCSD vs existing methods - Compare tCSD to GPCSD on shared across-trial evoked	April 19
More thorough data analysis of all data sources - Information flow in several auditory LFP sessions - Compare fitted evoked response among different locations - Neuropixel data: use 2D results to evaluate 1D model assumptions - Neuropixel data: other results (TBD)	May 10
Detailed outline of thesis document	May 17
Draft previous work, background, methods sections	May 31
Loose ends and writing up CSD methods and results	June
Anticipated thesis defense date	July

Table 1: Timeline for proposed work. More discussion of some items is given in the text.

References

Alvarez, M. A., Rosasco, L., Lawrence, N. D., et al. (2012). Kernels for vector-valued functions: a review. *Foundations and Trends® in Machine Learning*, 4(3):195–266.

Arieli, A., Sterkin, A., Grinvald, A., and Aertsen, A. (1996). Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses. *Science*, 273(5283):1868–1871.

Brincat, S. L. and Miller, E. K. (2015). Frequency-specific hippocampal-prefrontal interactions during associative learning. *Nature neuroscience*, 18(4):576.

Buzsáki, G., Anastassiou, C. A., and Koch, C. (2012). The origin of extracellular fields and currents: EEG, ECoG, LFP and spikes. *Nature reviews neuroscience*, 13(6):407.

Duvenaud, D. (2014). *Automatic model construction with Gaussian processes*. PhD thesis, University of Cambridge.

Einevoll, G. T., Kayser, C., Logothetis, N. K., and Panzeri, S. (2013). Modelling and analysis of local field potentials for studying the function of cortical circuits. *Nature Reviews Neuroscience*, 14(11):770.

Hardoon, D. R., Szedmak, S., and Shawe-Taylor, J. (2004). Canonical correlation analysis: An overview with application to learning methods. *Neural computation*, 16(12):2639–2664.

Herreras, O. (2016). Local field potentials: myths and misunderstandings. *Frontiers in neural circuits*, 10:101.

Hindriks, R., Schmiedt, J., Arsiwalla, X. D., Peter, A., Verschure, P. F., Fries, P., Schmid, M. C., and Deco, G. (2017). Linear distributed source modeling of local field potentials recorded with intra-cortical electrode arrays. *PloS one*, 12(12):e0187490.

Hyvärinen, A. (2005). Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6(Apr):695–709.

Kajikawa, Y. and Schroeder, C. E. (2011). How local is the local field potential? *Neuron*, 72(5):847–858.

Klein, N., Orellana, J., Brincat, S. L., Miller, E. K., and Kass, R. E. (submitted 2018). Torus graphs for multivariate phase coupling analysis. *Annals of Applied Statistics*.

Kropf, P. and Shmuel, A. (2016). 1D current source density (CSD) estimation in inverse theory: a unified framework for higher-order spectral regularization of quadrature and expansion-type CSD methods. *Neural computation*, 28(7):1305–1355.

Lachaux, J.-P., Rodriguez, E., Martinerie, J., and Varela, F. J. (1999). Measuring phase synchrony in brain signals. *Human brain mapping*, 8(4):194–208.

Lindén, H., Tetzlaff, T., Potjans, T. C., Pettersen, K. H., Grün, S., Diesmann, M., and Einevoll, G. T. (2011). Modeling the spatial reach of the LFP. *Neuron*, 72(5):859–872.

Marek, S., Tervo-Clemmens, B., Klein, N., Foran, W., Ghuman, A. S., and Luna, B. (2018). Adolescent development of cortical oscillations: Power, phase, and support of cognitive maturation. *PLoS Biology*, 16(11):e2004188.

Nicholson, C. (1973). Theoretical analysis of field potentials in anisotropic ensembles of neuronal elements. *IEEE Transactions on Biomedical Engineering*, (4):278–288.

Nicholson, C. and Freeman, J. A. (1975). Theory of current source-density analysis and determination of conductivity tensor for anuran cerebellum. *Journal of neurophysiology*, 38(2):356–368.

Nicholson, C. and Llinas, R. (1971). Field potentials in the alligator cerebellum and theory of their relationship to Purkinje cell dendritic spikes. *Journal of Neurophysiology*, 34(4):509–531.

Pettersen, K. H., Devor, A., Ulbert, I., Dale, A. M., and Einevoll, G. T. (2006). Current-source density estimation based on inversion of electrostatic forward solution: effects of finite extent of neuronal activity and conductivity discontinuities. *Journal of neuroscience methods*, 154(1-2):116–133.

Pitts, W. (1952). Investigations on synaptic transmission. In *Cybernetics, Trans. 9th Conf. Josiah Macy, New York*, pages 159–162.

Potworowski, J., Jakuczun, W., Łski, S., and Wójcik, D. (2012). Kernel current source density method. *Neural computation*, 24(2):541–575.

Rasmussen, C. E. and Williams, C. K. (2006). *Gaussian process for machine learning*. MIT press.

Rodu, J., Klein, N., Brincat, S. L., Miller, E. K., and Kass, R. E. (2018). Detecting multivariate cross-correlation between brain regions. *Journal of neurophysiology*.

Saatçi, Y. (2012). *Scalable inference for structured Gaussian process models*. PhD thesis, University of Cambridge.

Salvatier, J., Wiecki, T. V., and Fonnesbeck, C. (2016). Probabilistic programming in Python using PyMC3. *PeerJ Computer Science*, 2:e55.

Särkkä, S. (2011). Linear operators and stochastic partial differential equations in Gaussian process regression. In *International Conference on Artificial Neural Networks*, pages 151–158. Springer.

Schacke, K. (2004). On the Kronecker product. *Master's thesis, University of Waterloo*.

Szymanski, F. D., Rabinowitz, N. C., Magri, C., Panzeri, S., and Schnupp, J. W. (2011). The laminar and temporal structure of stimulus information in the phase of field potentials of auditory cortex. *Journal of Neuroscience*, 31(44):15787–15801.

A Appendix

A.1 One-dimensional *a priori* physical model details

Substituting Equation 5 into Equation 4 yields

$$\phi(x,y,z) = -\frac{1}{4\pi\sigma}\int_a^b \int\int_{x^2+y^2\leq R} \frac{g(z')}{\sqrt{(x-x')^2+(y-y')^2+(z-z')^2}}\,dx'\,dy'\,dz'. \tag{18}$$

We will assume x and y are inside the cylinder (as typically, we assume we observe ϕ at the center of the cylinder). Changing to polar coordinates, we define $r^2 = (x-x')^2+(y-y')^2$ as the variable radius inside the cylinder and use the substitution $dx'\,dy'\,dz' = r\,d\theta\,dr\,dz'$ to obtain

$$\phi(x,y,z) = -\frac{1}{4\pi\sigma}\int_a^b \int_0^R \int_0^{2\pi} \frac{rg(z')}{\sqrt{(z-z')^2+r^2}}\,d\theta\,dr\,dz' \tag{19}$$

$$= -\frac{1}{2\sigma}\int_a^b g(z') \int_0^R \frac{r}{\sqrt{(z-z')^2+r^2}}\,dr\,dz' \tag{20}$$

$$= -\frac{1}{2\sigma}\int_a^b g(z') \left[\sqrt{(z-z')^2+R^2} - \sqrt{(z-z')^2}\right] dz'. \tag{21}$$

Notice that after integration, this is no longer a function of x or y, so we can simply write

$$\phi(z) = -\frac{1}{2\sigma}\int_a^b g(z') \left[\sqrt{(z-z')^2+R^2} - \sqrt{(z-z')^2}\right] dz'. \tag{22}$$

To better understand how R affects the ϕ, I factor out R:

$$\phi(z) = -\frac{R}{2\sigma}\int_a^b g(z') \underbrace{\left[\sqrt{\left(\frac{r}{R}\right)^2+1} - \sqrt{\left(\frac{r}{R}\right)^2}\right]}_{b(r;R)} dz' \tag{23}$$

where $r = z - z'$ and $b(r;R)$ is a weight function with a maximum value of 1 when $r = 0$.

A.2 Kernel CSD (kCSD) details

The observed potential corresponding to one source, which we will call b_i, is

$$b_i(x,y,z) = \mathcal{A}C(x,y,z) \equiv \frac{1}{4\pi\sigma}\int\int\int \frac{\tilde{b}_i(x,y,z)}{\sqrt{(x-x')^2+(y-y')^2+(z-z')^2}} dx'dy'dz' \tag{24}$$

and the overall potential from all sources adds linearly:

$$V(x,y,z) = \mathcal{A}C(x,y,z) = \sum_{i=1}^{M} a_i b_i(x,y,z) \tag{25}$$

Defining $\mathbf{x} = (x,y,z)$, the kernel function between LFP basis functions is:

$$K(\mathbf{x},\mathbf{x}') \equiv \sum_{i=1}^{M} b_i(\mathbf{x}) b_i(\mathbf{x}')$$

This defines an RKHS where any function can be expressed as

$$f(\mathbf{x}) = \sum_{j=1}^{\ell} \alpha_j K(\mathbf{x}_j, \mathbf{x}) = \sum_{i=1}^{M} a_i b_i(\mathbf{x})$$

where $a_i = \sum_{j=1}^{\ell} \alpha_j b_i(\mathbf{x}_j)$. The minimum-norm solution that corresponds to interpolating the observed LFPs is given by

$$\boldsymbol{\beta} = \mathbf{K}^{-1}\mathbf{V}$$

where the estimated potentials are $V^*(\mathbf{x}) = \sum_{i=1}^{N} \beta_i K(\mathbf{x}_i, \mathbf{x})$. That is, we know from RKHS that $a_i = \sum_{k=1}^{N} \beta_k b_i(\mathbf{x}_k)$ such that

$$V^*(\mathbf{x}) = \sum_{i=1}^{M} a_i b_i(\mathbf{x}) = \sum_{i=1}^{M} b_i(\mathbf{x}) \sum_{k=1}^{N} \beta_k b_i(\mathbf{x}_k) = \sum_{k=1}^{N} \beta_k \sum_{i=1}^{M} b_i(\mathbf{x}_k) b_i(\mathbf{x}_k) = \sum_{k=1}^{N} \beta_k K(\mathbf{x}_k, \mathbf{x}).$$

Note that adding Tikhonov regularization simply adds λI inside the inverse and loosens the requirement that the LFPs are interpolated. Given this set of potentials V^* there is now a unique current distribution:

$$C^*(\mathbf{x}) = \sum_{j=1}^{M} a_j \tilde{b}_j(\mathbf{x}) = \sum_{i=1}^{N} \beta_i \sum_{j=1}^{M} b_j(\mathbf{x}_j) \tilde{b}_j(\mathbf{x}) \equiv \sum_{i=1}^{N} \beta_i \tilde{K}(\mathbf{x}_i, \mathbf{x})$$

where $\tilde{K}$ is the cross-kernel induced by the operator $\mathcal{A}$. Note that iCSD [Pettersen et al. 2006] is a special case of kCSD with either delta functions, piecewise-constant functions, or cubic spline functions used as the basis, but no regularization is applied, and that extensive comparisons of different kernels and regularization schemes was carried out in Kropf and Shmuel [2016].

A.3 Relating Gaussian process regression to RKHS regression

Under a zero-mean version of GPCSD, the form of the prediction in Equation 30 looks very similar to the prediction given by kCSD method discussed in Section 3.3, which happens because the two methods can be seen as two different frameworks for arriving at similar estimators. In kCSD, one specifies a set of basis functions, along with their centers and hyperparameters; in Potworowski et al. [2012], M Gaussian bumps with width ℓ were used, with a suggestion to use a large enough M to densely cover the electrode recording area, and to use wide enough ℓ so that the basis functions overlapped partially. From the basis functions, a kernel function was calculated in Equation 10 which functions in a similar manner as the Gaussian process covariance function of GPCSD. So the primary difference is in how the model is specified: through selection and placement of a number of basis functions or through direct specification of a suitable covariance function.

In fact, as shown in [Rasmussen and Williams 2006, p. 84], the squared exponential covariance function is equivalent to using infinitely many densely-spaced Gaussian-shaped basis functions, and the width of the Gaussian basis functions is proportional to the characteristic lengthscale of the covariance function. This implies that kCSD would give similar

results to a Gaussian process with squared exponential lengthscale if many densely-spaced basis functions were used. However, the Gaussian process approach precludes choosing M or using potentially computationally taxing cross-validation to select the width of the basis functions. Instead, maximum marginal likelihood can easily be used to select the lengthscale for the Gaussian process along with other hyperparameters. In addition, it is possible to specify more complex covariance functions, such as additive or product covariance functions, for which it may be difficult to directly specify the corresponding basis functions. The Gaussian process approach is also easily extendable to spatiotemporal processes, can incorporate a random mean function, can include multi-scale covariance functions, and provides a conditional distribution of CSD predictions given the LFP data (while kCSD only provides a point estimate).

A.4 Gaussian processes under linear operators

Consider applying a linear operator $\mathcal{A}_x$, such as differentiation or integration to a Gaussian process f. As discussed in Särkkä [2011], the result of applying a linear operator to $f(x)$ is still a Gaussian process; that is,

$$\mathcal{A}_x f(x) = h(x) \tag{26}$$

where $h(x)$ is a Gaussian process with covariance function influenced by the linear operator:

$$h(x) \sim GP(\mathcal{A}_x \mu(x), \mathcal{A}_x k(x, x') \mathcal{A}_x^T) \tag{27}$$

Then the joint distribution of vectors of observations $\mathbf{f_x}$ and $\mathbf{h_x}$ (for simplicity, at the same $\mathbf{x}$, though this is not essential) is:

$$\begin{bmatrix} \mathbf{f_x} \\ \mathbf{h_x} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_\mathbf{x} \\ \mathcal{A}_x \boldsymbol{\mu}_\mathbf{x} \end{bmatrix}, \begin{bmatrix} \mathbf{K_{x,x}} & \mathcal{A}_x \mathbf{K_{x,x}} \\ \mathbf{K_{x,x}} \mathcal{A}_x^T & \mathcal{A}_x \mathbf{K_{x,x}} \mathcal{A}_x^T \end{bmatrix} \right). \tag{28}$$

Suppose we have observations of $h(x)$ but want to predict values of $f(x)$. Under this framework, is natural to propose mean and covariance functions μ and k for f; application of the linear operator induces mean and covariance functions for the observed h. Then predictions (conditional on the observed values of h) can be made for any $\mathbf{x}'$ for either f or h. To be more realistic, the model could include additive noise on h which would amend the h-block of the covariance matrix to be $\mathcal{A}\mathbf{K_{x,x}}\mathcal{A}^T + \sigma^2\mathbf{I}$. With noise, due to properties of multivariate Gaussians, we have the following conditional distribution:

$$\mathbf{f_x} | \mathbf{h_x} \sim \mathcal{N}(\boldsymbol{\mu}^*, \mathbf{K}^*) \tag{29}$$

where

$$\boldsymbol{\mu^*}_\mathbf{x} = \boldsymbol{\mu}_\mathbf{x} + \mathcal{A}_x \mathbf{K_{x,x}} [\mathcal{A}\mathbf{K_{x,x}}\mathcal{A}^T + \sigma^2\mathbf{I}]^{-1} (\mathbf{h_x} - \mathcal{A}_x \boldsymbol{\mu}_\mathbf{x}) \tag{30}$$

$$\mathbf{K^*}_\mathbf{x,x} = \mathbf{K_{x,x}} - \mathcal{A}_x \mathbf{K_{x,x}} [\mathcal{A}_x \mathbf{K_{x,x}} \mathcal{A}_x^T + \sigma^2\mathbf{I}]^{-1} \mathbf{K_{x,x}} \mathcal{A}_x \tag{31}$$

so that predictions for $\mathbf{f_x}$ given $\mathbf{h_x}$ are given by $\boldsymbol{\mu^*}_\mathbf{x}$. Similar expressions can be derived for predictions for different inputs $\mathbf{x}'$ but are omitted for brevity.

In addition, using the maximum marginal likelihood is an attractive way to tune hyperparameters because, as discussed in [Rasmussen and Williams 2006, Ch. 5.2], it incorporates

a trade-off between model complexity and fit and can be interpreted as penalizing the complexity of the underlying function. In particular, suppose we have noisy observations of h at locations $\mathbf{x}$. Equation 28 shows that the joint distribution of $\mathbf{h_x}$ and $\mathbf{f_{x'}}$ is multivariate Gaussian. Then the marginal density for $\mathbf{h_x}$,

$$p(\mathbf{h_x}) = \int p(\mathbf{h_x}|\mathbf{f_{x'}})p(\mathbf{f_{x'}})\, d\mathbf{f_{x'}}, \tag{32}$$

using properties of multivariate Gaussians, is

$$\mathbf{h_x} \sim \mathcal{N}(\mathcal{A}_x \boldsymbol{\mu}_\mathbf{x},\, \mathcal{A}_x \mathbf{K_x} \mathcal{A}_x^T + \sigma^2 \mathbf{I}). \tag{33}$$

This marginal likelihood can be maximized with respect to hyperparameters of the mean and covariance functions.

A.5 Gaussian processes with Kronecker-product-plus-noise covariances

Fortunately, the separable covariance formulation also eases the computational burden. In particular, the covariance matrix evaluated at vectors of space-time points is a Kronecker product

$$\mathbf{K_{z,t;z',t'}} = \mathbf{K^z_{z,z'}} \otimes \mathbf{K^t_{t,t'}}. \tag{34}$$

If there is no additional additive noise, the inverse may be easily computed on $O(T^3 + D^3)$ time using properties of the Kronecker product [Schacke 2004]. However, if additive noise is included, other properties of Kronecker products must be used to obtain some computational speed-ups compared to naively inverting the matrix, as discussed in Saatçi [2012].

A.6 Simulation and data procedures

Exploratory data analysis of real LFPs I began by examining the average evoked responses to determine whether features of the evoked response appear to modulate with ISI. Because the ISI is very right-skewed, I used the natural logarithm of the ISI, then partitioned the trials into groups based on the quantiles of the log ISI and compute average evoked responses for each log ISI quantile. To explore the trial-to-trial variation in peak absolute amplitude of the evoked response by log ISI quantile, I first smoothed each trial with a Gaussian filter of width 1.5 ms in time and 1.5 μm in the spatial direction. Then, I found the space-time location of peak absolute amplitude on the average evoked response of all trials, then extracted the absolute amplitude from each trial at this space-time point to represent the peak absolute amplitude on a single-trial basis.

Simulation study As discussed in Section 3.2, the forward model depends on a parameter R which indicates how far the current source and sink activity extends in the directions perpendicular to the recording probe. In particular, Equation 6 shows that R affects not only the amplitude of the LFPs, but also the weight function inside the integral operator controlling the spatial smoothing of the LFPs.

To demonstrate this effect, I sampled a spatiotemporal CSD from a zero-mean Gaussian process with hyperparameters similar to those fit to real data and with $a = 0, b = 24$ representing the boundaries of the nonzero CSD in the z direction. I then applied the forward

model with $R \in \{0.1, 0.5, 1, 5, 20\}$ to generate noise-free LFPs corresponding to each R, where the conductivity scalar was taken to be $\sigma = 1$ and the spacing between electrodes was taken to be 1. The temporal resolution is the same as the original data, but the spatial resolution is higher than observed LFPs (300 points spaced evenly between a and b).

While these simulated LFPs are much more dense spatially than typical recordings, I applied tCSD to the LFPs at this high spatial resolution without any additional measurement noise to see how well tCSD could perform for each R under ideal conditions. I also applied GPCSD to the LFPs using the true hyperparameters and true R, but for computational reasons, subsampled the spatial dimension of the LFPs by a factor of 6 before predicting the latent CSD at the full spatial resolution.

GPCSD model applied to real LFPs with a single across-trial mean First, I fit the Gaussian process covariance function, the forward model parameter R, and a single mean function shared across all trials using maximum marginal likelihood. I parameterized the mean as a mixture of Gaussian components with possibly negative scaling factors α_j:

$$\mu(z,t) = \sum_{j=1}^{J} \alpha_j \exp\left(-\frac{(z-\mu_{z,j})^2}{2\sigma_{z,j}^2}\right) \exp\left(-\frac{(t-\mu_{t,j})^2}{2\sigma_{t,j}^2}\right). \tag{35}$$

I chose this formulation because it is straightforward to obtain surfaces that are nonstationary in time using prior information on the temporal location and duration of the evoked response. While fitting mixtures of Gaussians can be unstable, sufficiently large J initialized evenly across space should be able to approximate arbitrary evoked response shapes, and the large number of trials used to estimate $\mu(z,t)$ should make the resulting surface fairly stable to different initializations. I randomly selected 1000 of the trials and used $J = 30$ components in Equation 35 to fit the mean function surface (after checking that a range of $J > 20$ appeared to give similar results).

The spatial covariance function was taken to be squared exponential with unit variance, while the temporal covariance function was a sum of a squared exponential (initialized to bias it toward slow activity) and a Matérn (initialized to bias it toward fast activity). Letting the spacing between electrodes be $\Delta z = 1$ and the conductivity scalar be $\sigma = 1$, I used the assumption that all CSD activity is zero at depths of more than $2\Delta z$ outside the first and last electrode.

First, the average evoked response was subtracted to yield zero-mean data, and the marginal likelihood of Equation 33 was optimized using L-BFGS over the Gaussian process covariance hyperparameters and R. Then, given the estimated covariance hyperparameters and R, the marginal likelihood was optimized using L-BFGS over the mean function parameters. This process was repeated one more time to ensure the estimates did not change due to this step-wise process. For visualization of separate evoked current events, I thresholded the resulting mean function with $h_\mu = 0.25 \max(|\mu(z,t)|)$.

Hierarchical model for trial-to-trial variation applied on real LFPs As a proof of concept, I used a simplified version of the hierarchical model for across-trial variation in the evoked response which models the trial-specific mean as a shifted (in time) and scaled version of the fitted across-trial mean from the previous step. I used a modular fitting

approach, in which I first used maximum marginal likelihood to estimate R, the covariance hyperparameters, and a single mean function shared across all trials (given in Equation 35). Then, conditional on these point estimates, inference is performed on the quantities in the hierarchical model of Equation 15 via MCMC sampling of the posterior.

Because the inter-stimulus-interval (ISI) appears to affect the evoked response amplitude in binned LFP average evoked responses (Figure 7.D), I include a scalar covariate $x^{(n)}$ as part of the model for the trial-specific scale:

$$w^{(n)} = \beta_0 + \beta_1 x^{(n)} + \epsilon_w^{(n)} \tag{36}$$

Specifically, I specify the following priors and hyperpriors and estimate the posterior distribution using Metropolis-Hastings sampling with Normal proposal distributions using PyMC with an internal non-centered parameterization [Salvatier et al. 2016]:

$$\begin{aligned} \beta_0 \sim \mathcal{N}(0, 25),\ \beta_1 \sim \mathcal{N}(0, 25) \\ \epsilon_w^{(n)} \sim \mathcal{N}(0, \sigma_w^2),\ \sigma_w \sim \mathcal{HN}(0, 0.25) \\ \tau^{(n)} \sim \mathcal{N}(0, \sigma_\tau^2),\ \sigma_\tau \sim \mathcal{HN}(0, 0.25) \end{aligned} \tag{37}$$

where $\mathcal{HN}$ is a half-Normal distribution. For MCMC, two chains were used with 2,000 samples used to tune the proposal distributions then discarded before 5,000 samples were drawn from each chain, resulting in a total of 10,000 samples.