A Data-Driven Framework for Neural Field Modeling

D. R. Freestone [a,b,c,1,*], P. Aram[d,1], M. Dewar[c,e], K. Scerri[f], D. B. Grayden[a,b], V. Kadirkamanathan[d]

[a]*Department of Electrical and Electronic Engineering, University of Melbourne, Melbourne, VIC, Australia*
[b]*The Bionic Ear Institute, East Melbourne, VIC, Australia*
[c]*Institute for Adaptive and Neural Computation, University of Edinburgh, Edinburgh, UK*
[d]*Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield, UK*
[e]*Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY, USA*
[f]*Department of Systems and Control Engineering, University of Malta, Msida, MSD, Malta*

Abstract

This paper presents a framework for creating neural field models from electrophysiological data. The Wilson and Cowan or Amari style neural field equations are used to form a parametric model, where the parameters are estimated from data. To illustrate the estimation framework, data is generated using the neural field equations incorporating modeled sensors enabling a comparison between the estimated and true parameters. To facilitate state and parameter estimation, we introduce a method to reduce the continuum neural field model using a basis function decomposition to form a finite-dimensional state-space model. Spatial frequency analysis methods are introduced that systematically specify the basis function configuration required to capture the dominant characteristics of the neural field. The estimation procedure consists of a two-stage iterative algorithm incorporating the unscented Rauch-Tung-Striebel smoother for state estimation and a least squares algorithm for parameter estimation. The results show that it is theoretically possible to reconstruct the neural field and estimate intracortical connectivity structure and synaptic dynamics with the proposed framework.

Keywords: Neural Field Model, Nonlinear Estimation, Intracortical Connectivity, Nonlinear Dynamics

1. Introduction

Generating physiologically plausible neural field models is of great importance for studying brain dynamics at the mesoscopic and macroscopic scales. While our understanding of the function of neurons is well developed, the overall behaviour of the brain's mesoscopic and macroscopic scale dynamics remains largely theoretical. Understanding the brain at this level is extremely important since it is at this scale that pathologies such as epilepsy, Parkinson's disease and schizophrenia are manifested.

---

*Corresponding author
Email addresses:* `dfreestone@bionicear.org` (D. R. Freestone), `p.aram@sheffield.ac.uk` (P. Aram),
`mike.dewar@columbia.edu` (M. Dewar), `kenneth.scerri@um.edu.mt` (K. Scerri), `grayden@unimelb.edu.au` (D. B. Grayden),
`visakan@sheffield.ac.uk` (V. Kadirkamanathan)
[1]Authors contributed equally to this work.

Mathematical neural field models provide insights into the underlying physics and dynamics of electroencephalography (EEG) and magnetoencephalography (MEG) (see **??** for recent reviews). These models have demonstrated possible mechanisms for the genesis of neural rhythms (such as the alpha and gamma rhythms) (**??**), epileptic seizure generation (**???**) and insights into other pathologies (**??**) that would be difficult to gain from experimental data alone.

Unfortunately, the use of these models in the clinic has been limited, since they are constructed for "general" brain dynamics whereas pathologies almost always have unique underlying patient-specific causes. Patient-specific data from electrophysiological recordings is readily available in the clinical setting, particularly from epilepsy surgery patients, suggesting an opportunity to make the patient-specific link to models of cortical dynamics. Furthermore, recent technological advances have driven an increased level of sophistication in recording techniques, with dramatic increases in spatial and temporal sampling (**?**). However, the mesoscopic and macroscopic neural dynamic state is not directly observable in neurophysiological data, making predictions of the underlying physiology inherently difficult.

For models to be clinically viable, they must be patient-specific. A possible approach to achieve this would be to fit a general continuum neural field model, like the **?** (WC) or **?** models, or a neural mass model like the **?** model, to patient-specific data. Fitting the neural models to individuals is a highly non-trivial task. Recently, however, this task has been approached from a number of standpoints. The first paper on patient-specific modeling (to the authors' best knowledge) came from **?**, where they fit the neural mass model of **?** and **?** to EEG data using the local linearization filter (**???**). This paper demonstrated for the first time the feasibility of assimilating data with neural models of EEG.

Perhaps the most accepted estimation framework for neural mass models is dynamical causal modeling (DCM) (**??**), which has been proposed for studying evoked potential dynamics. This framework can be viewed as an extension to the work of **?** allowing for coupling of neural masses. Via a Bayesian inference scheme, DCM estimates the long range (cortico-cortical) connectivity structure between the specific isolated brain regions that best explains a given data set using the model of **?**.

Data-driven neural modeling was extended to the continuum approximation by **?**, where they proposed an estimation framework based on a linear damped wave equation. Using a Kalman filter maximum likelihood framework, they improved on the low resolution electromagnetic tomography (LORETA) method for solving the inverse EEG problem. Another continuum neural field model-based estimator was developed by **?**, as an extension to the DCM framework. They argue that using a spatially extended continuum field model is superior for explaining cortical function than the neural mass DCM, since field models can explain a richer repertoire of dynamics such as traveling waves and bump solutions.

Another recent approach estimates the parameters of a modified WC neural field model using an unscented Kalman filter (**?**). This work takes a systems theoretic approach to the neural estimation problem, successfully demonstrating that it is possible to perform state estimation and control on spatiotemporal

neural fields. This marks the first step in what has the potential to revolutionize the treatment of many neurological diseases where therapeutic electrical stimulation is viable. For other valuable contributions to data-driven neural modeling, see **?**, **?**, and **?**.

We present an extension to the work of **?** by establishing a framework for estimating the state of the WC equations for larger scale (more space) systems via a systematic model reduction procedure. In addition, a new method is presented for estimating the connectivity structure and the synaptic dynamics. Until now, model-based estimation of local intracortical connectivity has not been reported in the literature (to the best of the authors' knowledge). Our study also extends recent work which shows that it is possible to estimate local coupling of spatiotemporal systems using techniques from control systems theory and machine learning (**?**). The key development of this previous work was to represent a spatiotemporal system as a standard state-space model, with the number of states independent of the number of observations (recording electrodes in this case). In addition, the appropriate model selection tools have been developed (**?**) allowing for the application of the technique to neural fields. This paper extends the linear framework of **?** and **?** to the nonlinear case required for the neural field equations.

Modeling the neural dynamics within this framework has a distinct advantage over the more standard multivariate auto-regressive (MVAR) models: the number of parameters to define the spatial connectivity is considerably smaller than the number of AR coefficients typically required to achieve the model complexity.

In this paper, we demonstrate for the first time how an intracortical connectivity kernel can be inferred from data, based on a variant of the **?** neural field model. This work provides a fundamental link between the theoretical advances in neural field modeling and high resolution intracranial electrophysiological data. To illustrate the estimation framework, data is generated using the neural field equations incorporating modeled sensors enabling a comparison to be made between the estimated and true parameters. The paper proceeds by first describing the continuum neural field equations that are used as the cortical model. Then a finite-dimensional neural field model is derived. The model is reduced by approximating the neural field using a set of continuous basis functions, weighted by a finite dimensional state vector. The next section establishes conditions, using spatial frequency analysis, for both sensor and basis function spacing and width, such that the dominant dynamics of the neural field can be represented by the reduced model. The state and parameter estimation procedure is described in the following section. The results for the spatial frequency analysis and parameter estimation are then presented. Finally, the implications and limitations of this framework are discussed along with planned future developments.

2. Methods

2.1. Neural Field Model

Neural field models relate mean firing rates of pre-synaptic neural populations to mean post-synaptic membrane potentials. They are popular as they are parsimonious yet have a strong link with the underlying physiology. Each neural population represents a functional cortical processing unit, such as a column. The columnar organization of the cortex is continuous, where pyramidal cells are members of many columns. In general, cortical structure can be modeled in a physiologically plausible manner as being locally homogeneous (in short range intracortical connectivity) and heterogeneous (in long range cortico-cortical and corticothalamic connectivity) (**??**). In certain regions of the cortex, each column is thought to be connected locally via symmetric short range local excitation with surround inhibition (**?**). For example, this structural organization is most studied in the visual system, where the surrounding inhibition effectively tunes a cortical column to a particular receptive visual field (**?**). Neural field models are descriptive of a range of neurodynamics of the cortex such as evoked potentials, visual hallucinations and epileptic behaviour (**???**). Field models are also capable of generating complex spatial patterns of activity such as Turing patterns, spirals and traveling oscillations (**???**).

It is an implicit assumption that the neural field model (in equation 12) provides an apt description of the cortical dynamics recorded from a specific subject. Although models of this form are capable of describing a variety of cortical dynamics, there will be without doubt a mismatch between the cortex and the model. Nevertheless, there is a sufficient volume of interesting results from theoretical studies using the WC field equations that warrants the data assimilation framework. An advantage in using a lumped-parameter field model over a more detailed mathematical description is that the myriad of parameters that might influence excitability, such as specific ion concentrations, are considered to be lumped into parameters (connectivity kernel coefficients) that effectively describe the net system gain. Therefore, there are less parameters to estimate. A major challenge in model-based data analysis for the mass action of the brain is to make models sufficiently detailed, such that the parameters are meaningful, but simple enough to yield clear insights that can be related to theoretical studies (from both neuroscience and engineering perspectives).

2.2. Integro-Difference Equation Neural Field Model

The combination of modeling techniques in this paper leads to a large amount of notation, so a reference of the symbols used is provided in Table 1. The model relates the average number of action potentials $g(\mathbf{r}, t)$ arriving at time t and position $\mathbf{r}$ to the local post-synaptic membrane voltage $v(\mathbf{r}, t)$. The post-synaptic potentials generated at a neuronal population at location $\mathbf{r}$ by action potentials arriving from all other connected populations is described by

$$v\left(\mathbf{r}, t\right) = \int_{-\infty}^{t} h\left(t - t'\right) g\left(\mathbf{r}, t'\right) \, dt'. \tag{1}$$

The post-synaptic response kernel $h(t)$ is described by

$$h(t) = \eta(t) \exp\left(-\zeta t\right), \tag{2}$$

where $\zeta = \tau^{-1}$, τ is the synaptic time constant and $\eta(t)$ is the Heaviside step function. Non-local interactions between cortical populations at positions $\mathbf{r}$ and $\mathbf{r}'$ are described by

$$g\left(\mathbf{r}, t\right) = \int_{\Omega} w\left(\mathbf{r}, \mathbf{r}'\right) f\left(v\left(\mathbf{r}', t\right)\right) d\mathbf{r}', \tag{3}$$

where $f(\cdot)$ is the firing rate function, $w(\cdot)$ is the spatial connectivity kernel and Ω is the spatial domain representing a cortical sheet or surface. The connectivity kernel is typically a "Mexican hat" or "wizard hat" function, which describes local excitation and surround inhibition. An example of the Mexican hat connectivity kernel, that is a weighted sum of Gaussians, used in this paper is shown in Fig. 1. To demonstrate flexibility in the estimation algorithm and to model larger cortical regions, a third component of the kernel is introduced describing weak longer-range excitation. The exact shape of this kernel is assumed to vary across subjects, and hence needs to be inferred from subject-specific data.

The firing rate of the presynaptic neurons is related to the post-synaptic membrane potential by the sigmoidal activation function

$$f\left(v\left(\mathbf{r}', t\right)\right) = \frac{1}{1 + \exp\left(\varsigma\left(v_0 - v\left(\mathbf{r}', t\right)\right)\right)}. \tag{4}$$

The parameter v_0 describes the firing threshold of the neural populations and ς governs the slope of the sigmoid. By substituting equation 3 into equation 1 we get the spatiotemporal model

$$v\left(\mathbf{r}, t\right) = \int_{-\infty}^{t} h\left(t - t'\right) \int_{\Omega} w\left(\mathbf{r}, \mathbf{r}'\right) f\left(v\left(\mathbf{r}', t'\right)\right) d\mathbf{r}' dt'. \tag{5}$$

To obtain the standard integro-differential equation form of the model, we use the fact that the synaptic response kernel is a Green's function of a linear differential equation defined by the differential operator $D = d/dt + \zeta$. A Green's function satisfies

$$D h\left(t\right) = \delta\left(t\right), \tag{6}$$

where $\delta(t)$ is the Dirac-delta function. Applying the differential operator D to equation 1 gives

$$\mathrm{D}v\left(\mathbf{r},t\right) = \mathrm{D}\left(h * g\right)\left(\mathbf{r},t\right) \tag{7}$$

$$= \left(\mathrm{D}h * g\right)\left(\mathbf{r},t\right) \tag{8}$$

$$= \left(\delta\left(t\right) * g\right)\left(\mathbf{r},t\right) \tag{9}$$

$$= g\left(\mathbf{r},t\right) \tag{10}$$

where $*$ denotes the convolution operator. This gives the standard form of the model

$$\frac{dv\left(\mathbf{r},t\right)}{dt} + \zeta v\left(\mathbf{r},t\right) = \int_{\Omega} w\left(\mathbf{r},\mathbf{r}'\right) f\left(v\left(\mathbf{r}',t\right)\right) d\mathbf{r}'. \tag{11}$$

To arrive at the integro-difference equation (IDE) form of the model, we discretize time using a first-order Euler method (see Appendix A for derivation) giving

$$v_{t+T_s}\left(\mathbf{r}\right) = \xi v_t\left(\mathbf{r}\right) + T_s \int_{\Omega} w\left(\mathbf{r},\mathbf{r}'\right) f\left(v_t\left(\mathbf{r}'\right)\right) d\mathbf{r}' + e_t\left(\mathbf{r}\right), \tag{12}$$

where T_s is the time step, $\xi = 1 - T_s\zeta$ and $e_t(\mathbf{r})$ is an $i.i.d.$ disturbance such that $e_t(\mathbf{r}) \sim \mathcal{GP}(\mathbf{0},\gamma(\mathbf{r}-\mathbf{r}'))$. Here $\mathcal{GP}(\mathbf{0},\gamma(\mathbf{r}-\mathbf{r}'))$ denotes a zero mean Gaussian process with spatial covariance function $\gamma(\mathbf{r}-\mathbf{r}')$ (?). The disturbance is added to account for model uncertainty and unmodeled inputs. To simplify the notation, the index of the future time sample, $t + T_s$, shall be referred to as $t + 1$ throughout the rest of the paper.

The mapping between the membrane voltage and the electrophysiological data, denoted by $\mathbf{y}_t$, is modeled using the observation function that incorporates sensors with a spatial extent by

$$y_t(\mathbf{r}_n) = \int_{\Omega} m\left(\mathbf{r}_n - \mathbf{r}'\right) v_t\left(\mathbf{r}'\right) d\mathbf{r}' + \varepsilon_t(\mathbf{r}_n), \tag{13}$$

where $m\left(\mathbf{r}_n - \mathbf{r}'\right)$ is the observation kernel, $\mathbf{r}_n$ defines the location of the electrodes in the field, $n = 0,...,n_y - 1$ indexes the sensors and $\varepsilon_t(\mathbf{r}_n) \sim \mathcal{N}\left(0,\boldsymbol{\Sigma}_\varepsilon\right)$ denotes a multivariate normal distribution with mean zero and the covariance matrix $\boldsymbol{\Sigma}_\varepsilon = \sigma_\varepsilon^2 \mathbf{I}$, where $\mathbf{I}$ is the identity matrix. Since we are considering intracranial measurements recorded directly from the surface of the cortex or within the brain, the lead field is not modeled by the observation equation.

When considering the measurement of a point source in a conductive homogeneous field measured by a point sensor, the observation kernel will theoretically follow a Laplacian like shape, with an exponential decay in the amplitude of a measurement with an increasing distance between the source and sensor (?). However, when considering a source or sensor with a spatial extent, the peak of the observation kernel becomes smoothed and is more similar to a Gaussian shape (?). An experiment supporting this was performed and

the results are shown in Appendix B. Therefore, in this study the observation kernel, $m(\mathbf{r}-\mathbf{r}')$, that governs the sensor pick-up geometry is defined by the Gaussian

$$m\left(\mathbf{r}-\mathbf{r}'\right) = \exp\left(-\frac{(\mathbf{r}-\mathbf{r}')^\top(\mathbf{r}-\mathbf{r}')}{\sigma_m^2}\right),\tag{14}$$

where σ_m sets the kernel width. The superscript $\top$ denotes the transpose operator.

2.3. Derivation of Finite Dimensional State-Space Model

In order to implement standard estimation techniques, we use a decomposition of the field using a set of Gaussian basis functions. Decomposition allows a continuous field to be represented by a finite-dimensional state vector. This facilitates application of standard nonlinear state estimation methods such as the unscented Kalman filter. The field decomposition is described by

$$v_t\left(\mathbf{r}\right) \approx \boldsymbol{\phi}^\top\left(\mathbf{r}\right)\mathbf{x}_t,\tag{15}$$

where $\mathbf{x}_t$ is the state vector that scales the field basis functions $\boldsymbol{\phi}(\mathbf{r})$. The field basis functions are described by

$$\phi\left(\mathbf{r}-\mathbf{r}'\right) = \exp\left(-\frac{(\mathbf{r}-\mathbf{r}')^\top(\mathbf{r}-\mathbf{r}')}{\sigma_\phi^2}\right),\tag{16}$$

where σ_ϕ is the basis function width parameter. An example of a one-dimensional field decomposition is given in Fig. 2. The width and positioning of the basis functions can be determined by spectral analysis (explained in detail in the following section).

The connectivity kernel can also be decomposed as

$$w\left(\mathbf{r},\mathbf{r}'\right) = \boldsymbol{\psi}^\top\left(\mathbf{r},\mathbf{r}'\right)\boldsymbol{\theta},\tag{17}$$

where $\boldsymbol{\psi}(\mathbf{r},\mathbf{r}')$ is a vector of Gaussian basis functions and $\boldsymbol{\theta}$ is a vector of scaling parameters. A graphical demonstration of the connectivity kernel decomposition and the net shape is provided in Fig. 1. By exploiting the isotropy of the connectivity basis functions, they can be written as $\boldsymbol{\psi}(\mathbf{r}-\mathbf{r}')$. We will assume that we know the parametric form of the connectivity basis functions, where the scaling parameters $\boldsymbol{\theta}$ are unknown. Note, that although most examples presented in this paper use the Mexican hat connectivity kernel, the framework is also valid for other connectivity kernels, provided they are homogeneous and that they can be approximated by a set of isotropic basis functions parametrized by a coefficient vector $\boldsymbol{\theta}$. The mathematical formulation does not restrict the shape of the kernel, where anisotropic arbitrary shapes can be represented. Furthermore, the kernel basis functions need not be a Gaussian shape, but can be, for example, Laplacian to describe the "wizard hat" connectivity or B-spline functions to describe a kernel with a compact support.

Each connectivity basis function of the Mexican hat can, individually, be considered a layer in the Wilson and Cowan model. Rearranging the order of the spatial convolution in equation 12 and substituting in equations 15 and 17 to the right hand side we get the approximation of the dynamic field

$$
v_{t+1}(\mathbf{r}) \approx T_s \int_\Omega f(\boldsymbol{\phi}^\top(\mathbf{r}')\mathbf{x}_t)\boldsymbol{\psi}^\top(\mathbf{r}-\mathbf{r}')\,d\mathbf{r}'\boldsymbol{\theta}
$$
$$
+ \xi\boldsymbol{\phi}^\top(\mathbf{r})\mathbf{x}_t + \mathbf{e}_t(\mathbf{r}). \tag{18}
$$

Next we cross-multiply equation 18 by $\boldsymbol{\phi}(\mathbf{r})$ and integrate over the spatial domain, Ω, to get

$$
\int_\Omega \boldsymbol{\phi}(\mathbf{r})\,v_{t+1}(\mathbf{r})\,d\mathbf{r}
$$
$$
\approx T_s \int_\Omega \boldsymbol{\phi}(\mathbf{r}) \int_\Omega \boldsymbol{\psi}^\top(\mathbf{r}-\mathbf{r}')f(\boldsymbol{\phi}^\top(\mathbf{r}')\mathbf{x}_t)\,d\mathbf{r}'d\mathbf{r}\boldsymbol{\theta}
$$
$$
+ \xi\int_\Omega \boldsymbol{\phi}(\mathbf{r})\boldsymbol{\phi}^\top(\mathbf{r})d\mathbf{r}\mathbf{x}_t + \int_\Omega \boldsymbol{\phi}(\mathbf{r})e_t(\mathbf{r})\,d\mathbf{r}. \tag{19}
$$

The solution of this equation for $\mathbf{x}_t$ is equivalent to finding a Galerkin projection, where the field basis functions would be considered the projection functions and the states would be the amplitude coefficients. See Appendix C for more information. In the next step of forming the state-space model we substitute the field decomposition into the left hand side of equation 19 giving

$$
\int_\Omega \boldsymbol{\phi}(\mathbf{r})\boldsymbol{\phi}^\top(\mathbf{r})\mathbf{x}_{t+1}d\mathbf{r}
$$
$$
= T_s \int_\Omega \boldsymbol{\phi}(\mathbf{r}) \int_\Omega \boldsymbol{\psi}^\top(\mathbf{r}-\mathbf{r}')f(\boldsymbol{\phi}^\top(\mathbf{r}')\mathbf{x}_t)\,d\mathbf{r}'d\mathbf{r}\boldsymbol{\theta}
$$
$$
+ \xi\int_\Omega \boldsymbol{\phi}(\mathbf{r})\boldsymbol{\phi}^\top(\mathbf{r})d\mathbf{r}\mathbf{x}_t + \int_\Omega \boldsymbol{\phi}(\mathbf{r})e_t(\mathbf{r})\,d\mathbf{r}. \tag{20}
$$

Now defining the matrix

$$
\boldsymbol{\Gamma} \triangleq \int_\Omega \boldsymbol{\phi}(\mathbf{r})\boldsymbol{\phi}^\top(\mathbf{r})\,d\mathbf{r}, \tag{21}
$$

and substituting it into equation 20 and cross-multiplying by $\boldsymbol{\Gamma}^{-1}$ gives

$$
\mathbf{x}_{t+1} = T_s\boldsymbol{\Gamma}^{-1}\int_\Omega \boldsymbol{\phi}(\mathbf{r}) \int_\Omega \boldsymbol{\psi}^\top(\mathbf{r}-\mathbf{r}')f(\boldsymbol{\phi}^\top(\mathbf{r}')\mathbf{x}_t)d\mathbf{r}'d\mathbf{r}\boldsymbol{\theta}
$$
$$
+ \xi\mathbf{x}_t + \boldsymbol{\Gamma}^{-1}\int_\Omega \boldsymbol{\phi}(\mathbf{r})e_t(\mathbf{r})\,d\mathbf{r}. \tag{22}
$$

The analytic derivation of the inner product of n-dimensional Gaussians, which was used for calculating $\boldsymbol{\Gamma}$, is provided in Appendix D. Alternative field basis functions may be used to form the decomposition; if the basis is orthonormal, such as a Fourier basis, then Γ is simply the identity matrix. Equation 22 can be

simplified by exploiting the isotropy of the connectivity kernel basis functions where

$$\psi_i(\mathbf{r} - \mathbf{r}') = \psi_i(2\mathbf{c}_i + \mathbf{r}' - \mathbf{r}) \tag{23}$$

and $\mathbf{c}_i$ is the center of the ith basis function of the kernel decomposition. To make the simplification we define

$$[\mathbf{\Psi}(\mathbf{r}')]_{:i} \triangleq T_s \mathbf{\Gamma}^{-1} \int_\Omega \boldsymbol{\phi}(\mathbf{r}) \psi_i(2\mathbf{c}_i + \mathbf{r}' - \mathbf{r}) d\mathbf{r}, \tag{24}$$

where $[\mathbf{\Psi}(\mathbf{r}')]_{:i}$ denotes the ith column of $\mathbf{\Psi}(\mathbf{r}')$ which is a $n_x \times n_\theta$ matrix, where n_x is the number of basis functions (states) and n_θ is the number of connectivity kernel basis functions. This matrix is pre-defined analytically (see Appendix E for analytic convolution of two Gaussians), since it is constant with respect to the states and the unknown parameters. Substituting $\mathbf{\Psi}(\mathbf{r}')$ into equation 22 gives

$$\mathbf{x}_{t+1} = \int_\Omega \mathbf{\Psi}(\mathbf{r}') f(\boldsymbol{\phi}^\top(\mathbf{r}')\mathbf{x}_t) \, d\mathbf{r}' \boldsymbol{\theta}$$
$$+ \xi \mathbf{x}_t + \mathbf{\Gamma}^{-1} \int_\Omega \boldsymbol{\phi}(\mathbf{r}) e_t(\mathbf{r}) \, d\mathbf{r}. \tag{25}$$

This substitution swaps the convolution in equation 18, which is computationally demanding, with the convolution in equation 24, which is pre-defined analytically. This is of great importance as it lowers the computational demands of the estimation algorithm and provides a dramatic increase in estimation speed.

Now we define the state disturbance as

$$\mathbf{e}_t \triangleq \mathbf{\Gamma}^{-1} \int_\Omega \boldsymbol{\phi}(\mathbf{r}) e_t(\mathbf{r}) \, d\mathbf{r}, \tag{26}$$

which is a zero mean normally distributed white noise process with covariance (see Appendix F for the derivation)

$$\mathbf{\Sigma}_e = \mathbf{\Gamma}^{-1} \int_\Omega \int_\Omega \boldsymbol{\phi}(\mathbf{r}) \gamma(\mathbf{r} - \mathbf{r}') \boldsymbol{\phi}(\mathbf{r}')^\top \, d\mathbf{r}' d\mathbf{r} \mathbf{\Gamma}^{-\top}. \tag{27}$$

The observation equation of the reduced model is found by substituting equation 15 into equation 13 giving

$$\mathbf{y}_t = \int_\Omega m(\mathbf{r}_n - \mathbf{r}') \boldsymbol{\phi}^\top(\mathbf{r}') \mathbf{x}_t \, d\mathbf{r}' + \boldsymbol{\varepsilon}_t. \tag{28}$$

Note, the sensor index $\mathbf{r}_n$ has been omitted to delineate the reduced model observation equation from the full model. The observation equation is linear and can be written in the more compact form

$$\mathbf{y}_t = \mathbf{C} \mathbf{x}_t + \boldsymbol{\varepsilon}_t, \tag{29}$$

9

where each element of the observation matrix is

$$\mathbf{C}_{ij} \triangleq \int_\Omega m(\mathbf{r}_i - \mathbf{r}')\phi_j(\mathbf{r}')\,d\mathbf{r}'. \tag{30}$$

Now we have the final form of the state-space model where

$$\mathbf{x}_{t+1} = Q(\mathbf{x}_t) + \mathbf{e}_t \tag{31}$$

$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \boldsymbol{\varepsilon}_t \tag{32}$$

and

$$Q(\mathbf{x}_t) = \int_\Omega \boldsymbol{\Psi}(\mathbf{r}')f(\boldsymbol{\phi}^\top(\mathbf{r}')\mathbf{x}_t)\,d\mathbf{r}'\boldsymbol{\theta} + \xi\mathbf{x}_t. \tag{33}$$

2.4. Spectral Analysis

Spectral analysis has been used to identify both the number of sensors and the number of basis functions required to reconstruct the neural field from sampled observations (**??**). Based on a two-dimensional extension of Shannon's sampling theorem (**?**), the spatial bandwidth of the observed field can be used to provide a lower bound on both the number of sensors and the number of basis functions required to capture the dominant spatial spectral characteristics of the neural field.

Let the spectral representation of the post-synaptic membrane voltage field at time t be denoted by $V_t(\boldsymbol{\nu})$, where $\boldsymbol{\nu}$ is the spatial frequency. According to Shannon's sampling theorem, the field must be spatially band-limited for an accurate reconstruction using spatially discrete observations. However, an approximate reconstruction can be obtained if the field is approximately band-limited with

$$V_t(\boldsymbol{\nu}) \approx 0 \quad \forall \boldsymbol{\nu} > \boldsymbol{\nu}_c, \tag{34}$$

where $\boldsymbol{\nu}_c$ is a cutoff frequency (typically taken as the -3 dB point). Given such a band-limited field, the distance between adjacent sensors, Δ_y, must satisfy

$$\Delta_y \leq \frac{1}{2\rho_y\boldsymbol{\nu}_c}, \tag{35}$$

where $\rho_y \in \mathbb{R} \geq 1$ is an oversampling parameter. A derivation of the sampling theorem that is based on **?** and **?** is provided in Appendix G. The condition in equation 35 must be satisfied to avoid spatial spectral aliasing effects when reconstructing the hidden dynamic field, $v_t(\mathbf{r})$, using the sampled observations, $\mathbf{y}_t$.

In practice, it is difficult to estimate the bandwidth of the cortex using traditional electrophysiological measurements, possibly preventing placement of sensors in accordance to equation 35. However, we envisage it may be possible to estimate the spatial bandwidth using other modalities with higher spatial resolution

such as fMRI, NIRS or other optical imaging techniques (**?**). The expected bandwidth of the human neocortex is approximately 0.1-0.4 cycles/mm (**?**).

Spectral aliasing can be avoided by a proper choice of the spatial sampling distance given the sensors' spectral characteristics. The spatial extent of the sensors results in a spectral low-pass action, thus providing spatial anti-aliasing filtering. Such sensors limit the spatial bandwidth of the observation to a value denoted by $\boldsymbol{\nu}_{cy}$. Aliasing can then be avoided by positioning the sensors according to equation 35, replacing $\boldsymbol{\nu}_c$ with $\boldsymbol{\nu}_{cy}$.

Although such sensors avoid errors due to aliasing, they attenuate the high spatial frequency variations in the observations. Therefore, any procedure applied to estimate the original field or the underlying connectivity structure from these band-limited observations has the potential to underestimate the high spatial frequency components of both the field and the connectivity kernel. This motivates the use of sensors with wider bandwidths (narrower in space). This choice would require more sensors in order to satisfy Shannon's sampling theorem for a given spatial region. Therefore, a compromise should be found between the bandwidth of the sensor response, the accuracy of the estimation results, the number of sensors and the computational demands of the estimation procedure when designing experiments.

Similar considerations need to be made regarding the representation of the dynamic field, $v_t(\mathbf{r})$, using the basis function decomposition. Again, applying Shannon's sampling theorem, the maximum distance between basis functions to accurately reconstruct the neural field must satisfy

$$\Delta_\phi \leq \frac{1}{2\rho_\phi \boldsymbol{\nu}_{cy}} \tag{36}$$

where $\rho_\phi \in \mathbb{R} \geq 1$ is an oversampling parameter to determine the basis function separation.

The field basis function widths can also be inferred using spectral considerations (**??**). To demonstrate this, we begin with a two-dimensional Gaussian basis function located at the origin, defined as

$$\phi(\mathbf{r}) = \exp\left(-\frac{1}{\sigma_\phi^2} \mathbf{r}^\top \mathbf{r}\right). \tag{37}$$

The corresponding Fourier transform is

$$\boldsymbol{\Phi}(\boldsymbol{\nu}) = \left(\frac{1}{\pi \sigma_\nu^2}\right) \exp\left(-\frac{1}{\sigma_\nu^2} \boldsymbol{\nu}^\top \boldsymbol{\nu}\right), \tag{38}$$

where

$$\sigma_\nu^2 = \frac{1}{\pi^2 \sigma_\phi^2}. \tag{39}$$

To obtain a 3 dB attenuation at $\boldsymbol{\nu}_{cy}$, the basis function width, σ_ϕ^2, should be chosen such that

$$\sigma_\phi^2 = \frac{1}{\pi^2 \sigma_{\nu_{cy}}^2}, \tag{40}$$

where

$$\sigma_{\nu_{cy}}^2 = \frac{2\boldsymbol{\nu}_{cy}^\top \boldsymbol{\nu}_{cy}}{\ln 2}. \tag{41}$$

This ensures that the basis functions can represent the neural field with frequency content up to $\boldsymbol{\nu}_{cy}$. Derivations for equations 38 and 41 are given in Appendix H.

Alternatively, the basis function width can be set to provide a desired level of smoothness in the approximated field by reducing the spatial bandwidth. This allows for an approximate representation by a lower order model (fewer basis functions), thus reducing the computational demands of the estimation process. The spatial cutoff frequency of this smoothed field becomes a design choice. To calculate the cutoff frequency imposed by the basis function width, we rearrange equations 40 and 41 giving

$$\nu_{c\phi} = \frac{1}{\pi \sigma_\phi} \sqrt{\frac{\ln 2}{2}}. \tag{42}$$

Note that for a spatially homogeneous and isotropic field, the basis functions can be placed on a regular grid. Thus, the knowledge of the distance between basis functions can be used to determine the total number of basis functions required to represent a known spatial region.

2.5. State and Parameter Estimation

In this section, we describe the procedure for estimating the states, $\mathbf{x}_t$, the connectivity kernel parameters, $\boldsymbol{\theta}$, and the synaptic dynamics, ξ. The framework assumes that the parameters of the model (and brain) are stationary over the estimation period. This is a reasonable assumption when applying the framework to an evoked-potential experiment over a short time period in a controlled environment.

The estimation process is a two part iterative algorithm, consisting of a state estimation step followed by a parameter estimation step. Once initialized, at each iteration the sequence of estimated state vectors is used to update the parameter set. The resulting parameters are then used when estimating the new state vector sequence for the next iteration. The procedure stops when the parameter estimates converge. The algorithm is initialized using a bounded random state vector sequence, guaranteeing that the initial estimated parameter set forms a stable kernel.

The additive form of the unscented Rauch-Tung-Striebel smoother (URTSS) (**?**) was used for the state estimation. The URTSS incorporates an unscented Kalman filter (UKF) (**??**) in a forward step to estimate posterior states, $\hat{\mathbf{x}}_t^f$, followed by a backward step to compute the smoothed state estimates, $\hat{\mathbf{x}}_t^b$. The first and second order moments of the predicted state are captured by propagating the so-called sigma points

through the state equation. The sigma points, $\mathcal{X}_i$, are calculated using the unscented transform as follows:

$$\mathcal{X}_0 = \bar{\mathbf{x}} \tag{43}$$

$$\mathcal{X}_i = \bar{\mathbf{x}} + \left(\sqrt{(n_x + \lambda)\mathbf{P}_x} \right)_i, \quad i = 1, \ldots, n_x \tag{44}$$

$$\mathcal{X}_i = \bar{\mathbf{x}} - \left(\sqrt{(n_x + \lambda)\mathbf{P}_x} \right)_{i-n_x}, \quad i = n_x + 1, \ldots, 2n_x \tag{45}$$

where $\bar{\mathbf{x}}$ represents either $\hat{\mathbf{x}}_t^f$ or $\hat{\mathbf{x}}_t^b$, $\mathbf{P}_x$ is the corresponding covariance matrix for filtering or smoothing, $\left(\sqrt{(n_x + \lambda)\mathbf{P}_x} \right)_i$ is the ith column of the scaled matrix square root of $\mathbf{P}_x$, and n_x is the dimension of the state space. The total number of sigma points is $2n_x + 1$. The scaling parameter, λ, is defined as

$$\lambda = \alpha^2(n_x + \kappa) - n_x, \tag{46}$$

The positive constant α, that determines the spread of the sigma points around $\bar{\mathbf{x}}$, was set to 10^{-3}. It is typically set arbitrary small to minimise higher-order effects (?). The other scaling parameter, κ, was set to $3 - n_x$ in accordance to ?.

The sigma vectors are propagated through the system equations and weighted to form the predicted mean and covariance. The weights are calculated by

$$\mathbf{W}_0^{(m)} = \frac{\lambda}{n_x + \lambda} \tag{47}$$

$$\mathbf{W}_0^{(c)} = \frac{\lambda}{n_x + \lambda} + (1 - \alpha^2 + \beta) \tag{48}$$

$$\mathbf{W}_i^{(m)} = \mathbf{W}_i^{(c)} = \frac{1}{2(n_x + \lambda)} \quad i = 1, \ldots, 2n_x, \tag{49}$$

where the superscripts m and c denote mean and covariance and β incorporates prior knowledge of the distribution of the states, $\mathbf{x}$ (for a Gaussian disturbance, β should be set to 2 (?)). Since the observation equation is linear (equation 29), the standard Kalman Filter update equations are used to correct the predicted states. The state estimates from forward filtering are used to form a new set of sigma points for the smoother, as described above. The cross-covariance matrix of the states, $\mathbf{M}$, is also required for computing the smoother gain, as described in the summary of the steps in the URTSS algorithm given in Table 2.

Although the system is nonlinear, the parameters of the system are linear with respect to the state. This is exploited by our procedure where the parameter estimation uses a least squares (LS) method that minimizes the sum of the squared errors (of a predicted state update) with each new estimate of the state

vector sequence. To create the least squares parameter estimator, we first define the $n_x \times n_\theta$ matrix

$$\mathbf{q}(\mathbf{x}_t) = \int_\Omega \boldsymbol{\Psi}(\mathbf{r}')f(\boldsymbol{\phi}^\top(\mathbf{r}')\mathbf{x}_t)\,\mathrm{d}\mathbf{r}'. \tag{50}$$

Now given a state estimate sequence from the initialization or an iteration of the URTSS, we can write

$$\hat{\mathbf{x}}_1^b = \mathbf{q}(\hat{\mathbf{x}}_0^b)\boldsymbol{\theta} + \xi\hat{\mathbf{x}}_0^b + \mathbf{e}_0$$

$$\hat{\mathbf{x}}_2^b = \mathbf{q}(\hat{\mathbf{x}}_1^b)\boldsymbol{\theta} + \xi\hat{\mathbf{x}}_1^b + \mathbf{e}_1$$

$$\vdots$$

$$\hat{\mathbf{x}}_T^b = \mathbf{q}(\hat{\mathbf{x}}_{T-1}^b)\boldsymbol{\theta} + \xi\hat{\mathbf{x}}_{T-1}^b + \mathbf{e}_{T-1}.$$

This can be written in the compact form

$$\mathbf{Z} = \mathbf{X}\mathcal{W} + \mathbf{e}, \tag{51}$$

where

$$\mathbf{Z} = \begin{bmatrix} \hat{\mathbf{x}}_1^b \\ \hat{\mathbf{x}}_2^b \\ \vdots \\ \hat{\mathbf{x}}_T^b \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} \mathbf{q}(\hat{\mathbf{x}}_0^b) & \hat{\mathbf{x}}_0^b \\ \mathbf{q}(\hat{\mathbf{x}}_1^b) & \hat{\mathbf{x}}_1^b \\ \vdots & \vdots \\ \mathbf{q}(\hat{\mathbf{x}}_{T-1}^b) & \hat{\mathbf{x}}_{T-1}^b \end{bmatrix}$$

and

$$\mathcal{W} = \begin{bmatrix} \boldsymbol{\theta} \\ \xi \end{bmatrix}, \quad \mathbf{e} = \begin{bmatrix} \mathbf{e}_0 \\ \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_{T-1} \end{bmatrix}.$$

Following this, the LS parameter estimates (**?**), $\hat{\mathcal{W}}$, are

$$\hat{\mathcal{W}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Z}. \tag{52}$$

3. Results

We now provide demonstrations of the estimation procedure using the new methods that are described above. Data was generated using the neural field model, described in Section 2.2, to demonstrate the state and parameter estimation framework. All parameters for the model are given in Table 3, unless otherwise explicitly stated for particular experiments. Note that the parameters for the spatial and temporal aspects of the model can be considered arbitrary when demonstrating the estimation methodology. This is because the temporal aspects scale with the sampling rate and the length of the times series, and the spatial aspects scale with the size of the neural field and the spatial step size used for numerical simulation. When generating

the simulated data, we assumed free boundary conditions for the neural field. An example of the simulated field at a single point in time can be seen in Fig. 3(A). Fig. 3(B) shows an example of data from the modeled sensors and Fig. 3(C) shows the power spectral density (PSD) of the simulated time-series.

The sensors were spaced on a 14×14 square grid and the sensor width was set to $\sigma_m^2 = 0.81$. This provides a width of 1.5 mm at half the maximum amplitude of the observation kernel. The distance between sensors was set to $\Delta_y = 1.5$ mm. This arrangement modeled the recording as having some cross talk (from volume conduction), which is typical for neurophysiological recordings.

3.1. Spatial Frequency Analysis

The spectral analysis was used to confirm that the spacing and width of the sensors was adequate to capture the significant dynamics of the field. Fig. 4(A) shows the spatial frequency of the neural field and Fig. 4(B) shows the spatial frequency of the reconstructed neural field from the basis function decomposition. The cutoff frequency (-3 dB point) of the neural field is $\boldsymbol{\nu}_c \approx 0.24$ cycles/mm. From equation 35, this yielded a maximum sensor separation of 2.08 mm. This confirms that the separation of $\Delta_y = 1.5$ mm was sufficient to mitigate problems associated with spatial aliasing of the dominant dynamics.

The spatial frequency of the observed field is shown in Fig. 5(A). The cutoff frequency is $\boldsymbol{\nu}_{cy} \approx 0.2$ cycles/mm (-3 dB point). The observation cutoff frequency is lower than the field cutoff due to the low-pass filtering effect of the sensors. The spatial frequency response of the sensors can be seen in Fig. 5(B).

As described in Section 2.4, the basis function configuration can be systematically chosen to account for the full spatial bandwidth from the observations, or alternatively, the configuration can be chosen to represent the neural field up to a specified bandwidth to limit the computational complexity of the estimation procedure. In the examples provided in this paper, the later option was chosen, where the neural field cutoff frequency was further reduced to $\boldsymbol{\nu}_{c\phi} = 0.12$ cycles/mm. The band-limiting effect on the field reconstruction can be seen by comparing Fig. 4(A) to Fig. 4(B). This effect is due to the spatial frequency response of the basis functions, which is shown in Fig. 5(C). Due to the slow roll-off in the frequency response of Gaussian basis functions an oversampling parameter of $\rho_\phi = 1.67$ was chosen. Given the spectral response of the basis functions, the over-sampling parameter, and the spatial domain an equally spaced grid of 9×9 basis functions was used to decompose the field, which resulted in 81 states.

3.2. State and Parameter Estimates

Several experiments were performed to demonstrate the performance of the state and parameter estimation algorithm. The first experiment evaluated the estimation performance using a Monte Carlo approach for a single set of parameters. In this experiment we demonstrate convergence of the algorithm, show the distribution of estimated parameters and the accuracy of the state estimates for a single set of parameters.

The second experiment evaluated the performance with a selection of different connectivity kernel parameters. In addition, results are shown where the support of the connectivity kernels differs between actual data and the estimator. The third experiment shows how the estimation accuracy depends on the parameters of the sigmoidal nonlinearity. The final experiment shows the effect of the width of the observation kernel on the estimation performance.

3.2.1. Experiment I: Monte Carlo Simulation with Fixed Parameters

A Monte Carlo approach was employed where 150 realizations of the data were generated, such that distributions of the parameters estimates could be formed. Each realization consisted of 500 ms of data (sampled at 1 kHz) and the estimation was applied to the final 400 ms, allowing the model's initial transients to die out. The initial state and parameters were unknown to the estimator. Ten iterations of the estimation algorithm were used to ensure that the parameter estimates converged. Fig. 6 shows the rate of convergence, confirming that the algorithm converged to steady parameter values, with the changes in the parameter errors (between consecutive iterations) falling to less than 10^{-4} after 6 iterations.

Histograms of the parameter estimates obtained over the 150 realizations are shown in Fig. 7. In each of the subplots of the figure the true parameter value is shown by the solid line and the mean parameter estimate is shown by the dashed line. The actual parameter values, the mean parameter estimates, standard deviations, and biases for this parameter combination are given as the first entry in Table 4, where other entries in the table show results from Experiment II. The biases show the percentage differences between the actual values and the mean parameter estimates. For this parameter combination the mean estimates are in good agreement (within 1 standard deviation) with the actual values of the connectivity kernel parameters. However, the time constant parameter, ξ, is outside the estimated distribution (with bias of 2.67%), with a small variance in parameter estimates. We attribute this small bias to the loss of the high frequency components in the smoother estimated field with respect to the actual field. These results indicate that choosing to limit the frequency content of the reconstructed neural field has only minor effects on the accuracy of the estimated kernel while allowing for significant gains in the computational demand.

The accuracy of the state estimates was evaluated by comparing the field reconstruction to the true field using the mean (over iterations) of the root mean square error (MRMSE) over space. The MRMSE of the field estimates over time for 150 realizations is shown in Fig. 8 with the 95 % confidence interval. At each time point, the error is consistently around 0.5 mV, which is approximately 10.73 % of the field voltage range ([-2.20, 2.46] mV). The error in the reconstructed field is again mostly due to the loss of high spatial frequency information from the basis function decomposition as illustrated in Fig. 9. The figure shows a snapshot at a single time instant of the true neural field, the estimated field, and the corresponding error, which is the difference between the true and estimated neural fields. It can be seen in Fig. 9(B) that the estimated field contains the basic structure of the true field, shown in Fig. 9(A), but is missing the

higher-frequency detail, which can be see in the error shown in Fig. 9(C).

3.2.2. Experiment II: Results for Different Connectivity Kernel Parameters

To study the robustness of the estimation algorithm, the performance was evaluated for various connectivity kernel parameters. The reconstructed connectivity kernels from the estimated parameters are shown in Fig. 10. For Fig. 10(A-C) Mexican hat connectivity kernels were used. For each subset of parameters, the mean of the reconstructed kernel over 50 realizations is shown as a dotted line and the 95 % confidence interval is indicated by the grey shaded region. In each case, the actual kernel was within the 95 % confidence interval of the kernel estimate. The statistics for the kernel parameter estimates for each of the parameter sets are shown in Table 4.

For each of the Mexican hat connectivities, the support of the kernel was assumed to be known. However, in Fig. 10(D-F) the kernel support in the estimator differed from the actual data. In these cases the data sets were generated using a single positive (excitatory) Gaussian kernel, as seen by the solid line in each of the subplots of the figure. The estimation algorithm was constructed allowing for the Mexican hat connectivity of the previous examples. Each plot shows that the actual kernel lies within the confidence intervals of the reconstructed estimates for the majority of the kernel space, until the amplitude is relatively low.

Table 5 shows the statistics of the parameter estimates. Although the net reconstructed kernel estimates are in good accordance with the actual kernels, the individual kernel basis function weights are not in agreement with the actual parameters. This result demonstrates that the interpretation of the results should be based on the net shape of the estimated connectivity kernel and not on the estimated parameter values individually. These results indicate that the solution to the system may not be unique in non-ideal, practical situations. Nevertheless, the net shape of the connectivity kernel can be determined reasonably accurately.

3.2.3. Experiment III: Results for Varying the Parameters of the Sigmoidal Activation Function

Fig. 11 demonstrates the effect of the sigmoid parameters on the estimation accuracy. The RMSE of the reconstructed connectivity kernel was calculated for various parameters of the sigmoid by varying the slope, ς, and the threshold, v_0. The other model parameters remained fixed and were set according to the Table 3.

From Fig. 11 it can be seen that the estimation accuracy depends on the parameters of the sigmoid function. This is expected since that by altering the parameters we are changing the nonlinearity and the saturation region of the sigmoid. For example, when the sigmoid slope is very gentle, most of the neural activity falls in the central, nearly linear region of the sigmoid, thus having minor effects on the estimation procedure. On the other hand, sharp slopes result in more of the activity falling in the saturation region of the sigmoid with adverse effects on the estimation. Similar arguments can be made for variations to the sigmoid threshold.

3.2.4. Experiment IV: Results for Varying the Sensor Kernel Width

The accuracy of the estimation algorithm was tested for different observation kernels by varying the width parameter, σ_m. Fig. 12 shows the RMSE of the connectivity kernel estimates for the various observation kernel widths. As the theory presented earlier in Section 2.4 states, the observation kernel acts as a spatial filter. When the observation kernel is too narrow, high frequency components in the neural field are present but under sampled. These high frequency components result in aliasing errors in the estimated field, thus affecting the kernel estimation accuracy. On the other hand, for the very wide observation kernels considered, important neural dynamics are not represented in the estimated field due to the filtering out of the important frequency components by the sensors, resulting in poor accuracy in the kernel estimate. As indicated in Fig. 12, the best results are obtained when the sensor width is chosen to minimize both sources of errors.

4. Discussion

4.1. Implications for Experimental Design

As seen in Fig. 7 and Fig. 10, the results show good estimation accuracy for the connectivity kernels, where the true kernels lie within the 95 % confidence intervals of the kernel estimates. The estimates of the connectivity kernels show relatively broad distributions over the different realizations, which can be expected considering the system's input is modeled as a random disturbance and the observations are corrupted by noise. To overcome the effect of noise, the use of a controlled input stimulus should be considered when applying the estimation algorithm to electrophysiological recordings, where given knowledge of the input and the appropriate model, it is expected that the variance of the parameter estimates would decrease. Even when using a known input, our results imply that we should only expect to learn distributions of parameters, rather than uncover a 'true' parameter set. By designing an input signal carefully and using an evoked-potential paradigm where stimuli are repeated many times, recordings could be averaged over a sufficient number of trials (realizations), such that a distribution of the parameter estimates and resulting connectivity kernels can be formed.

The frequency analysis specifies the minimum sensor arrangement that is required to capture the dominant cortical dynamics. As pointed out in numerous publications by Freeman and colleagues (see ? for example), a detailed description of the spatial properties of responses to stimuli is essential in understanding cortical dynamics. Importantly, the spatial frequency analysis tools presented in this paper can be used to design high density intracranial electrophysiological recording systems to avoid spatial aliasing, ensuring the essential spatial dynamics can be observed. In addition, the frequency analysis highlights the relationship between the width of the sensors and the information in the data. This analysis is becoming increasingly relevant as recording systems are becoming more sophisticated with higher density electrodes (?).

The work of **?** provides direct evidence that the spatial frequency of the neural field is in the range of 0.1-0.4 cycles/mm and spacing required to prevent aliasing is approximately 1.25 mm. Micro-electrodes systems that satisfy this requirement are now currently in use for brain computer interface and epilepsy monitoring applications (**?**), which suggests that the framework can be applied to human data with existing technology.

4.2. Computational Complexity

The closest competitor to our 'grey-box' integro-difference equation approach is perhaps the multivariate auto-regressive (MVAR) model, which is a 'black-box' approach. Examples of MVAR models for estimating functional connectivity include Granger causality (**?**), the direct transfer function (**?**), and partial directed coherence (**?**). A distinct advantage of the integro-difference equation based technique proposed in this current study is that the number of parameters required to model the system is considerably less than a typical MVAR model. Furthermore, by using the neural field equations as a parametric form of a model, the estimated quantities have physical meaning and, therefore, can be clinically relevant. An advantage of the MVAR-based methods is that they allow for inhomogeneous connectivity, which is present in the longer range cortico-cortical fibres.

Previous work by **?** demonstrated for the first time the applicability of the unscented Kalman filter (UKF) for state estimation of discrete (in space and time) nonlinear spatiotemporal systems. Within their framework, each discrete spatial location formed an element of the state vector to be estimated with the UKF. A limitation of this framework was that an increase in spatial resolution and/or the physical size of the neural field resulted in an increased state dimension. This, in turn, leads to increases in uncertainty in the state estimates. In addition, directly linking the complexity of the underlying system to the spatial resolution of the observation process implies that the system's dynamics become increasingly more complex the closer we look, which may not be an accurate modeling assumption.

The current study builds on Schiff and Sauer's approach by estimating a continuous field whose complexity is independent of the resolution of the observation process and spatial resolution of the model. Instead, the dimensionality is governed by the expected complexity of the underlying field, which we measure using the frequency analysis presented in the Methods section. These methods provide a rigorous framework for determining the dimensionality of the system using inherent spatial properties of the field. It should be noted that Schiff and Sauer suggest the possibility of using a Galerkin projection, or similar decomposition that is used in this paper, to perform efficient estimation as a direction for future work. Indeed, we have shown that this enables estimation of a larger field with a lower dimensional state vector than would be otherwise possible.

The isotropy of the connectivity kernel basis functions means that the mixing effect of the kernel on the firing rates can be written as a convolution of the kernel with the field basis functions. This transformation

greatly simplifies the computational complexity of the estimator, which is important for the estimator to run in a real-time scenario.

4.3. Patient-Specific Modeling

Generating data-driven neural field models has the potential to have significant impact on several areas of neuroscientific research and clinical neurology. Specifically, these subject-specific neural field models will provide new avenues for the application of engineering techniques to neural systems. For example, state tracking could be used in brain-computer interfaces or epileptic seizure prediction, where specific regions of the state-space or certain parameter combinations may indicate intent of movement or imminent seizures, respectively. In addition, a patient-specific state-space model will allow application of electrical stimulation to robustly prevent the neurodynamics entering pathological regions of state-space, thereby possibly preventing seizure onset or abating seizure activity.

Estimation of the connectivity structure or the synaptic time constant may also have significant impact on the treatment of disease. For example, recent theoretical work has demonstrated that seemingly similar electrographic seizures in absence epilepsy may arise from either abnormal excitatory or inhibitory mechanisms (**?**). These very different mechanisms would require very different medications for successful treatment. However, the underlying mechanisms would be hidden from the clinician in the normal clinical setting. Successful parameter estimation has the potential to reveal these hidden mechanisms allowing for improved treatment strategies.

The development of the estimation framework also has implications for theoretical neuroscience, where the ability to estimate subject-specific properties will allow testing of hypotheses that have been generated in computational modeling studies. This has the potential to validate neural field models and, in doing so, provide new insights that may lead to a greater understanding of mesoscopic and macroscopic neurodynamics. For example, using the proposed model-based framework, we can reconstruct a neural field from data and estimate parameters to check if they correspond to theoretically derived values where certain solutions may exist. In addition, the parameters and the state-space can be explored during cognitive tasks and states of arousal such as sleep and anaesthesia.

4.4. Extensions to the Framework

In this study, an assumption was made that the chosen kernel basis functions can exactly represent the actual connectivity kernel. In future work, we plan to relax this assumption and show how the support of the connectivity kernel can also be inferred from data. In addition, we have assumed that the neural field is isotropic and homogeneous. It is generally accepted that the intracortical connectivity structure can be accurately modeled as being isotropic. However, the assumption of homogeneity is quite strong as the size of the neural field becomes larger. Future work should focus on developing estimation algorithms where

assumptions of homogeneity can be relaxed. Other natural extensions of this framework are the inclusion of a post-synaptic response kernel with a finite rise and decay time, and a finite distant-dependant action potential propagation velocity into the parametric neural field equations.

The assumption was made that the parameters of the model are stationary over the estimation period. However, it is well known that factors that influence excitability are non-stationary over longer time periods (see **?**). Therefore, the estimation framework must be altered if it is to be applied to data over extended periods of time. A possible way to achieve this is to use a more sophisticated model that accounts for longer-term time dependancies (higher temporal order). Alternatively, the parameters of the model could be considered time-varying, where changes in excitability would result in different connectivity kernel parameters. Furthermore, the model may be considered heteroscedastic, where the disturbance is also time-varying. This is common in spatio-temporal econometric studies, where methods have been developed to test for heteroscedasticity (**?**). Following this, the disturbance properties of the system will also need to be estimated. A possible approach allowing for time-varying parameters is to perform a window-based or fixed-lag estimation, which would require the assumption that the parameters are quasi-stationary over the duration of the window.

The estimation framework that we have presented is inherently a batch algorithm designed to operate after data has been collected. A further extension of the framework is to allow for the algorithm to learn online. A filter that is designed to be implemented in an online setting can be used in place of the smoother along with an online parameter estimator to update the parameters after a new observation arrived. A real-time implementation can be achieved using the UKF (instead of the URTSS) and an online recursive LS estimator **?**.

5. Conclusion

This paper has presented a framework for creating data-driven neural field models by developing a method for estimating intracortical connectivity and synaptic dynamics from electrophysiological data. In doing so, we have highlighted important aspects of the experiment design process, demonstrating the relationship between sensor bandwidth and the complexity of the field that can be inferred. We have derived a finite-order state-space representation of the continuum neural field equations that allows inference to be performed efficiently. Importantly, the mathematical formulation that we have presented scales well with the number of sensors. The major contribution of this framework is to formally bridge a gap between experimental and theoretical neuroscience in order to gain a better understanding of the mesoscopic dynamics of the brain. A validation of the estimation framework on real data is now required.

6. Acknowledgments

Tables

Symbol	Quantity	Units
Domain and indices		
Ω	spatial domain	n.a.
$\mathbf{r}$	spatial location	[mm, mm]
t	time	s
Model		
y_t	observation	mV
$v(\mathbf{r}, t)$	mean membrane potential	mV
$g(\mathbf{r}, t)$	weighted firing rate function	spikes.s^{-1}
$f(\mathbf{r}, t)$	firing rate function	spikes.s^{-1}
$e(\mathbf{r}, t)$	field disturbance, with covariance function $\gamma(\mathbf{r} - \mathbf{r}')$	mV
$\boldsymbol{\varepsilon}_t$	observation noise, with covariance matrix Σ_ε	mV
$m(\mathbf{r}_n, \mathbf{r}')$	observation kernel where n is sensor index $n = 1, .., n_y$ (width σ_m)	n.a.
$w(\mathbf{r}, \mathbf{r}')$	spatial connectivity kernel	n.a.
$h(t)$	post-synaptic response kernel	mV
ζ, ξ	inverse synaptic time constant & time constant parameter	s^{-1}, n.a.
$\eta(t)$	Heaviside step function	n.a.
$\delta(t)$	Dirac-delta function	n.a.
D	temporal differential operator	n.a.
Reduced Model		
$\boldsymbol{\phi}(\mathbf{r})$	vector of n_x Gaussian basis functions	n.a.
$\mathbf{x}_t$	state vector at time t	mV
$\boldsymbol{\psi}(\mathbf{r} - \mathbf{r}')$	vector of n_θ connectivity kernel basis functions	n.a.
$\boldsymbol{\Psi}(\mathbf{r}')$	decomposition connectivity matrix	n.a.
$\boldsymbol{\Gamma}$	inner product of field basis functions	n.a.
$Q()$	state transition function	n.a.
$\mathbf{e}_t$	state disturbance vector, with covariance Σ_e	mV
$\mathbf{C}$	observation matrix	n.a.
Frequency Analysis		
$\boldsymbol{\nu}, \boldsymbol{\nu}_c$	spatial frequency and spatial cutoff frequency	cycles/mm
σ_ν^2	variance of Fourier transformed Gaussian basis function	(cycles/mm)^2
Estimation		
$\mathcal{X}$	matrix of sigma vectors	mV
W	matrix of sigma vector weights (scaling parameter λ)	n.a.
$\hat{\mathbf{x}}_t^{f-}, \hat{\mathbf{x}}_t^f$	forward prior and posterior state estimates	mV
$\hat{\mathbf{x}}_t^{b-}, \hat{\mathbf{x}}_t^b$	backward prior and posterior state estimates	mV
P_t^{f-}, P_t^f	forward prior and posterior covariance matrices	mV^2
P_t^{b-}, P_t^b	backward prior and posterior covariance matrices	mV^2
M_t	cross-covariance matrix	mV^2
$\mathcal{K}_t$	filter gain	n.a.
S_t	smoother gain	n.a.
$q()$	nonlinear state function (used in LS estimator)	n.a.
$\hat{\mathcal{W}}$	vector of parameter estimates (used in LS estimator)	n.a.

Table 1: **Notation.** The first column of the table gives the symbols for all the notation used in the paper. The second column provides a brief description of the quantity, and the third column provides the SI units where relevant.

1. Forward initialization
$$\hat{\mathbf{x}}_0, \mathbf{P}_0$$

2. Forward iteration: for $t \in \{0, \cdots, T\}$, calculate the sigma points $\mathcal{X}_{i,t}^f$ using equations 43-46 and propagate through equation 33

$$\mathcal{X}_{i,t+1}^{f-} = Q(\mathcal{X}_{i,t}^f) \quad i = 0, \dots, 2n_x$$

Calculate the predicted state and the predicted covariance matrix

$$\hat{\mathbf{x}}_{t+1}^{f-} = \sum_{i=0}^{2n_x} W_i^{(m)} \mathcal{X}_{i,t+1}^{f-}$$

$$\mathbf{P}_{t+1}^{f-} = \sum_{i=0}^{2n_x} W_i^{(c)} (\mathcal{X}_{i,t+1}^{f-} - \hat{\mathbf{x}}_{t+1}^{f-})(\mathcal{X}_{i,t+1}^{f-} - \hat{\mathbf{x}}_{t+1}^{f-})^\top + \mathbf{\Sigma}_e$$

Compute the filter gain, the filtered state and the filtered covariance matrix using the standard Kalman filter update equations

$$\mathcal{K}_{t+1} = \mathbf{P}_{t+1}^{f-} \mathbf{C}^\top (\mathbf{C}\mathbf{P}_{t+1}^{f-}\mathbf{C}^\top + \mathbf{\Sigma}_\varepsilon)^{-1}$$

$$\hat{\mathbf{x}}_{t+1}^f = \hat{\mathbf{x}}_{t+1}^{f-} + \mathcal{K}_{t+1}(\mathbf{y}_{t+1} - \mathbf{C}\hat{\mathbf{x}}_{t+1}^{f-})$$

$$\mathbf{P}_{t+1}^f = (\mathbf{I} - \mathcal{K}_{t+1}\mathbf{C})\mathbf{P}_{t+1}^{f-}$$

3. Backward initialization

$$\mathbf{P}_T^b = \mathbf{P}_T^f, \quad \hat{\mathbf{x}}_T^b = \hat{\mathbf{x}}_T^f$$

4. Backward iteration: for $t \in \{T-1, \cdots, 0\}$ calculate the sigma points $\mathcal{X}_{i,t}^b$ and propagate them through equation 33

$$\mathcal{X}_{i,t+1}^{b-} = Q(\mathcal{X}_{i,t}^b) \quad i = 0, \dots, 2n_x$$

Calculate the predicted state and the predicted covariance matrix

$$\hat{\mathbf{x}}_{t+1}^{b-} = \sum_{i=0}^{2n_x} W_i^{(m)} \mathcal{X}_{i,t+1}^{b-}$$

$$\mathbf{P}_{t+1}^{b-} = \sum_{i=0}^{2n_x} W_i^{(c)} (\mathcal{X}_{i,t+1}^{b-} - \hat{\mathbf{x}}_{t+1}^{b-})(\mathcal{X}_{i,t+1}^{b-} - \hat{\mathbf{x}}_{t+1}^{b-})^\top + \mathbf{\Sigma}_e$$

$$\mathbf{M}_{t+1} = \sum_{i=0}^{2n_x} W_i^{(c)} (\mathcal{X}_{i,t}^b - \hat{\mathbf{x}}_t^f)(\mathcal{X}_{i,t+1}^{b-} - \hat{\mathbf{x}}_{t+1}^{b-})^\top$$

Compute the smoother gain, the smoothed state and the smoothed covariance matrix

$$\mathbf{S}_t = \mathbf{M}_{t+1} \left[\mathbf{P}_{t+1}^{b-}\right]^{-1}$$

$$\hat{\mathbf{x}}_t^b = \hat{\mathbf{x}}_t^f + \mathbf{S}_t \left[\hat{\mathbf{x}}_{t+1}^b - \hat{\mathbf{x}}_{t+1}^{b-}\right]$$

$$\mathbf{P}_t^b = \mathbf{P}_t^f + \mathbf{S}_t \left[\mathbf{P}_{t+1}^b - \mathbf{P}_{t+1}^{b-}\right] \mathbf{S}_t^\top$$

Table 2: **Algorithm for the Unscented RTS Smoother.** This table shows the steps in the unscented Rauch-Tung-Striebel smoother algorithm. The steps are iterated 10 times for our state estimation procedure. The least squares algorithm is run after each iteration to update the parameter estimates.

Symbol	Quantity	Value	Units
	Domain and indices		
Δ	spatial discretisation step	0.5	mm
T_s	time step	0.001	s
T	number of time steps used in estimation	400	n.a.
	Model		
τ	synaptic time constant	0.01	s^{-1}
ς	slope of sigmoidal activation function	0.56 (?)	mV^{-1}
v_0	firing threshold	1.8 (?)	mV
$\boldsymbol{\theta}$	vector of connectivity kernel parameters	$\begin{bmatrix} 100 & -80 & 5 \end{bmatrix}^\top$	$\mathrm{mV.spike}^{-1}$
$\sigma_{\psi_1}, \sigma_{\psi_2}, \sigma_{\psi_3}$	connectivity kernel width parameters	1.8, 2.4, 6	mm
n_y	number of sensors	196	n.a.
Δ_y	distance between adjacent sensors	1.5	mm
σ_m	observation kernel width parameter	0.9	mm
Σ_ε	observation noise variance	$0.1 \times I_{n_y}$	mm^2
σ_γ	disturbance spatial covariance width	1.3	mm
σ_d	disturbance variance	0.1	mV
	Reduced Model		
n_x	number of basis functions	81	n.a.
Δ_ϕ	distance between field basis functions	2.5	mm
σ_ϕ	width of field basis functions in spatial domain	1.58	mm^2
	Frequency Analysis		
ρ_y	sensor over-sampling parameter	1	n.a.
ρ_ϕ	field basis function over-sampling parameter	1.67	n.a.
	Estimation		
α	determines the spread of sigma points	0.001	n.a.
β	prior knowledge of the sigma points distribution	2	n.a.
κ	scaling parameter $(3 - n_x)$	-78	n.a.
λ	scaling parameter	-80.99	n.a.

Table 3: **Parameters.** Parameters for the neural field model, the reduced model, the frequency analysis and the estimation procedure.

Experiment	Parameter	Actual	Mean Est.	Std. Dev.	Bias (%)
I and II-A	θ_0	100	101.75	21.30	1.75
	θ_1	-80	-81.00	14.82	1.25
	θ_2	5	4.76	0.65	4.8
	ξ	0.9	0.924	0.003	2.67
II-B	θ_0	90	103.86	23.14	15.4
	θ_1	-80	-89.05	16.57	11.31
	θ_2	5	5.23	0.74	4.6
	ξ	0.9	0.923	0.003	2.56
II-C	θ_0	110	98.90	16.48	10.1
	θ_1	-80	-73.13	11.72	8.58
	θ_2	5	4.37	0.54	12.6
	ξ	0.9	0.925	0.003	2.78

Table 4: **Results from varying the Mexican hat connectivity kernel parameters**

Experiment	Parameter	Actual	Mean Est.	Std. Dev.	Bias (%)
II-D	θ_0	15	35.68	16.72	137.87
	θ_1	0	-18.16	11.77	-
	θ_2	0	0.81	0.53	-
	ξ	0.9	0.923	0.002	2.56
II-E	θ_0	30	40.43	11.47	34.77
	θ_1	0	-12.01	7.8	-
	θ_2	0	0.35	0.36	-
	ξ	0.9	0.924	0.003	2.67
II-F	θ_0	50	61.46	11.51	22.92
	θ_1	0	-15.03	7.56	-
	θ_2	0	0.27	0.29	-
	ξ	0.9	0.924	0.003	2.67

Table 5: **Results from varying the Gaussian connectivity kernel parameter**

Figure Legends

Figure 1. Mexican-hat connectivity kernel. The kernel is a composite of three Gaussian basis functions (dashed lines), representing short-range excitation, mid-range inhibition and long-range excitation. The kernel is rotationally symmetric (isotropic) about zero, hence a cross-section captures the important aspect of the connectivity kernel's shape. The kernel decays asymptotically to zero.

Figure 2. Example of a one-dimensional field decomposition. A continuous, one-dimensional field decomposed by a finite number of basis functions scaled by the state vector. The dashed line depicts the field, the black solid lines show the Gaussian basis functions and the vertical lines show the position and amplitude of the states.

Figure 3. Example of the spatiotemporal properties of the model and synthetically generated data. (A) The neural field at a single time instant. **(B)** Data generated by the full neural field model at the first 5 sensors. **(C)** Mean power spectral density of the data generated by the observations.

Figure 4. Spatial frequency analysis: the neural field. **(A)** The average (over time) power in dB of the spatial frequency of the neural field. **(B)** The average (over time) power in dB of the spatial frequency of the reconstructed neural field from the basis function decomposition. The reconstructed field shows a narrower bandwidth as a result of the band limiting properties of the sensors and basis functions.

Figure 5. Spatial frequency analysis: observations and basis functions. The dashed line shows the cutoff frequency (-3 dB point). **(A)** The average (over time) power in dB of the spatial frequency of the observations. **(B)** The spatial frequency response of a one-dimensional model sensor. The symmetry of the sensors and field basis functions allow for the one-dimensional representation of the spectral characteristics. **(C)** The spatial frequency response of a one-dimensional neural field basis function, illustrating a narrower band width than the sensors.

Figure 6. Convergence of parameters. The mean of the absolute error between true and estimated parameters averaged over 150 realizations. All parameters converge after 6 iterations, where the mean difference between errors in consecutive iterations falls below 10^{-4}.

Figure 7. Histograms of the parameter estimates over 150 realizations. In each of the subplots the solid lines show the actual parameter values and the dotted lines show the means of the estimated parameters. **(A)** The histogram of the central excitatory connectivity kernel basis function parameter estimates, θ_0. **(B)** The histogram of the surround inhibition connectivity kernel basis function parameter estimates, θ_1. **(C)** The histogram of the longer range excitatory connectivity kernel basis function parameter estimates, θ_2. **(D)** The histogram of the ξ estimates where $\xi = 1 - \zeta T_s$ and ζ is the reciprocal of the synaptic

time constant.

Figure 8. Error in the field reconstruction. The mean (over iterations) of average (over space) RMSE of the estimated field over 150 realizations (solid line) and the 95 % confidence interval (grey region).

Figure 9. Examples of the actual and estimated neural fields and the error. (**A**) The actual field at a single time instant from the neural field model that was used to generate the data. (**B**) The reconstructed field of the reduced model, showing the effect of the basis function decomposition on the frequency content of the estimated field. (**C**) The error between the true and reconstructed fields.

Figure 10. Cross-section of connectivity kernel estimates. In each of the subplots the actual kernels are shown by the solid lines, the mean kernel estimates (over 50 realizations) are shown by the dotted lines and the 95 % confidence intervals are shown by the shaded grey regions.

Figure 11. RMSE of connectivity kernel estimates with various sigmoid parameters. Log of the RMSE of the connectivity kernel for different values of firing threshold, v_0, and the slope of the sigmoid, ς. Mean kernel estimates are obtained over 50 realizations in each case. White spaces indicate cases where the kernel can not be estimated due to numerical problems.

Figure 12. RMSE of connectivity kernel estimates with various observation kernel widths. RMSE of the connectivity kernel for different values of observation width, σ_m. Mean kernel estimates are obtained over 50 realizations in each case.

Figure 13. Observation kernel experimental setup. Photograph through a microscope of the Utah array recording electrode and the twisted bipolar current source.

Figure 14. Observation kernel experimental results. Amplitude of the envelope of the signal recorded from the electrode at the center of the array vs. the source position.
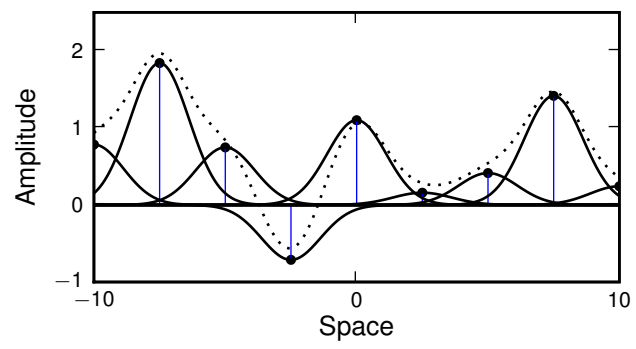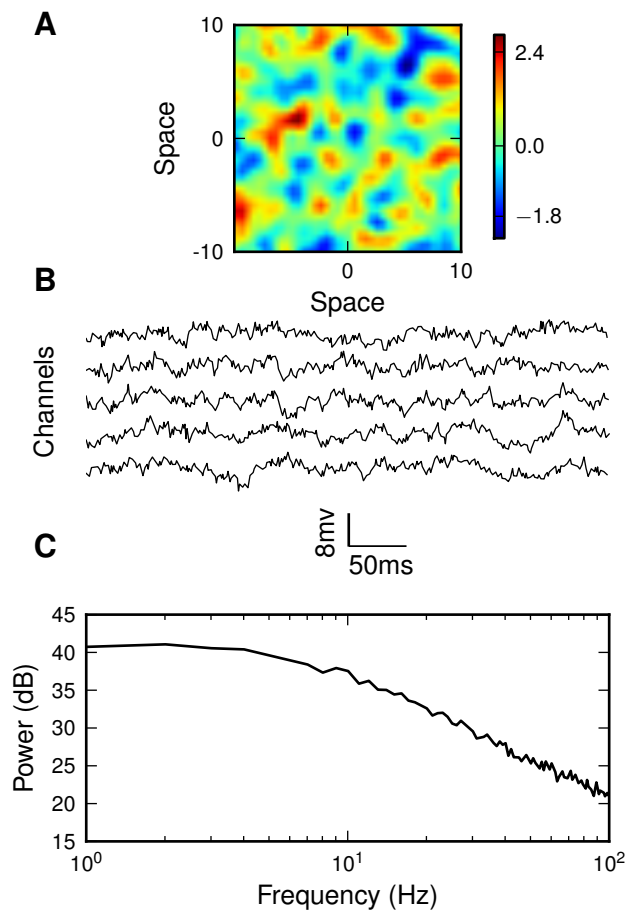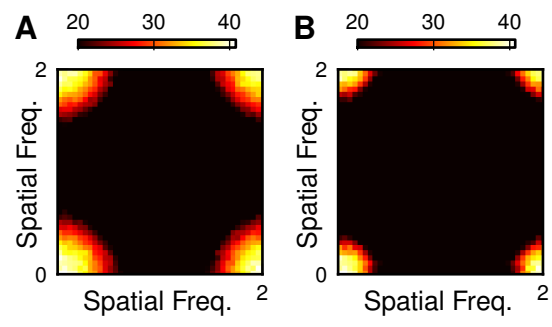
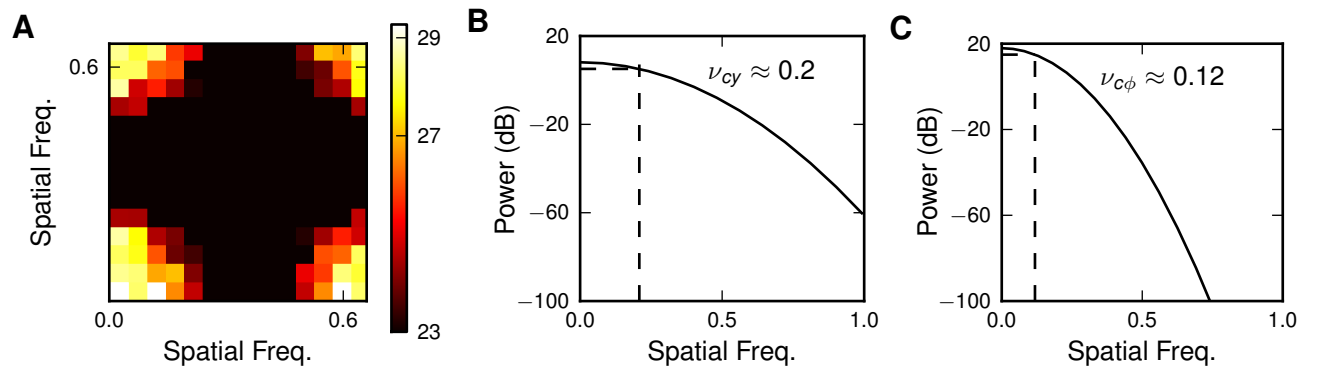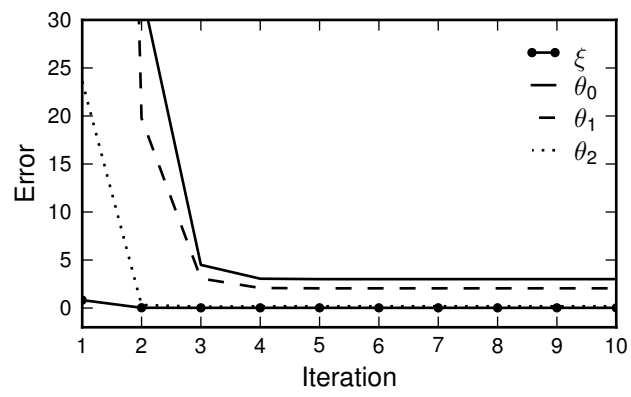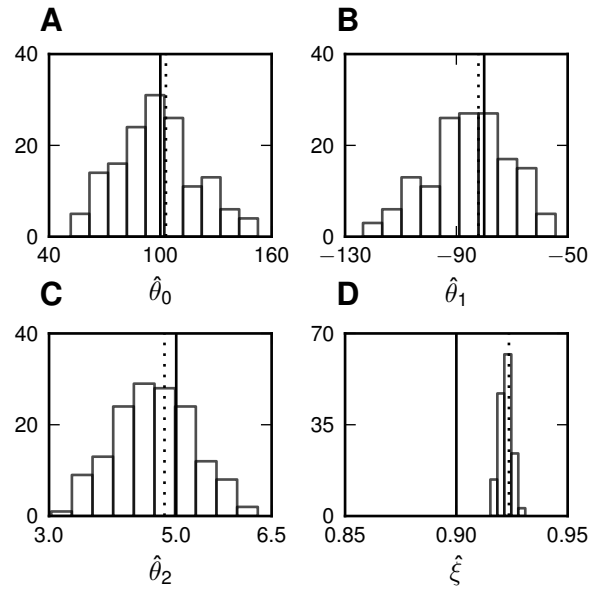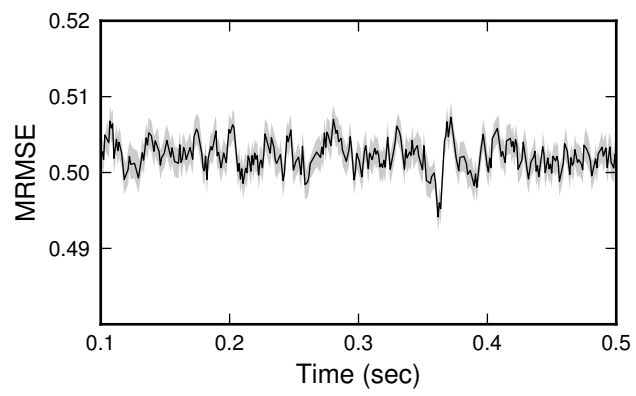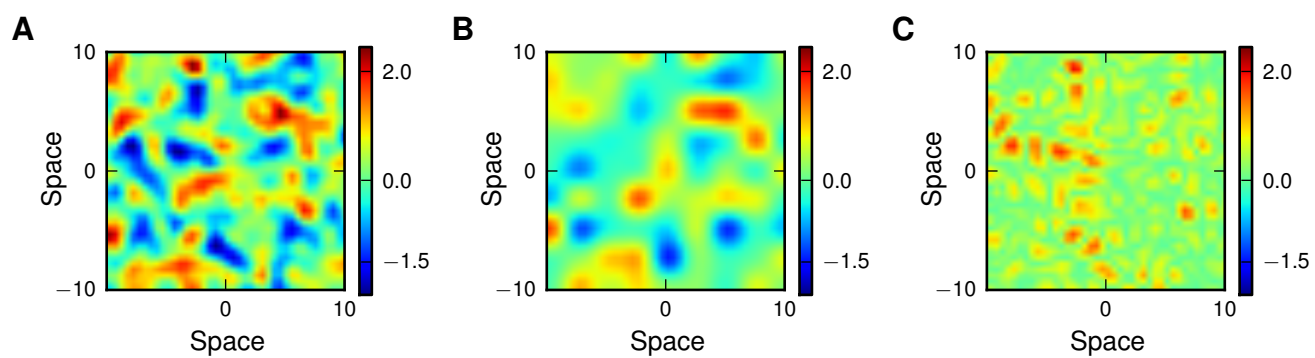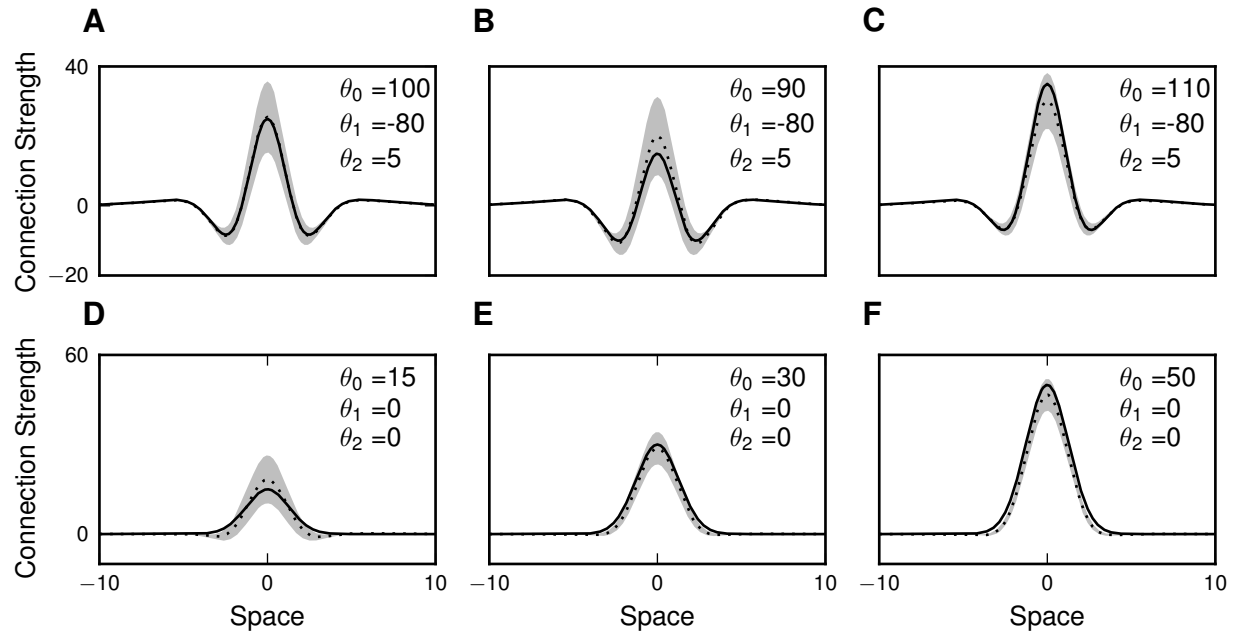Figure 1:

Figure 2:

A

B

C
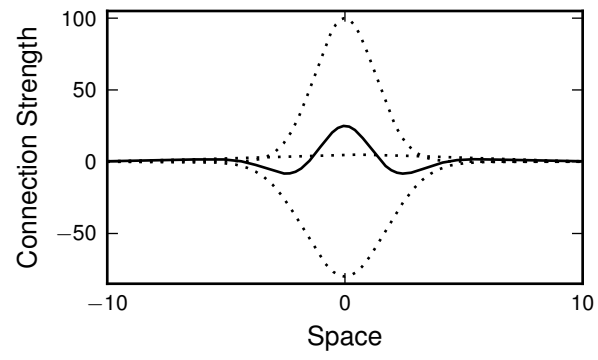
Figure 3:

Figure 4:

Figure 5:

Figure 6:

Figure 7:

Figure 8:

A

B

C

Figure 9:

A

$\theta_0 = 100$
$\theta_1 = -80$
$\theta_2 = 5$

B

$\theta_0 = 90$
$\theta_1 = -80$
$\theta_2 = 5$

C

$\theta_0 = 110$
$\theta_1 = -80$
$\theta_2 = 5$

D

$\theta_0 = 15$
$\theta_1 = 0$
$\theta_2 = 0$

E

$\theta_0 = 30$
$\theta_1 = 0$
$\theta_2 = 0$

F

$\theta_0 = 50$
$\theta_1 = 0$
$\theta_2 = 0$

Connection Strength
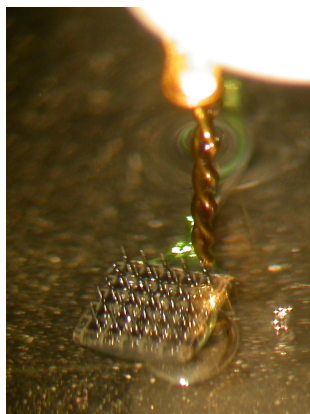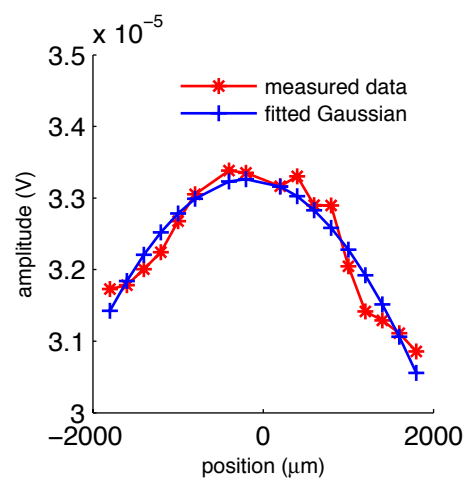
Space
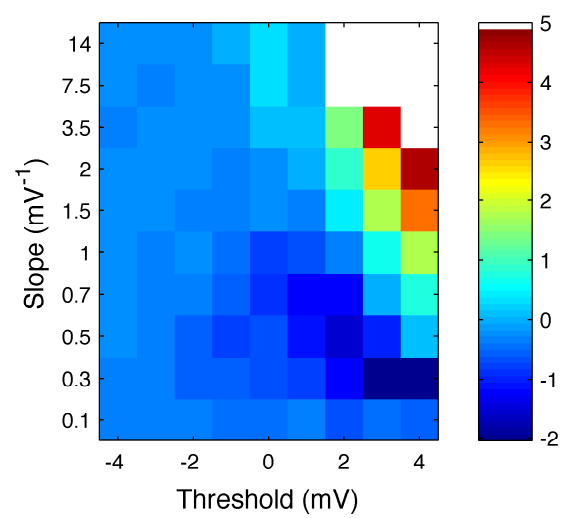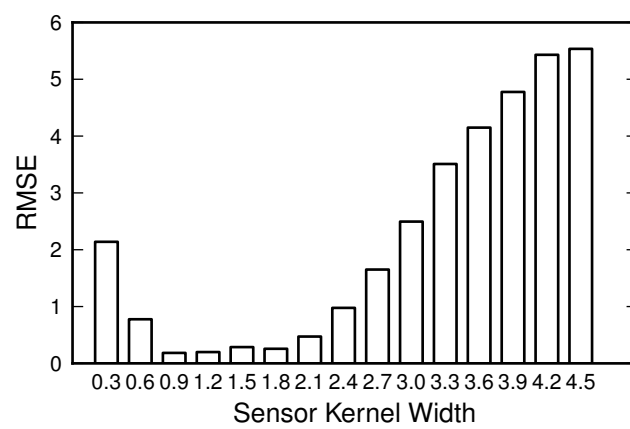
Figure 10:

Figure 11:

Figure 12:

Figure 13:

Figure 14:

Appendix A. Time Discretization of the Neural Field Model

This appendix provides a description of the discretization of the neural field model to form the integro-difference equation (IDE). To perform the discretization, we used a one-step Euler method, where equation 11 can be approximated by

$$\frac{v\left(\mathbf{r}, t + T_s\right) - v\left(\mathbf{r}, t\right)}{T_s} = -\zeta v\left(\mathbf{r}, t\right)$$
$$+ \int_\Omega w\left(\mathbf{r} - \mathbf{r}'\right) f\left(v\left(\mathbf{r}', t\right)\right) d\mathbf{r}'. \tag{A.1}$$

For clarity, we shall index time points in the discrete time form of the model using the subscript t and the next time point as $t + 1$. Rearranging equation A.1, we get

$$v_{t+1}\left(\mathbf{r}\right) = v_t\left(\mathbf{r}\right) - T_s \zeta v_t\left(\mathbf{r}\right) + T_s \int_\Omega w\left(\mathbf{r} - \mathbf{r}'\right) f\left(v_t\left(\mathbf{r}'\right)\right) d\mathbf{r}'. \tag{A.2}$$

Now making simplifications, the discrete time form of the model is

$$v_{t+1}\left(\mathbf{r}\right) = \xi v_t\left(\mathbf{r}\right) + T_s \int_\Omega w\left(\mathbf{r} - \mathbf{r}'\right) f\left(v_t\left(\mathbf{r}'\right)\right) d\mathbf{r}', \tag{A.3}$$

where $\xi = 1 - T_s \zeta$.

Note, the discretization using the first order Euler method has the potential to affect the system's behaviour, where the difference between the continuous and discrete-time systems can potentially be significant. The sampling period of 1 ms was used in this paper which is ten times smaller than the synaptic time constant, therefore the discrepancy between the continuous and its discrete approximation is small. This yielded a stable discrete-time system and good estimation results. An analysis of the error introduced by Euler discretization method can be found in **?**. An alternative to Euler's method is the local linearization method (**?**).

Appendix B. Observation Kernel

The Gaussian shape of the observation kernel was verified experimentally. This was achieved by moving a dipole across a recording electrode submerged in saline. A photograph through a microscope of the submerged electrodes can be seen in Fig. 13. The measurement sensor was the most central electrode of a 5×5 Utah array (Blackrock Microsystems) which had electrode spacing of 400 μm. The current source

was a twisted wire electrode (Plastics One Inc.) with tip-tip spacing of 60 μm and tip diameter of 200 μm. The twisted wire electrode was moved in the saline solution using a 3 dimensional manipulator (Sutter Instrument Company, MPC-100) in increments of 200 μm. The twisted wire electrode and the Utah array were co-registered visually using the microscope. The stimulation waveform for the current source consisted of a 40 Hz sine wave with a duration of 1 s. The recording from the central electrode in the array was used to estimate the shape of the sensor kernel. To extract the amplitude of the signal, the data was first band-pass filtered with a center frequency of 40 Hz. The envelope of the signal was found by computing the mean instantaneous amplitude over the 1 s stimulation duration by

$$a = \frac{1}{T} \sum_t \sqrt{(y_t^2 - \mathcal{H} y_t^2)} \tag{B.1}$$

where $\mathcal{H}$ denotes the Hilbert transform operator, t indexes samples and T is the total number of samples in the time series. Fig. 14 shows the mean amplitude of the recorded envelope versus the distance between the twisted wire electrode and the recording electrode. A Gaussian fitted to the amplitudes is also shown. The figure shows that a Gaussian is indeed a good model for the sensor kernel in the homogeneous solution.

Appendix C. Equivalence to Galerkin's Projection

To show how equation 19 is equivalent to the Galerkin projection, we define the residual between the field and the decomposition as

$$\begin{aligned}
\Xi(\mathbf{x}_t, v_{t+1}(\mathbf{r})) = {}& v_{t+1}(\mathbf{r}) \\
& - T_s \int_\Omega f(\boldsymbol{\phi}^\top(\mathbf{r}')\mathbf{x}_t)\boldsymbol{\psi}^\top(\mathbf{r} - \mathbf{r}')\, d\mathbf{r}'\boldsymbol{\theta} \\
& - \xi\boldsymbol{\phi}^\top(\mathbf{r})\mathbf{x}_t - \mathbf{e}_t(\mathbf{r}).
\end{aligned} \tag{C.1}$$

The Galerkin projection finds $\mathbf{x}_t$, such that the inner product of the residual with all of the basis functions satisfies

$$G_i = \int_\Omega \Xi(\mathbf{x}_t, v_{t+1}(\mathbf{r}))\phi_i(\mathbf{r})d\mathbf{r} = 0, \tag{C.2}$$

where $i = 0, 1, \ldots, n_x - 1$, and n_x is the number of states and basis functions. If equation C.2 is satisfied the basis functions are orthogonal to the residual. By substituting equation C.1 into equation C.2 and rearranging we get equation 19 of the main text. Note, if Galerkin's method was used to estimate the states, n_x coupled ODEs would need to be solved independently (each G_i) to estimate the state vector, which is

a standard method in fluid dynamics and economics modeling. The method in this paper takes a different approach by forming the state-space model, allowing for more efficient estimation.

Appendix D. Inner Product of Two n-Dimensional Gaussians

The derivation in this appendix was used for calculating the matrix $\boldsymbol{\Gamma}$ from equation 21. Consider two Gaussian basis functions

$$\varphi_i(\mathbf{r}) = \exp\left(-\frac{1}{\sigma_i^2}(\mathbf{r} - \boldsymbol{\mu}_i)^\top (\mathbf{r} - \boldsymbol{\mu}_i)\right) \tag{D.1}$$

and

$$\varphi_j(\mathbf{r}) = \exp\left(-\frac{1}{\sigma_j^2}(\mathbf{r} - \boldsymbol{\mu}_j)^\top (\mathbf{r} - \boldsymbol{\mu}_j)\right). \tag{D.2}$$

The inner product of the two n-dimensional Gaussians is defined as

$$\varphi_i \otimes \varphi_j \triangleq \int_{\mathbb{R}^n} \varphi_i(\mathbf{r})\varphi_j(\mathbf{r})d\mathbf{r}. \tag{D.3}$$

A Gaussian basis function is symmetric with respect to its center, therefore

$$\varphi_i(\boldsymbol{\mu}_i + \mathbf{r}) = \varphi_i(\boldsymbol{\mu}_i - \mathbf{r}). \tag{D.4}$$

Letting $\mathbf{r} = \mathbf{r} - \boldsymbol{\mu}_i$, equation D.4 becomes

$$\varphi_i(\mathbf{r}) = \varphi_i(2\boldsymbol{\mu}_i - \mathbf{r}). \tag{D.5}$$

Now substituting equation D.5 into equation D.3 gives

$$\begin{aligned}
\varphi_i \otimes \varphi_j &= \int_{\mathbb{R}^n} \varphi_i(\mathbf{r})\varphi_j(\mathbf{r})d\mathbf{r} \\
&= \int_{\mathbb{R}^n} \varphi_i(2\boldsymbol{\mu}_i - \mathbf{r})\varphi_j(\mathbf{r})d\mathbf{r} \\
&= (\varphi_i * \varphi_j)(2\boldsymbol{\mu}_i).
\end{aligned} \tag{D.6}$$

Now we solve the convolution using the formula derived in Appendix E, and evaluating it at $\mathbf{r} = 2\boldsymbol{\mu}_i$ gives

$$
\begin{aligned}
\varphi_i \otimes \varphi_j &= \left(\frac{\pi \sigma_i^2 \sigma_j^2}{\sigma_i^2 + \sigma_j^2} \right)^{\frac{n}{2}} \\
&\quad \times \exp\left(-\frac{1}{\sigma_i^2 + \sigma_j^2} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^\top (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j) \right).
\end{aligned}
\tag{D.7}
$$

This completes the derivation.

Appendix E. Convolution of Two n-Dimensional Gaussians

In this section, we provide a derivation for the convolution of two n-dimensional Gaussian basis functions. This derivation is used in the analytic calculation of $\boldsymbol{\Psi}(\mathbf{r})$ and the elements of the observation matrix, $\mathbf{C}$. Consider the two Gaussian basis functions from equations D.1 and D.2. The derivation proceeds by assuming the Gaussian basis functions are positioned at the origin ($\boldsymbol{\mu}_i = \boldsymbol{\mu}_j = \mathbf{0}$). The convolution theorem states that the Fourier transform of the convolution of two functions is equal to the product of their Fourier transforms. Using the formula from the derivation of the Fourier transform of an n-dimensional Gaussian from equation H.4 in Appendix H, we have

$$
\mathcal{F} \left(\varphi_i * \varphi_j \right) (\mathbf{r}) = \left(\pi^2 \sigma_i^2 \sigma_j^2 \right)^{\frac{n}{2}} \exp\left(-(\sigma_i^2 + \sigma_j^2) \pi^2 \boldsymbol{\nu}^\top \boldsymbol{\nu} \right),
\tag{E.1}
$$

where $\mathcal{F}$ denotes the Fourier transform. Multiplying and dividing by $\left(\pi \left(\sigma_i^2 + \sigma_j^2 \right) \right)^{\frac{n}{2}}$, equation E.1 can be written as

$$
\begin{aligned}
\mathcal{F} \left(\varphi_i * \varphi_j \right) (\mathbf{r}) &= \left(\pi \sigma_i^2 \sigma_j^2 \right)^{\frac{n}{2}} \left(\frac{1}{\sigma_i^2 + \sigma_j^2} \right)^{\frac{n}{2}} (\pi(\sigma_i^2 + \sigma_j^2))^{\frac{n}{2}} \\
&\quad \times \exp\left(-(\sigma_i^2 + \sigma_j^2) \pi^2 \boldsymbol{\nu}^\top \boldsymbol{\nu} \right).
\end{aligned}
\tag{E.2}
$$

Taking the inverse Fourier transform, the convolution becomes

$$
\left(\varphi_i * \varphi_j \right) (\mathbf{r}) = \left(\frac{\pi \sigma_i^2 \sigma_j^2}{\sigma_i^2 + \sigma_j^2} \right)^{\frac{n}{2}} \exp\left(-\frac{1}{\sigma_i^2 + \sigma_j^2} \mathbf{r}^\top \mathbf{r} \right).
\tag{E.3}
$$

In the case where $\boldsymbol{\mu}_i \neq \boldsymbol{\mu}_j \neq \mathbf{0}$, equation E.3 becomes

$$(\varphi_i * \varphi_j)(\mathbf{r}) = \left(\frac{\pi \sigma_i^2 \sigma_j^2}{\sigma_i^2 + \sigma_j^2}\right)^{\frac{n}{2}}$$
$$\times \exp\left(-\frac{1}{\sigma_i^2 + \sigma_j^2}(\mathbf{r} - \boldsymbol{\mu})^\top (\mathbf{r} - \boldsymbol{\mu})\right), \tag{E.4}$$

where $\boldsymbol{\mu} = \boldsymbol{\mu}_i + \boldsymbol{\mu}_j$.

Appendix F. Mean and Covariance of the State Disturbance

In this section, we show that the state disturbance vector is $\mathbf{e}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_e)$ if $e_t(\mathbf{r}) \sim \mathcal{GP}(\mathbf{0}, \gamma(\mathbf{r} - \mathbf{r}'))$. We begin the derivation by acknowledging the state disturbance vector is a linear function of $e_t(\mathbf{r})$. Therefore, $\mathbf{e}_t$ also follows a Gaussian distribution. From equation 26, the state disturbance vector is defined as

$$\mathbf{e}_t \triangleq \boldsymbol{\Gamma}^{-1} \int_\Omega \boldsymbol{\phi}(\mathbf{r}) e_t(\mathbf{r}) d\mathbf{r}. \tag{F.1}$$

The expected value of $\mathbf{e}_t$ is given by

$$\mathbf{E}[\mathbf{e}_t] = \boldsymbol{\Gamma}^{-1} \int_\Omega \boldsymbol{\phi}(\mathbf{r}) \mathbf{E}[e_t(\mathbf{r})] d\mathbf{r} = \mathbf{0}. \tag{F.2}$$

The covariance of $\mathbf{e}_t$ is

$$\begin{aligned}
\boldsymbol{\Sigma}_e &= \mathbf{E}[\mathbf{e}_t \mathbf{e}_t^\top] \\
&= \boldsymbol{\Gamma}^{-1} \mathbf{E}[\int_\Omega \boldsymbol{\phi}(\mathbf{r}) e_t(\mathbf{r}) d\mathbf{r} \int_\Omega \boldsymbol{\phi}^\top(\mathbf{r}') e_t(\mathbf{r}') d\mathbf{r}'] \boldsymbol{\Gamma}^{-\top} \\
&= \boldsymbol{\Gamma}^{-1} \int_\Omega \int_\Omega \boldsymbol{\phi}(\mathbf{r}) \mathbf{E}[e_t(\mathbf{r}) e_t(\mathbf{r}')] \boldsymbol{\phi}^\top(\mathbf{r}') d\mathbf{r}' d\mathbf{r} \boldsymbol{\Gamma}^{-\top} \\
&= \boldsymbol{\Gamma}^{-1} \int_\Omega \int_\Omega \boldsymbol{\phi}(\mathbf{r}) \gamma(\mathbf{r} - \mathbf{r}') \boldsymbol{\phi}^\top(\mathbf{r}') d\mathbf{r}' d\mathbf{r} \boldsymbol{\Gamma}^{-\top}. \tag{F.3}
\end{aligned}$$

Appendix G. Derivation of Sampling Theorem

This appendix provides a derivation for equations 35 and 36, which provide conditions on distance between the sensors and the field basis functions. The derivation is adapted from ? and ?. Consider the cortical sheet defined on the surface Ω with sensors placed on a regular grid at positions $\mathbf{r}_n$ (where $n = 0, ..., n_y - 1$ indexes the sensors) that are separated by the distance Δ_y. Noise-free observations can be

considered samples from a continuous field that is filtered (convolved) with the observation kernel described by

$$y(\mathbf{r}_n) = \delta(\mathbf{r} - \mathbf{r}_n)v_m(\mathbf{r}), \tag{G.1}$$

where $v_m(\mathbf{r}) = v_t(\mathbf{r}) * m(\mathbf{r})$. The observations are described in matrix form by

$$\mathbf{y} = \sum_{n=0}^{n_y-1} \delta(\mathbf{r} - \mathbf{r}_n)v_m(\mathbf{r}). \tag{G.2}$$

The Fourier transform of $\mathbf{y}$ is

$$Y(\boldsymbol{\nu}) = \frac{1}{\Delta_y^2} \sum_{n=0}^{n_y-1} \delta\left(\boldsymbol{\nu} - \frac{\mathbf{k}_n}{\Delta_y}\right) * V_m(\boldsymbol{\nu}) \tag{G.3}$$

$$= \frac{1}{\Delta_y^2} V_m(\boldsymbol{\nu}) + \frac{1}{\Delta_y^2} \sum_{n=1}^{n_y-1} V_m\left(\boldsymbol{\nu} - \frac{\mathbf{k}_n}{\Delta_y}\right), \tag{G.4}$$

where $\mathbf{k}$ is an $n_y \times 2$ matrix, spatially indexing the sensors (the columns are different spatial dimensions such that $\mathbf{r}_n = \Delta_y \mathbf{k}_n$). It should be clear that $Y(\boldsymbol{\nu})$ contains $V_m(\boldsymbol{\nu})$ plus $n_y - 1$ copies (aliases) of $V_m(\boldsymbol{\nu})$ repeated on the regular lattice, where each copy is centered at $\mathbf{k}_n \Delta_y^{-1}$. Due to the lowpass action of the observation kernel, the filtered dynamic field is band-limited with a cutoff frequency $\boldsymbol{\nu}_{cy}$. In order for $V_m(\boldsymbol{\nu})$ to be disjointed from its aliases, the amount of shift Δ_y^{-1} must be at least $2\boldsymbol{\nu}_{cy}$, since each alias has a spectral range of

$$\boldsymbol{\nu} = \left[\frac{\mathbf{k}_n}{\Delta_y} - \boldsymbol{\nu}_{cy}, \frac{\mathbf{k}_n}{\Delta_y} + \boldsymbol{\nu}_{cy}\right]. \tag{G.5}$$

Therefore, to avoid aliasing the distance between observations must satisfy

$$\frac{1}{\Delta_y} \geq 2\boldsymbol{\nu}_{cy}. \tag{G.6}$$

Since the observation kernel is not an ideal filter, an oversampling parameter should be included to further separate the aliases. Rearranging equation G.6 and adding the oversampling parameter, ρ, gives

$$\Delta_y \leq \frac{1}{2\rho\boldsymbol{\nu}_{cy}}, \tag{G.7}$$

where $\rho \in \mathbb{R} \geq 1$.

Appendix H. Spatial Cut-off Frequency of the Field Basis Functions

This section provides a derivation for the width of the Fourier transform of the field basis functions in terms of their spatial cut-off frequency (equations 38 and 41). We define an n-dimensional Gaussian centered at the origin as

$$\phi(\mathbf{r}) = \exp\left(-\frac{1}{\sigma_\phi^2}\mathbf{r}^\top\mathbf{r}\right).\tag{H.1}$$

Now applying the n-dimensional Fourier transform yields

$$\begin{aligned}
\mathbf{\Phi}(\boldsymbol{\nu}) &= \int_{\mathbb{R}^n} \exp\left(-\frac{1}{\sigma_\phi^2}\mathbf{r}^\top\mathbf{r}\right)\exp\left(-2\pi i\boldsymbol{\nu}^\top\mathbf{r}\right)d\mathbf{r} \\
&= \int_{\mathbb{R}^n} \exp\left(-\frac{1}{\sigma_\phi^2}\left[\mathbf{r}+\sigma_\phi\pi i\boldsymbol{\nu}\right]^\top\left[\mathbf{r}+\sigma_\phi\pi i\boldsymbol{\nu}\right]\right)d\mathbf{r} \\
&\quad\times \exp\left(-\sigma_\phi^2\pi^2\boldsymbol{\nu}^\top\boldsymbol{\nu}\right).
\end{aligned}\tag{H.2}$$

Evaluating the integral on the RHS of equation H.2 gives

$$\int_{\mathbb{R}^n} \exp\left(-\frac{1}{\sigma_\phi^2}\left[\mathbf{r}+\sigma_\phi\pi i\boldsymbol{\nu}\right]^\top\left[\mathbf{r}+\sigma_\phi\pi i\boldsymbol{\nu}\right]\right)d\mathbf{r} = \left(\pi\sigma_\phi^2\right)^{\frac{n}{2}}.\tag{H.3}$$

Substituting equation H.3 into equation H.2 gives the scaled Gaussian

$$\mathbf{\Phi}(\boldsymbol{\nu}) = \left(\frac{1}{\pi\sigma_\nu^2}\right)^{\frac{n}{2}}\exp\left(-\frac{1}{\sigma_\nu^2}\boldsymbol{\nu}^\top\boldsymbol{\nu}\right),\tag{H.4}$$

where

$$\sigma_\nu^2 = \frac{1}{\pi^2\sigma_\phi^2}.\tag{H.5}$$

For 3 dB attenuation (1/2 to peak power) at the cut-off frequency, $\boldsymbol{\nu}_c$, we need to set

$$|\mathbf{\Phi}(\boldsymbol{\nu}_c)|^2 = \frac{1}{2}|\mathbf{\Phi}(\mathbf{0})|^2,\tag{H.6}$$

since the maximum of the Gaussian described in equation H.4 is at the origin. Solving for σ_ν^2, we get

$$\sigma_\nu^2 = \frac{2\boldsymbol{\nu}_c^\top\boldsymbol{\nu}_c}{\ln 2}.\tag{H.7}$$