Estimating the unmeasured membrane potential of neuronal populations from the EEG using a class of deterministic nonlinear filters

Michelle Chong, Romain Postoyan, Dragan Nesic, Levin Kuhlmann, Andrea Varsavsky

Estimating the unmeasured membrane potential of neuronal populations from the EEG using a class of deterministic nonlinear observers

Michelle Chong[1], Romain Postoyan[2]‡, Dragan Nešić[1], Levin Kuhlmann[1,3] and Andrea Varsavsky[1]

[1] Department of Electrical and Electronic Engineering, The University of Melbourne, Australia.
[2] Centre de Recherche en Automatique de Nancy, UMR 7039, Nancy-Université, CNRS, France.
[3] Department of Optometry and Vision Science, The University of Melbourne, Australia.

E-mail: chongms@unimelb.edu.au, romain.postoyan@cran.uhp-nancy.fr, dnesic@unimelb.edu.au, levink@unimelb.edu.au, ava@unimelb.edu.au

Abstract. We present a model-based estimation method to reconstruct the unmeasured membrane potential of neuronal populations from a single channel electroencephalographic (EEG) measurement. We consider a class of neural mass models that share a general structure, specifically the models by Stam et. al. in [1], Jansen et. al. in [2] and Wendling et. al. in [3]. Under idealised assumptions, we prove the global exponential convergence of our observer. Then, under more realistic assumptions, we investigate the robustness of our observer against model uncertainties and disturbances. Analytic proofs are provided for all results and our analyses are further illustrated via simulations.

PACS numbers: 87.19.le, 87.19.lj, 87.19.ll, 87.19.lr

Submitted to: *Journal of Neural Engineering*

‡ This work was carried out while the author was in the Department of Electrical and Electronics Engineering, The University of Melbourne, Australia.

1. Introduction

Brain studies involving the EEG typically employ signal processing strategies, which can be categorised into measure-based or model-based methods. The measure-based method involves extracting key measures from the EEG signal to classify brain phenomena. This approach is prevalent for the purpose of seizure detection (e.g. [4, 5]) and prediction (e.g. [6, 7]) in epilepsy. These methods are continuously under development due to the strict conditions in which they can be considered successful [8].

On the other hand, model-based estimation methods can provide insight into the complex nature of the brain via mathematical models (e.g. [9, 10, 11, 12] to name a few). These models describe brain activity at different scales, from a *single neuron* to *populations of neurons* and these activities are captured by states of a dynamical system. These states can be physiologically relevant and most are unmeasurable by conventional means. The model-based approach of estimating the unmeasured brain activity is called *state estimation*. In systems theory literature, the algorithm to estimate states from a measurement is commonly known as 'observer', 'filter' or 'estimator' [13, 14]. We will use the term 'observer' in this paper.

The objective of this study is to propose a model-based algorithm to estimate online the hidden states of neural mass models formulated as deterministic ordinary differential equations in continuous time. As such, this scheme similar in flavour to Dynamic Causal Modelling [15, 16, 17, 18, 19, 20], which aims to recover the hidden states and parameters of an underlying dynamical system from observed time series. In this paper, we focus on estimating hidden states, noting that this is useful for model identification or parameter estimation. The ability to recover hidden states and parameters from electrophysiological data is of clear importance; for example, in the diagnosis, classification and anticipation of seizures in epilepsy.

Most of the work in state estimation for neuroscientific and neurological studies are developed under the stochastic framework, using the Kalman filter and its variants, or by employing the expectation maximisation algorithm [21, 22, 23, 24, 25, 26]. However, the usage of stochastic observers on a nonlinear model such as [9] and [3] requires that the observer be initialised close to the true initial condition, which is realistically unknown. Moreover, the convergence of the estimates to the true states is not guaranteed for every trajectory.

Deterministic state estimation circumvents these issues. Nevertheless, the main difficulty lies in that there exists no generic method of designing deterministic observers for nonlinear models. As existing techniques do not apply to the models we consider, we developed our own observer and provide analytical proofs for the convergence of state estimates to the true states.

Deterministic state estimation methods at the microscopic level (single neuron) include works by Tokuda et. al. in [27], Totoki et. al. in [28], Tyukin et. al. in [29] and Mao et. al. in [30]. As our measurement of choice is the EEG, which is known to reflect the average behaviour of *populations of neurons* [31], our work differs by taking a macroscopic view. We focus on designing deterministic observers for estimating the unmeasured mean membrane potential of *neuronal populations*. At present, it is only possible to measure the membrane potential of neuronal populations through invasive implantation of electrodes. The positioning and size of electrodes need to be accurate enough such that the measurements are not influenced by nearby populations. To this end, we consider physiologically relevant models that describe the interaction between

neuronal populations, known as *neural mass models* (see [32] for review), that are known to replicate patterns typically seen in the EEG (e.g. [9, 10, 11, 3, 1, 2] to name a few). We consider three neural mass models available in the literature that share a general mathematical structure, they describe: (i) alpha rhythms [1], (ii) alpha rhythms in the cerebral cortex [2] and (iii) epileptic activity in the hippocampus [3]. We rewrite these neural mass models in state space form, specifically, in state coordinates that are amenable for observer design. We also identify the common model features that allow us to prove convergence of state estimates to the true states in a unified manner.

In the development of our observer, we first assume that there are no uncertainties and external disturbances that affect our model. We later relax these assumptions and show that the proposed class of observers is robust to uncertainty in parameters, additive disturbance, as well as noise and artefacts in the input and EEG measurement. The robustness of this class of observers to parameter uncertainty provides a starting point in the final aim of building an online observer for both the states and parameters of the model that will form the further work of this study. The contributions of this paper are:

- The design of a class of deterministic nonlinear observers for neural mass models that guarantee convergence of the estimated states to the true states for any initial condition and any known input. Analytical proofs using system theoretic methods are provided in the Appendix.
- Robustness analysis of our proposed observers towards common issues faced in estimating brain activity. These include measurement noise, uncertainty in the parameters and the measured input. Simulation studies confirm the analyses performed.

The rest of the paper is organised as follows. We start by introducing the notations and formally state the problem in Section 2. Section 3 contains a more detailed exposition on the three neural mass models of interest. Then in Section 4, we show how to convert these neural mass models into a state space form that is amenable for observer design and error convergence analysis. Section 5 contains the main simplifying assumptions. Section 6 contains the main result of the paper that states the exponential convergence properties of the state estimates to the true states. We then relax some of the earlier strong assumptions and show that the observer is robust under a more realistic set of assumptions. We present simulation results in Section 7. In Section 8, We provide some discussions and conclude the paper.

Notation

- The matrix $\begin{bmatrix} A_1 & 0 & \dots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & A_n \end{bmatrix}$ where A_i for $i \in \{1, \dots, n\}$ are $m \times m$ matrices, is denoted as $\text{diag}(A_1, \dots, A_n)$.
- The set $\mathcal{L}_\infty$ denotes the set of functions $f : \mathbb{R} \to \mathbb{R}^n$, for some $n \in \mathbb{Z}$, such that for any $0 \leq t_1 \leq t_2 < \infty$ there exists $r \geq 0$ so that $\|f\|_{[t_1,t_2]} := \text{ess sup}_{\tau \in [t_1,t_2]} |f(\tau)| \leq r$.

- A vector $u \in \mathbb{R}^n$, where each element is drawn from a Gaussian distribution with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 \in \mathbb{R}$ is denoted $u \sim \mathcal{N}(M, \sigma^2 \mathbb{I})$, where $M = (\mu, \ldots, \mu) \in \mathbb{R}^n$ and $\mathbb{I} \in \mathbb{R}^{n \times n}$ is an identity matrix.

2. Problem statement

We consider several classes of neural mass models described in [1], [2] and [3]. These models have some common features and can all be represented by the same general functional block diagram that is shown in Figure 1. The input to the model u represents all the external influences from afferent neuronal populations or structures. The output of the model y is taken to be the EEG measurement, which is proportional to the aggregated average postsynaptic potentials of the pyramidal neurons. The state x is a vector whose entries are the afferent population specific contributions to the mean membrane potential of each neuronal population and its derivative. The parameter θ is a vector whose entries are the connectivity strengths and the synaptic gain of each neuronal population. The state x, output y and parameter θ have subtly different meaning for each model we consider and these differences are discussed in more detail in the following sections.

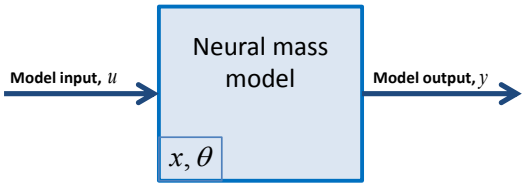

Figure 1. Block diagram of the neural mass model.

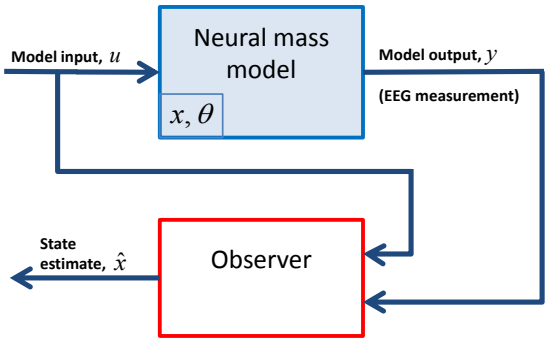

Figure 2. Block diagram of the neural mass model with our observer.

Our goal is to design a model-based algorithm (observer) which uses measurements from u and y to compute the state estimate $\hat{x}$ as illustrated in Figure 2. We show how to design the observer for a general class of systems in state space form that covers many neural mass models. The observer is designed to provide a state estimate $\hat{x}$ that converges to the real state x in exponential time, for any initial

condition and any input. In the next section, we describe the neural mass models considered in more detail.

3. Neural mass models

We consider several classes of neural mass models that describe the interconnection between neuronal populations as shown in Figures 3-5. These block diagrams describe a functional relationship between various neuronal populations. They are developed by: (i) Stam et. al. in [1], (ii) Jansen and Rit in [2] and (iii) Wendling et. al. in [3]. These models are based on the models by Wilson and Cowan in [33] as well as Lopes da Silva in [34]. In the following sections, the classes of neural mass models are introduced with increasing level of complexity. Noting that the models by Stam et. al. as well as Jansen and Rit can be obtained from the model by Wendling et. al. by omitting certain neuronal populations, a detailed block diagram is presented only for the Wendling et. al. model in Section 3.3.

3.1. Neural mass model by Stam et. al.

We begin with the model by Stam et. al., which includes an excitatory and inhibitory population to replicate alpha rhythms as seen in Figure 3. The emergence of alpha rhythms in the EEG is related to the human subject being in a relaxed state with eyes closed. Our observer is used to estimate the membrane potential of each population and its derivatives from the EEG and input measurements. Estimating the unmeasured membrane potential of neuronal populations when alpha rhythms are seen on the EEG measurement could allow for the better understanding of the visual pathway while in an idle state.

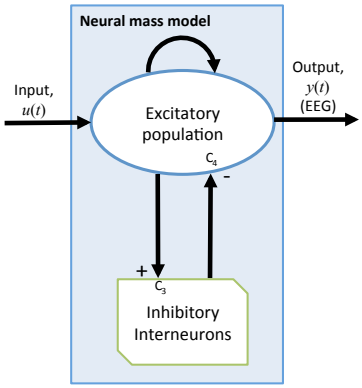

Figure 3. Neural mass model by Stam et. al. in [1]

3.2. Neural mass model by Jansen and Rit

The interconnections between the pyramidal neurons, excitatory and inhibitory populations are described in this model to investigate the generation of visual evoked potentials in the cerebral cortex, as illustrated in Figure 4. We design observers to

estimate the unmeasured membrane potential of neuronal populations, which could give insight into the stimulation of a sensory pathway.

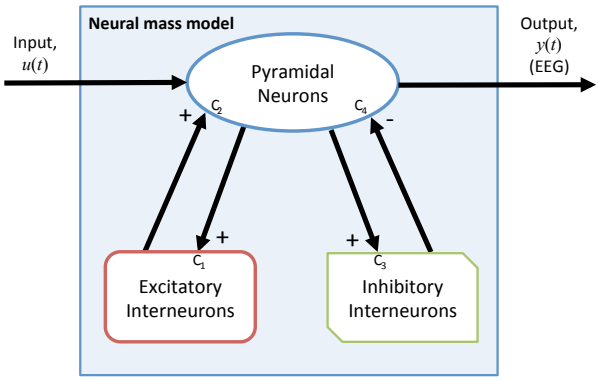

Figure 4. Neural mass model by Jansen and Rit in [2]

3.3. Neural mass model by Wendling et. al.

Wendling et. al. built upon the Jansen and Rit model described in Section 3.2. Four neuronal populations (with one population being a subset of another) are included in this model as shown in Figure 5. They are the pyramidal neurons, the excitatory population (included in the pyramidal neurons), the slow and fast inhibitory populations. The fast somatic projection of the inhibitory population is introduced in this model as it is hypothesised to play a role in the fast oscillatory pattern seen in the EEG during epileptic seizure onsets. Estimating the membrane potential contribution of one population to another could provide insight into the underlying dynamics of seizure activity.

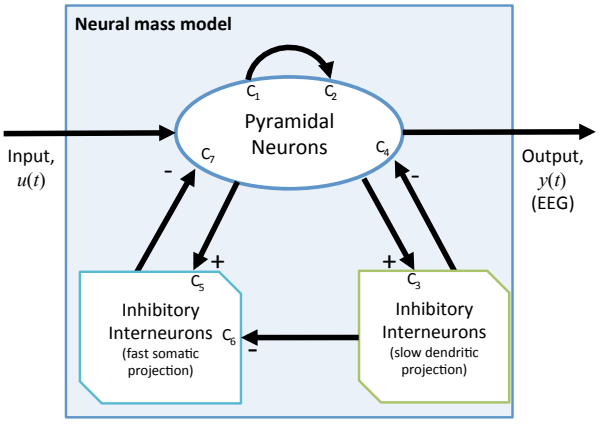

Figure 5. Neural mass model by Wendling et. al. in [3].

Figure 6 describes the interaction between populations of neurons in greater detail, which consists of postsynaptic membrane potential (PSP) kernels h_e, h_i and

h_g, sigmoid functions $S : \mathbb{R} \to \mathbb{R}$ and connectivity constants C_1 to C_7.

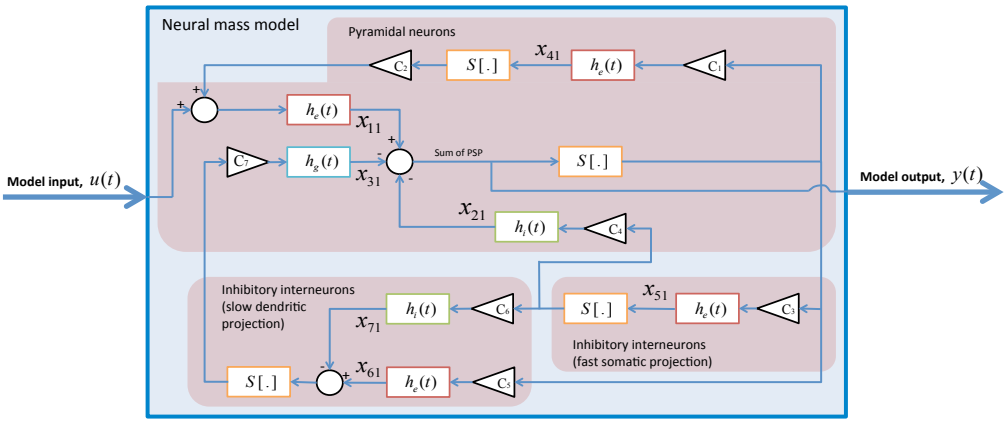

Figure 6. Detailed block diagram of the Wendling et. al. model. Reproduced from Figure 4 in [3].

The firing rate of the afferent population is converted into an excitatory, slow or fast inhibitory postsynaptic membrane potential via the following kernels, for $t \geq 0$:

- The excitatory population:

$$h_e(t) = \theta_A a t \exp(-at). \tag{1}$$

- The slow inhibitory population:

$$h_i(t) = \theta_B b t \exp(-bt). \tag{2}$$

- The fast inhibitory population:

$$h_g(t) = \theta_G g t \exp(-gt). \tag{3}$$

Parameters θ_A, θ_B and θ_G in (1)-(3) correspond to the synaptic gains of the excitatory, slow and fast inhibitory populations respectively. These parameters characterise the observed pattern in the EEG. For example, the values of θ_A, θ_B and θ_G that distinguish between seizure and non-seizure activities have been identified in [3].

Internal variables $x_{11}, \ldots, x_{71}$ are introduced as shown in Figures 3-6. They describe the membrane potential contribution from one population to another. For example as seen in Figure 6, the mean membrane potential of the pyramidal neurons is $x_{11} - x_{21} - x_{31}$, which reflects the membrane potential contribution from the excitatory, slow and fast inhibitory populations respectively. The mean membrane potential of a population is converted into the average firing rate of all the neurons in that population using a sigmoid function S:

$$S(z) = \frac{\alpha_2}{1 + \exp\big(-r_2(z - V_2)\big)}, \qquad \text{for } z \in \mathbb{R}, \tag{4}$$

where α_2 is the maximum firing rate of the population, r_2 is the slope of the sigmoid and V_2 is the threshold of the population's mean membrane potential.

The neuronal populations are connected with connectivity strengths C_1 to C_7, which represents the average number of synaptic contacts between the neuronal populations concerned.

4. Neural mass models in state space form

Our observer design is most conveniently carried out using the state space form of the neural mass models. However, some of the neural mass models of interest presented in Section 3 were in block diagram form (e.g. Figure 6) in the given references [1, 2, 3]. In [2] and [3], state space forms were provided but not in the convenient state coordinates where the techniques we use for proving convergence of state estimates can be applied. Therefore, we illustrate how this can be done on the model by Wendling et. al., whose detailed block diagram can be found in Figure 6. The other models considered are special cases of the model by Wendling et. al. and hence, can easily be obtained from the derivation below.

We will show that all neural mass models from Section 3 can be written in the following state space form:

$$\dot{x} = Ax + G(\theta)\gamma(Hx) + \sigma(u, Cx, \theta), \tag{5}$$

and the output of the model is:

$$y = Cx, \tag{6}$$

where the state vector is $x \in \mathbb{R}^n$, input is $u \in \mathbb{R}$, output/EEG measurement is $y \in \mathbb{R}$, parameter vector is $\theta \in \mathbb{R}^p$, $G : \mathbb{R}^p \to \mathbb{R}^{n\times m}$, nonlinearity $\gamma = (\gamma_1, \ldots, \gamma_m)$ with $\gamma_i : \mathbb{R} \to \mathbb{R}$ for $i \in \{1, \ldots, m\}$ and nonlinearity $\sigma = (\sigma_1, \ldots, \sigma_n)$ with $\sigma_i : \mathbb{R}\times\mathbb{R}\times\mathbb{R}^m \to \mathbb{R}$ for $i \in \{1, \ldots, n\}$. The number of states n, number of parameters p and number of scalar nonlinear functions m differs for each model. These are defined in Sections 4.3-4.5.

4.1. Essential features

We highlight the following important essential features that our neural mass models possess. These are crucial for the applicability of the nonlinear observer proposed in Section 6 to the neural mass models we consider in this paper:

(i) The eigenvalues of the matrix A in (5) have strictly negative real parts.

(ii) The nonlinearity γ in (5) is globally Lipschitz [35, Theorem 3.2] i.e. there exists $\rho > 0$ such that $|\gamma(u)-\gamma(v)| \leq \rho|u-v|$ for u, $v \in \mathbb{R}$. Moreover, it is also bounded: $\gamma(z) \leq \alpha$, for all $z \in \mathbb{R}$. For instance, this nonlinearity γ can take the form of (4), which satisfies both properties.

(iii) The nonlinearity σ in (5) is bounded, i.e. for $i \in \{1, \ldots, n\}$, there exists $M > 0$ such that $|\sigma_i(z)| \leq M$, for all $z \in \mathbb{R}$.

The subsystems x_i from (10) for $i \in \{1, \ldots, 7\}$ that make up the full model (5) are interconnected in a unique manner as shown in Figure 6. This interconnection leads to a 'cascaded' structure of the error dynamics system that is crucial in the convergence proofs stated in Appendix A. This cascaded structure will be introduced in Appendix A.

4.2. Physiological interpretation

Physiologically, the first term in (5) implements the postsynaptic potential (PSP) kernels from (1), (2) and (3). This is effectively a convolution of the pre-synaptic firing rates arriving from other populations with the appropriate PSP response functions. These firing rates are modelled in the second and third term in (5) that incorporates

the sigmoid firing rate function of the depolarisation of contributing populations. The second term, $G(\theta)\gamma(Hx)$, reflects the influence of all states except the membrane potential of pyramidal population Cx. While the third term, $\sigma(u, Cx, \theta)$, reflects the influence of the mean membrane potential of the pyramidal cells Cx and the exogenous input u.

In the following sections, we present the state space form for each model for ease of observer design in Section 6. Detailed derivations are first shown for the model by Wendling et. al., as it is the most complex model and then the subtle differences in derivations are described for the other models.

4.3. State space form for the model by Wendling et. al.

We write the Wendling et. al. model in state space form by introducing the state variables x_{i1} for $i \in \{1, \ldots, 7\}$ as the membrane potential contribution from one population to another and x_{i2} for $i \in \{1, \ldots, 7\}$ as its derivative. The states x_{i1} is introduced at the outputs of all the impulse responses h_e, h_i and h_g blocks as shown in Figure 6. Recalling that the Laplace transform of the impulse responses h_e, h_i and h_g (as described by (1), (2) and (3)) are second order transfer functions, by performing the inverse Laplace transform, each transfer function is represented by a second order ordinary differential equation (ODE). We show this transformation for h_e from (1) as an example. Let the input to the h_e block be $\bar{u}$ and output be $\bar{y}$. We denote the Laplace transform of signal v as $\mathcal{L}(v)$. Hence, the Laplace transform of h_e is:

$$\mathcal{L}(h_e(t)) = \mathcal{L}(\theta_A a t \exp(-at)) = \frac{\theta_A a}{(s+a)^2}. \tag{7}$$

Recalling that $\mathcal{L}(h_e) = \frac{\mathcal{L}(\bar{y})}{\mathcal{L}(\bar{u})}$, we obtain:

$$\mathcal{L}(\bar{y})s^2 + 2a\mathcal{L}(\bar{y})s + a^2\mathcal{L}(\bar{y}) = \theta_A a \mathcal{L}(\bar{u}). \tag{8}$$

By taking the inverse Laplace transform, we obtain a second order ODE as follows:

$$\ddot{\bar{y}} + 2a\dot{\bar{y}} + a^2\bar{y} = \theta_A a \bar{u}. \tag{9}$$

The x_{i2} states are defined as $x_{i2} = \dot{x}_{i1}$ for $i \in \{1, \ldots, 7\}$ to rewrite the second order ODEs as two first order ODEs for each impulse response block.

We illustrate this for the $h_e(t)$ block in the fast inhibitory population, then the output of that block is $\bar{y} = x_{51}$ and the input is $\bar{u} = C_3 S(x_{11} - x_{21} - x_{31})$. Taking $x_{52} = \dot{x}_{51}$, (9) can be written as two first order ODE as follows:

$$\begin{aligned}
\dot{x}_{51} &= x_{52} \\
\dot{x}_{52} &= -2ax_{52} - a^2 x_{51} + \theta_A a C_3 S(x_{11} - x_{21} - x_{31}).
\end{aligned}$$

Hence, each impulse response h_e, h_i and h_g will each introduce a first order ODE in the following general state space form by taking $x_i = [x_{i1},\, x_{i2}]^T$ for $i \in \{1, \ldots, 7\}$:

$$\dot{x}_i = A_i x_i + \left[0, \vartheta_i S(\mu_i) + \varphi_i\right]^T, \tag{10}$$

where μ_i is the input to the respective sigmoid functions, $A_i = \begin{bmatrix} 0 & 1 \\ -k_{i1}k_{i2} & -(k_{i1}+k_{i2}) \end{bmatrix}$, for $i = \{1, \ldots, 7\}$ with $k_{11} = k_{41} = k_{51} = k_{61} = a$, $k_{12} = k_{42} = k_{52} = k_{62} = a$, $k_{21} = k_{71} = b$, $k_{22} = k_{72} = b$, $k_{31} = g$ and $k_{32} = g$. ϑ_i and φ_i are defined as such $\vartheta_1 = \theta_A a C_2$, $\vartheta_2 = \theta_B b C_4$, $\vartheta_3 = \theta_G g C_7$, $\vartheta_7 = \theta_B b C_6$, $\vartheta_4 = \vartheta_5 = \vartheta_6 = 0$ and $\varphi_1 = \theta_A a u$, $\varphi_2 = \varphi_3 = \varphi_7 = 0$, $\varphi_4 = \theta_A a C_1 S(y)$, $\varphi_5 = \theta_A a C_3 S(y)$, $\varphi_6 = \theta_A a C_5 S(y)$. Constants a, b and g are strictly positive. S

is a sigmoid function described by (4). All constants discussed in this section are summarised in Appendix C.

The subsystems defined in (10) are put together to be written compactly in state space form (5)-(6) for ease of observer design in Section 6.

We take the state vector in (5) and (6) to be $x = [x_1, \ldots, x_7]^T$ where x_i for $i = \{1, \ldots, 7\}$ satisfy (10). The states x_1, x_2 and x_3 capture the membrane potential contribution and its derivative of the excitatory, slow and fast inhibitory populations to the pyramidal neurons respectively. The states x_4, x_5 and x_6 capture the membrane potential contribution and its derivative of the pyramidal neurons to the excitatory, slow and fast inhibitory populations respectively. The output is $y = x_{11} - x_{21} - x_{31}$. The specific matrices in (5) and (6) are denoted as:

- The parameter vector is $\theta = [\theta_A, \theta_B, \theta_G]^T$,
- The matrix $A = \text{diag}(A_1, \ldots, A_7)$,
- $\gamma = [S, S, S]^T$ from (4), (11)
- $\sigma = (0, \theta_A a u, 0, 0, 0, 0, 0, \theta_A a C_1 S(y), 0, \theta_A a C_3 S(y), 0, \theta_A a C_5 S(y), 0, 0)$, where S is described by (4),
- $C = [\ 1 \quad 0 \quad -1 \quad 0 \quad -1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0\ ]$, (12)
- $G = \begin{bmatrix} 0 & 0 & 0 \\ \theta_A a C_2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \theta_B b C_4 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \theta_G g C_7 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \theta_B b C_6 & 0 \end{bmatrix}$ (13)

  and
- $H = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 \end{bmatrix}$.

4.4. State space form for the model by Jansen and Rit in [2]

We write the model in state space form by taking the state vector in (5) to be $x = [x_1, x_2, x_4, x_5]^T$, where x_i for $i = \{1, 2, 4, 5\}$ satisfy (10). States x_1 and x_2 are respectively, the membrane potential contribution and its derivative of the excitatory and inhibitory populations to the pyramidal neurons. States x_4 and x_5 capture the membrane potential contribution and its derivative of the pyramidal neurons to the excitatory and inhibitory populations respectively. The output is $y = x_{11} - x_{21}$. The specific matrices in (5) and (6) are denoted as:

- The parameter vector is $\theta = [\theta_A, \theta_B, C_1, C_2, C_3, C_4]^T$,
- $A = \text{diag}(A_1, A_2, A_4, A_5)$,

- $\gamma = [S, S]^T$ where S is defined in (4), (14)
- $\sigma = [0, \theta_A a u, 0, 0, 0, \theta_A a C_1 S(y), 0,\ \theta_A a C_3 S(y)]^T$,
- $C = \begin{bmatrix} 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$, (15)
- $G = \begin{bmatrix} 0 & 0 \\ \theta_A a C_2 & 0 \\ 0 & 0 \\ 0 & \theta_B b C_4 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$ (16)

  and
- $H = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$.

4.5. State space form for the model by Stam et. al. in [1]

The model is written in state space form by taking the state vector in (5) as $x = [x_1, x_2, x_5]^T$ where x_i for $i = \{1, 2, 5\}$ satisfy (10). State x_1 represents the contribution and its derivative of the excitatory population's activity to itself. States x_2 and x_5 represent the contribution and its derivative of the inhibitory population to the excitatory population and vice versa respectively. The output is $y = x_{11} - x_{21}$.

This model differs from the models by Wendling et. al. and Jansen and Rit in the sense that the firing rate of the population is converted to postsynaptic potential via different kernels from (1) and (2), for $t \geq 0$:

- Excitatory population:
$$h_e(t) = \theta_A[\exp(-a_1 t) - \exp(-a_2 t)]. \tag{17}$$
- Inhibitory population:
$$h_i(t) = \theta_B[\exp(-b_1 t) - \exp(-b_2 t)]. \tag{18}$$

As performed in Section 4.3, by taking Laplace transformations of (17) and (18) and taking the inverse Laplace transform, the kernels can be written as second order ODEs. They can then be rewritten as two first order ODEs by introducing extra state variables, x_{i2} for $i \in \{1, 2, 5\}$, in a similar fashion as in Section 4.3.

Also, the sigmoid function that converts the postsynaptic potential to the firing rate of the population differs from (4) for the models by Wendling et. al. and Jansen and Rit, as follows:

$$S_1(z) = \begin{cases} \alpha_1 \exp\big(r_1(z - V_1)\big) & z \leq V_1, \\ \alpha_1\Big(2 - \exp\big(-r_1(z - V_1)\big)\Big) & z > V_1, \end{cases} \tag{19}$$

where α_1 is the maximum firing rates of the population, r_1 is the slope of the sigmoid and V_1 is the threshold of the population's mean membrane potential.

Nevertheless, rewriting (17)-(19) into state space form does not differ from the derivation presented in Section 4. The specific matrices in (5) are denoted as:

- The parameter vector is $\theta = [C_3, C_4]^T$,
- $A = \mathrm{diag}(A_1, A_2, A_5)$,

- $\gamma = S_1$ from (19), which still satisfies item (ii) in Section 4.1. (20)
- $\sigma = [0, \theta_A(a_2 - a_1)u, 0, 0, 0, \theta_A(a_2 - a_1)C_3 S_1(y)]^T$,
- $C = [\ 1 \quad 0 \quad -1 \quad 0 \quad 0 \quad 0\ ]$, (21)
- $G = [0, 0, 0, \theta_B(b_2 - b_1)C_4, 0, 0]^T$ (22)
  and
- $H = [\ 0 \quad 0 \quad 0 \quad 0 \quad 1 \quad 0\ ]$.

5. Assumptions

Our aim is to design a class of nonlinear observers for the neural mass models considered in Section 3, such that the state estimates converge to the true states in exponential time, for any initial condition and any input. We first present the idealised assumptions in Section 5.1 that are used in Section 6.1. We then state the relaxed assumptions in Section 5.2 and we show the robustness of the observers in Section 6.2 under these relaxed assumptions.

5.1. Assumptions under the ideal scenario

The convergence of the state estimates provided by our observers are proven in Section 6.1 under the idealised assumptions as illustrated in Figure 2 and stated as follows:

Assumption 1. *The synaptic gain of each neuronal population $\theta_A, \theta_B, \theta_G$ and the connectivity strengths C_1, C_2, C_3, C_4 are parameters that are constant and known.*

These parameters are typically slowly-varying during a particular brain activity, such that they can be considered constant over the time period observed [35, Section 9.6]. When the brain transitions from one activity to another however, the parameters C_3, C_4 for the model by Stam et. al. in [1], parameters θ_A, θ_B, C_1, C_2, C_3 and C_4 in for the model by Jansen and Rit in [2] and parameters θ_A, θ_B and θ_G for the model by Wendling et. al. in [3] will change. If the goal is solely to estimate the parameters, system identification methods such as least-squares estimation may be used [36]. However, it remains a hard problem due to the nonlinearity of these models. For this study, we only consider the case where the brain is in one particular brain activity and that the parameters are known, that is for the case where the parameters have been identified a priori. We analyse the robustness of the observer to parameter uncertainty in Section 6.2.

Assumption 2. *The input u is measured.*

The input from afferent populations is hard to quantify in practice, hence this assumption is not justified in general. In Section 6.2, we relax this assumption by allowing the input to be uncertain. We show that good estimates can still be obtained provided that the L_∞ norm of the difference between the true and assumed input is small. This is further illustrated with simulations in Section 7.2.2. It is important to note that it is not an easy task to guarantee exponential convergence of the error estimates without assuming that the input is known exactly. Please refer to [37] and references therein.

Assumption 3. *The measured EEG y is noise-free.*

The measured EEG considered here is recorded using intracranial electrodes, which we first assume to be clean data such that it can be considered noise-free. We will show later in Section 6.2 that the estimation error remains bounded despite noisy measurements. Assuming noise-free measurements makes it easier to design a deterministic observer and it is the first step to explain our methodology; we later relax our assumptions and show that the designed observer is robust to measurement noise.

Assumption 4. *No external influences from other neuronal populations.*

All the neuronal populations except for the pyramidal neurons or the excitatory population receiving the input u included in the models in Section 3 are not affected by adjacent populations. In Section 6.2, we will introduce the external influences as an additive disturbance and show that the estimation error remains bounded with bounded disturbance.

5.2. Relaxed assumptions

The assumptions set out in Section 5.1 are strong, but provide us with a good first step in proving the estimation error convergence of the observers. We now relax these assumptions to characterise their effects in Section 6.2 on the convergence property of the class of observers we propose in Section 6. The relaxed conditions are captured in Figure 7 and we restate our relaxed assumptions as follows:

Assumption 5 (Relaxation of Assumption 1)**.** *The parameters θ_A, θ_B, θ_G, C_1, C_2, C_3 and C_4 of the model (5) are known with error.*

We characterise these uncertainties in parameters by introducing bounded, time-varying signal $\epsilon_\theta(t)$ to the parameters of the model (5) as shown in (26), so that the true parameter values in the model become $\theta + \epsilon_\theta$.

Assumption 6 (Relaxation of Assumption 2)**.** *The input is uncertain.*

We introduce bounded, time-varying disturbances $\epsilon_u(t)$ to the input available to the observer as shown in (27). Thus, the real input to (5) is modelled by $u + \epsilon_u$, whereas the observer is fed by an assumed input u. The introduction of ϵ_u relaxes Assumption 2 by allowing the real input to be unknown to the observer.

Assumption 7 (Relaxation of Assumption 3)**.** *The measured EEG is affected by measurement noise.*

Disturbance ϵ_y is introduced to characterise measurement noise. The measurement with noise is denoted as $y + \epsilon_y$.

Assumption 8 (Relaxation of Assumption 4)**.** *The neural mass model is inaccurate.*

Additive disturbance $\epsilon_{sys}(t)$ characterises the possible influence from other populations to the neuronal populations included in the model.

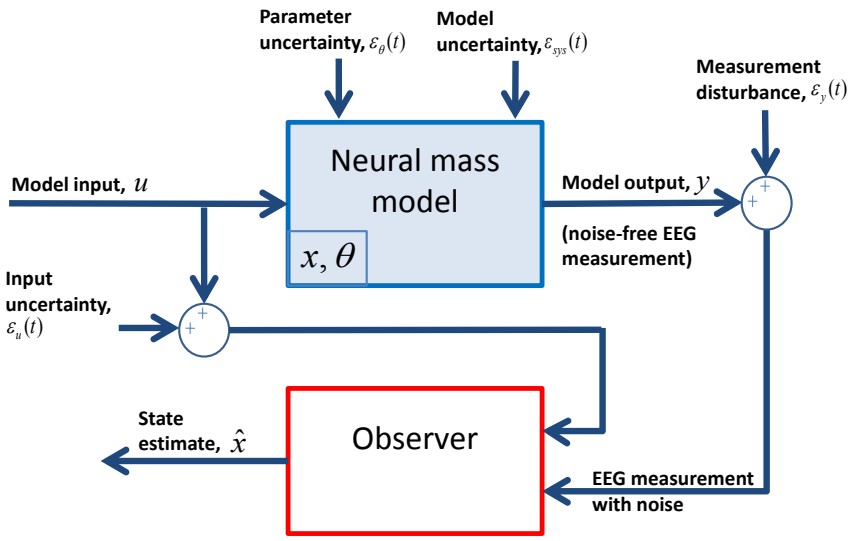

Figure 7. Model and observer setup under the relaxed assumptions.

6. Main results

We first propose a class of nonlinear observers in Theorem 1 under the ideal Assumptions 1-4. The ideal assumptions are later relaxed as stated in Assumptions 5-8. Theorem 2 uses the relaxed assumptions and show the robustness of the designed observers against disturbances and uncertainties.

6.1. Observer design under the ideal scenario

We propose the following class of observers:

$$\dot{\hat{x}} = A\hat{x} + G(\theta)\gamma(H\hat{x} + K(C\hat{x} - y)) + L(C\hat{x} - y) + \sigma(u, y, \theta), \qquad (23)$$

where $\hat{x} \in \mathbb{R}^n$ is the state estimate and observer matrices $K \in \mathbb{R}^m$, $L \in \mathbb{R}^n$ are introduced via linear output injection terms $K(C\hat{x} - y)$ and $L(C\hat{x} - y)$.

The observer matrices L and K in (23) are the choice of the user. They are introduced to provide flexibility over the convergence rate of the estimates. We establish in Theorem 1 that for sufficiently small norms of matrices K and L, we obtain convergence of the estimates to the true states in exponential time, for any initial conditions.

Theorem 1. *Consider the model in general form (5) under Assumptions 1 to 4 and observer matrices K and L for (23) are chosen such that:*

$$\rho|K||G| + |L| < \nu\frac{1}{|C|}, \qquad (24)$$

where ρ is the Lipschitz constant of γ, matrices G and C are from the model (5) and $\nu > 0$ is constructed in the proof.

Specifically, γ is (11) for the model by Wendling et. al., (14) for the model by Jansen and Rit and (20) for the model by Stam et. al. Also, G is (13) and C is (12)

for the model by Wendling et. al., G is (16) and C is (15) for the model by Jansen and Rit, and G is (22) and C is (21) for the model by Stam et. al..

Then, the observer (23) is a global exponential observer for the model (5), i.e. for any $u \in L_\infty$ and denoting the estimation error as $e := x - \hat{x}$,

$$|e(t)| \leq k \exp(-\lambda t)|e(0)| \qquad \forall t \geq 0, \forall e(0) \in \mathbb{R}^n, \tag{25}$$

where constants k, $\lambda > 0$.

Proof. The proof is given in Appendix A. □

Remark 1. *Note that Theorem 1 applies to all neural mass models considered in Section 3, where matrices G and C are different for each model considered. Consequently, the obtained observer matrices K and L as well as ν from (24) and k, λ from (25) will differ for each model.*

The exponential convergence property of the observer is desirable in practice because it means that the estimates are guaranteed to converge to the true states in exponential time. Additionally, the validity of the convergence property for any initial conditions is highly desirable because the initial conditions of the neuronal populations are usually unknown.

Theorem 1 provides us with a condition (24) that gives a class of nonlinear observers parameterised by observer matrices K and L. This condition (24) is conservatively obtained in Appendix A. However, we see that the choice of $L = 0$ and $K = 0$ fulfils condition (24) and therefore, admits an open-loop observer for the model (5) [38]. The drawback of an open-loop observer lies in that the convergence speed of the state estimation error cannot be controlled by the user. Nevertheless, it remains a useful observer provided that the error converges sufficiently fast. In Section 7.1, we see in simulations that the open-loop observer for the model by Wendling et. al. [3] provides estimates that converge to the true states in a reasonable timeframe. Non-zero choices of K and L fulfilling condition (24) provide tuneability of the estimation error's convergence speed. However, condition (24) does not provide a priori information about the convergence rate of the observer. In Section 7.1, we show in simulations that some choices of non-zero matrices K and L that satisfy condition (24) can lead to faster convergence rate for the state estimation error.

Remark 2. *The observer structure (23) shares the same mathematical structure as other nonlinear observers in the literature, that is, the high gain [39] (with $K = 0$) and circle criterion observers [40]. These observers employ special techniques in obtaining observer matrices L and K that are not satisfied by the model we consider (5). Therefore, we prove the existence of K and L matrices using a different analysis from that in [39, 40].*

6.2. Robustness Analysis

We show in Theorem 2 that under Assumptions 5-8, i.e. uncertainties in modelling, parameters, measurement and input, the estimates provided are close to the true states, where the 'closeness' is determined by the L_∞ norm of the uncertainties ϵ_y, ϵ_θ, ϵ_u and ϵ_{sys}. The introduction of the uncertainties and disturbances to the ideal setup is illustrated in Figure 7.

Therefore, we obtain the following perturbed systems (26) and (27) from (5) and (23) as illustrated in Figure 7:
From (5):

$$\begin{aligned}\dot{x} &= Ax + G(\theta+\epsilon_\theta)\gamma(Hx) + \sigma(u,\, Cx,\, \theta+\epsilon_\theta) + \epsilon_{sys},\\ y &= Cx.\end{aligned} \tag{26}$$

From (23):

$$\begin{aligned}\dot{\hat{x}} = {} & A\hat{x} + G(\theta)\gamma\big(H\hat{x} + K(C\hat{x} - (y+\epsilon_y))\big)\\ & + L(C\hat{x} - (y+\epsilon_y)) + \sigma(u+\epsilon_u,\, y+\epsilon_y,\, \theta).\end{aligned} \tag{27}$$

Theorem 2. *Consider the perturbed estimation error system (26) and (27) under Assumptions 5 to 8. The observer (27) with observer matrices K and L are chosen such that (24) is satisfied.*

Specifically, γ is (11) for the model by Wendling et. al., (14) for the model by Jansen and Rit and (20) for the model by Stam et. al. Also, G is (13) and C is (12) for the model by Wendling et. al., G is (16) and C is (15) for the model by Jansen and Rit, and G is (22) and C is (21) for the model by Stam et. al..

This guarantees that for all $e(0) \in \mathbb{R}^n$, $t \geq 0$ and for all u, ϵ_y, ϵ_θ, ϵ_u and $\epsilon_{sys} \in L_\infty$, the error system satisfies the following:

$$|e(t)| \leq \tilde{k}\exp(-\tilde{\lambda}t)|e(0)| + \gamma_y(\|\epsilon_y\|_{[0,t]}) + \gamma_\theta(\|\epsilon_\theta\|_{[0,t]}) + \gamma_u(\|\epsilon_u\|_{[0,t]}) + \gamma_{sys}(\|\epsilon_{sys}\|_{[0,t]}),$$

where $\tilde{k}$, $\tilde{\lambda} > 0$ and γ_y, γ_θ, γ_u, $\gamma_{sys} : \mathbb{R} \to \mathbb{R}$ are continuous positive increasing functions that are zero at the origin.

Proof. The proof for Theorem 2 is provided in Appendix B. □

We see that with no uncertainties and disturbance, i.e. $\epsilon_u = 0$, $\epsilon_\theta = 0$ and $\epsilon_y = 0$, we recover the results of Theorem 1. Loosely speaking, Theorem 2 states that small uncertainties and disturbance implies small estimation error.

Remark 3. *By setting $\epsilon_u = -u$, we have that the input to the observer (23) is 0, that is we consider the proposed observer (23) as an unknown-input observer. The estimation error of the proposed observer (23) converges with some error that depends upon the L_∞ norm of ϵ_u. This result is advantageous because the mean firing rate of the afferent neuronal populations to a particular brain region is hard to measure in reality. We will further demonstrate this observation in Section 7.2.2 via simulations.*

7. Simulation Results

We illustrate the performance of our observers using simulations. First, we perform simulations to show the convergence of the estimates under the ideal conditions stated in Section 5 and test the influence of the observation gains on the speed of convergence of the estimation error. Next, we show the robustness of our observers against uncertainties in modelling, parameters, input and measurement in Section 7.2.1. We also show that our observers can provide estimates that are close to to the true states when the input to the observer is unknown in Section 7.2.2. The simulation results provided are for the model of the hippocampus by Wendling et. al. in [3]. Similar results can be obtained for other models introduced in Section 3.

We select synaptic gains $\theta_A = 5$, $\theta_B = 25$ and $\theta_G = 10$ that correspond to seizure activity as identified in [3]. The initial condition of the model (5) is set to

$x(0) = [6, 0.5, 6, 0.5, 6, 0.5, 6, 0.5, 6, 0.5, 6, 0.5, 6, 0.5]^T$. We initialise the observer (23) to $\hat{x}(0) = 0$. This choice is arbitrary as we have shown that the convergence of the estimation error is valid for all initial conditions in Theorems 1 and 2.

In this section, we denote the observer matrices as $K = k \times I_k$ and $L = l \times I_l$ where $I_k = [1, 1, 1]^T$ and $I_l = [1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]^T$. We call k and l the observer gains. For each of the scenarios in the following sections, the performance of two observers are evaluated:

(i) The open-loop observer $l = 0$, $k = 0$.

(ii) The observer with observer gains $k = 0.1$ and $l = -0.2$.

All other constants used in simulation are as specified in Appendix C.

7.1. Simulations under Assumptions 1 to 4: ideal scenario

We consider the ideal case where the parameters are constant and known, input u and measurement y are known and unperturbed as well as no modelling errors. Both model (5) and observer (23) are supplied with the same Gaussian noise input with mean 90mV and variance 30mV used in [3].

As stated in Theorem 1 and illustrated in Figures 8 and 9, the state estimation error $e := x - \hat{x}$ converges to 0 asymptotically in exponential time. We observe in simulations that for both observers, the convergence rate is in general faster than the duration of a specific brain activity (see [3]): by $t = 0.3s$, all state estimation errors have converged to 0. As seen in Figure 8, the plot of the error norm shows a large overshoot initially (approximately 1200). However, the plots on Figures 9 show that the estimates of the membrane potential contributions from one population to another $\hat{x}_{i1}$ are reasonably close to the true values x_{i1}. The large initial error is due to the estimation error in the other states, i.e. the error of the time derivative of the membrane potential contributions $|x_{i2} - \hat{x}_{i2}|$, which are not the physical quantities of concern in practice.

As discussed in Section 6, the choice of non-zero observer gains provides flexibility over the convergence rate of the estimation error. Figure 8 shows that for a selection of observer gains k and l, the norm of the estimation error converges faster than with the case of $k = l = 0$ (open-loop observer).

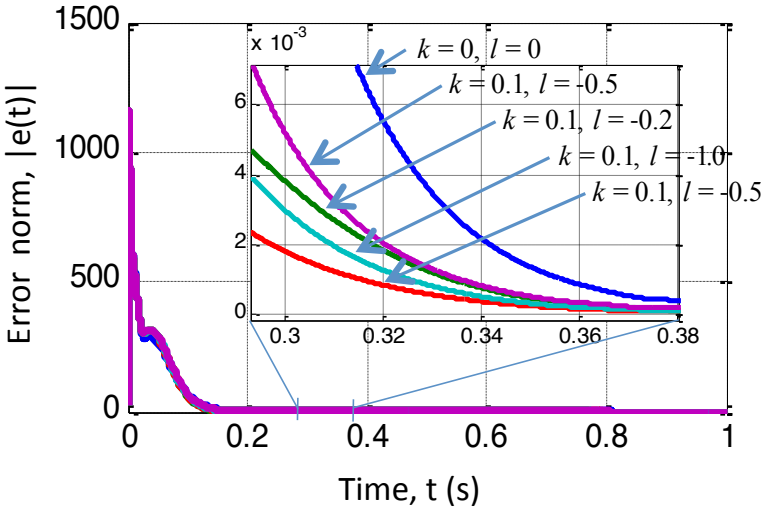

Figure 8. Norm of error over time for a selection of K and L.

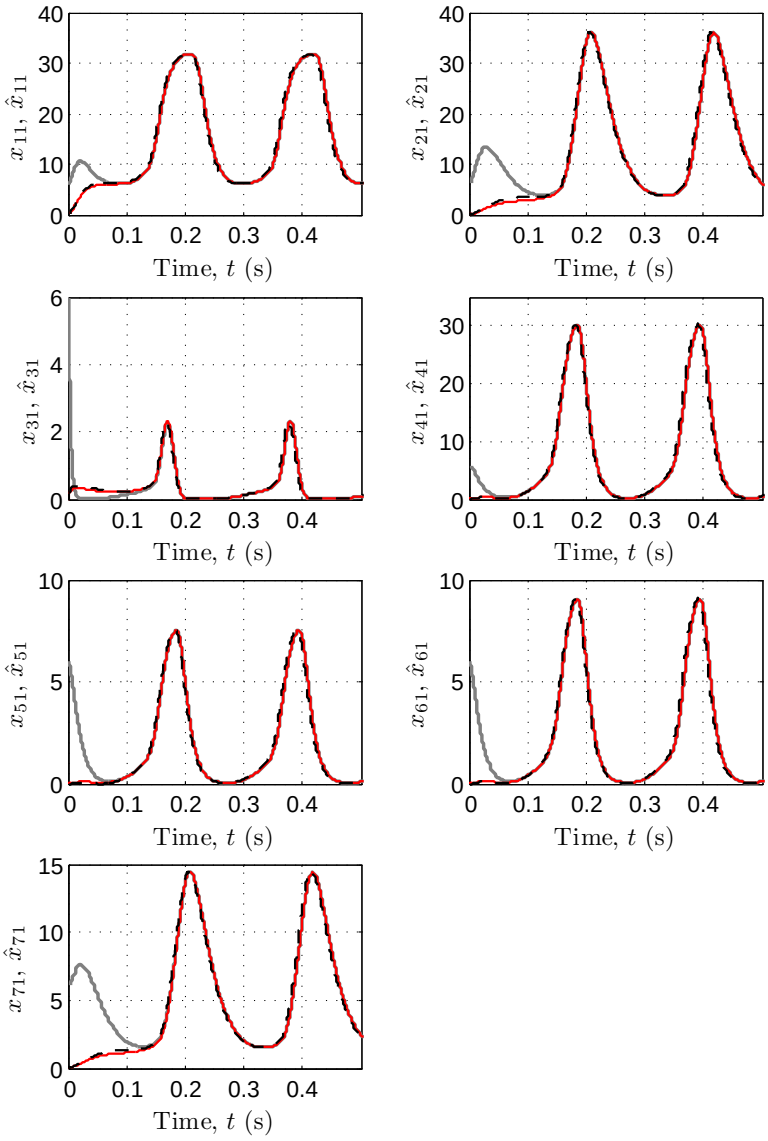

Figure 9. Membrane potential contribution x_{i1} (grey solid line) and estimated membrane potential contribution $\hat{x}_{i1}$ (red solid line: $l = k = 0$, black dashed line: $k = 0.1$, $l = -0.5$) under the ideal scenario (Assumptions 1 to 4).

7.2. Simulations for the practical scenario

7.2.1. Simulations under Assumptions 5 to 8: uncertainty in parameters, input and measurement as well as additive disturbance.

Next, we test the performance of the observers under uncertainties in the parameters θ_A, θ_B, and θ_G, disturbances in the input and output of the observer, as well as additive disturbance.

We simulate the perturbed systems (26) and (27) with the perturbations: constant parameter uncertainty $\epsilon_\theta = (0.5, 2.5, 1)$, Gaussian input uncertainty $\epsilon_u \sim \mathcal{N}(0, 0.3^2)$, Gaussian measurement noise $\epsilon_y \sim \mathcal{N}(0, 0.1^2)$ and Gaussian model uncertainty $\epsilon_{sys} \sim \mathcal{N}(0, 1^2\mathbb{I})$. The performance of both observers are similar. Figures 10 and 11 show that the estimation error converges to a neighbourhood of the origin. This observation agrees with Theorem 2.

Figure 10 shows that the estimates for the states of interest x_{i1} do converge reasonably close to the true states. It can be seen in Figure 11 that the relative error of the states is quite low (under 10%).

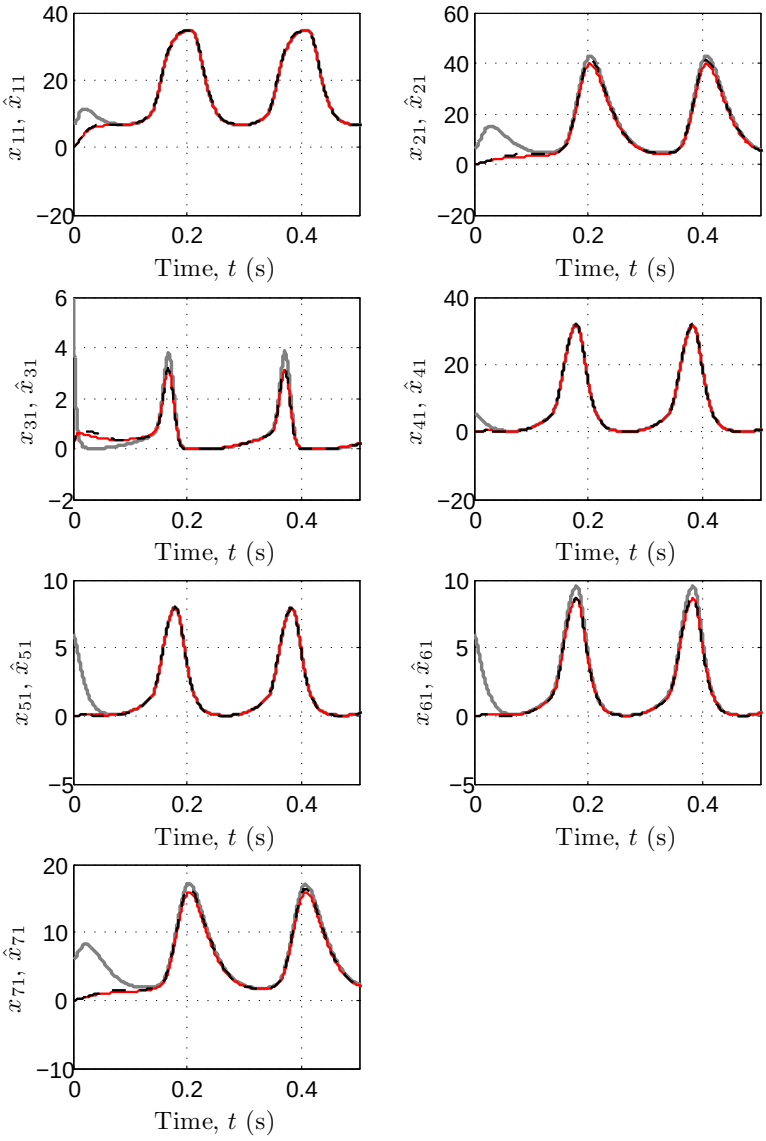

Figure 10. True membrane potential contribution x_{i1} (grey solid line) and estimated membrane potential contribution $\hat{x}_{i1}$ (red solid line: $l = k = 0$, black dashed line: $k = 0.1$, $l = -0.5$) under the practical scenario (Assumptions 5 to 8).

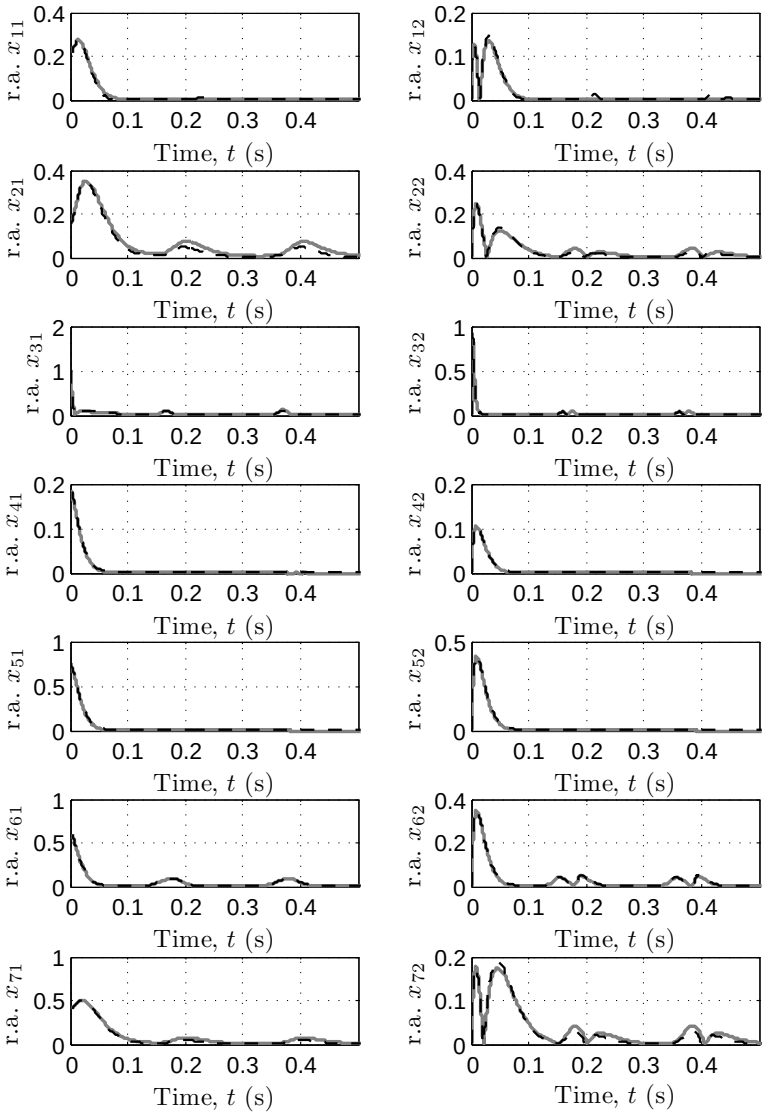

Figure 11. Relative state estimation error $\frac{|x_i - \hat{x}_i|}{\max(x_i) - \min(x_i)}$ for $i \in \{1, \ldots, 7\}$ under the practical scenario (Assumptions 5 to 8) for the open-loop observer (grey solid line) and the observer with $k = 0.1$, $l = -0.5$ (black dashed line).

7.2.2. Simulations under Assumptions 1, 3, 4 and 6: uncertain-input observer.
The input to the brain is usually hard to quantify under realistic conditions. To gauge the performance of the observers under this condition, we perform another set of simulations where the input is set to $u = 0$ for the observers with no uncertainty in parameters and measurement noise. This case is formally stated in Remark 3.

For both observers, Figures 13 to 12 illustrate that the estimation error converges to a neighbourhood of the origin as expected. In Figure 13, the estimation error for subsystems $(x_{21},\ x_{22})$, $(x_{31},\ x_{32})$, $(x_{41},\ x_{42})$, $(x_{51},\ x_{52})$, $(x_{61},\ x_{62})$ and $(x_{71},\ x_{72})$ converges to a reasonably small neighbourhood of the origin. However, the $(x_{11},\ x_{12})$ subsystem exhibits a much larger steady state estimation error than the other subsystems due to the input u directly affecting it. Nevertheless, our simulations in Figure 14 show that the states can be reasonably reconstructed when the input is unknown.

In Figure 12, the open-loop observer outperforms the observer with $k = 0.1$ and $l = -0.5$, in the sense that the open-loop observer converges to a smaller neighbourhood around the origin $e = 0$ (with $|e| \leq 35$ for the open-loop observer and $|e| \leq 85$ for the observer with $l = -0.5$ and $k = 0.1$). A decision regarding the tradeoff between the convergence speed of the state estimation error and the accuracy of the estimates needs to be made when employing these observers. In this simulation scenario, the error converges to a neighbourhood in a reasonable timeframe ($t = 0.2$s) for both observers. Hence, the open-loop observer is the observer of choice due to the steady-state accuracy of its estimates.

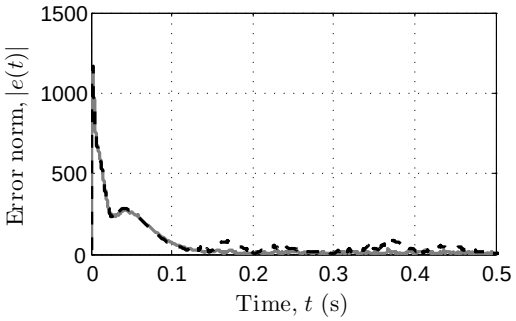

Figure 12. Norm of state estimation error $|e|$ for the open-loop observer (grey solid line) and the observer with $k = 0.1$, $l = -0.5$ (black dashed line) under Assumptions 1, 3, 4 and 6.

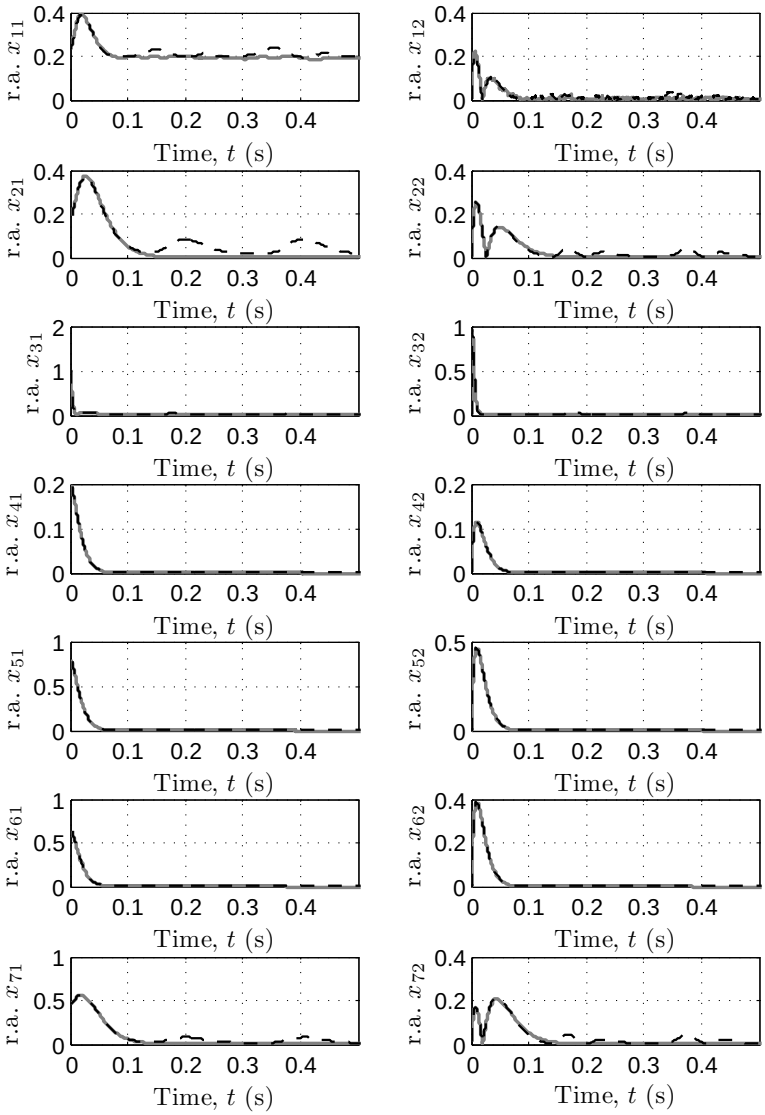

Figure 13. Absolute state estimation error $\frac{|x_i - \hat{x}_i|}{\max(x_i) - \min(x_i)}$ for $i \in \{1, \ldots, 7\}$ for the open-loop observer (grey solid line) and the observer with $k = 0.1$, $l = -0.5$ (black dashed line) under Assumptions 1, 3, 4 and 6.

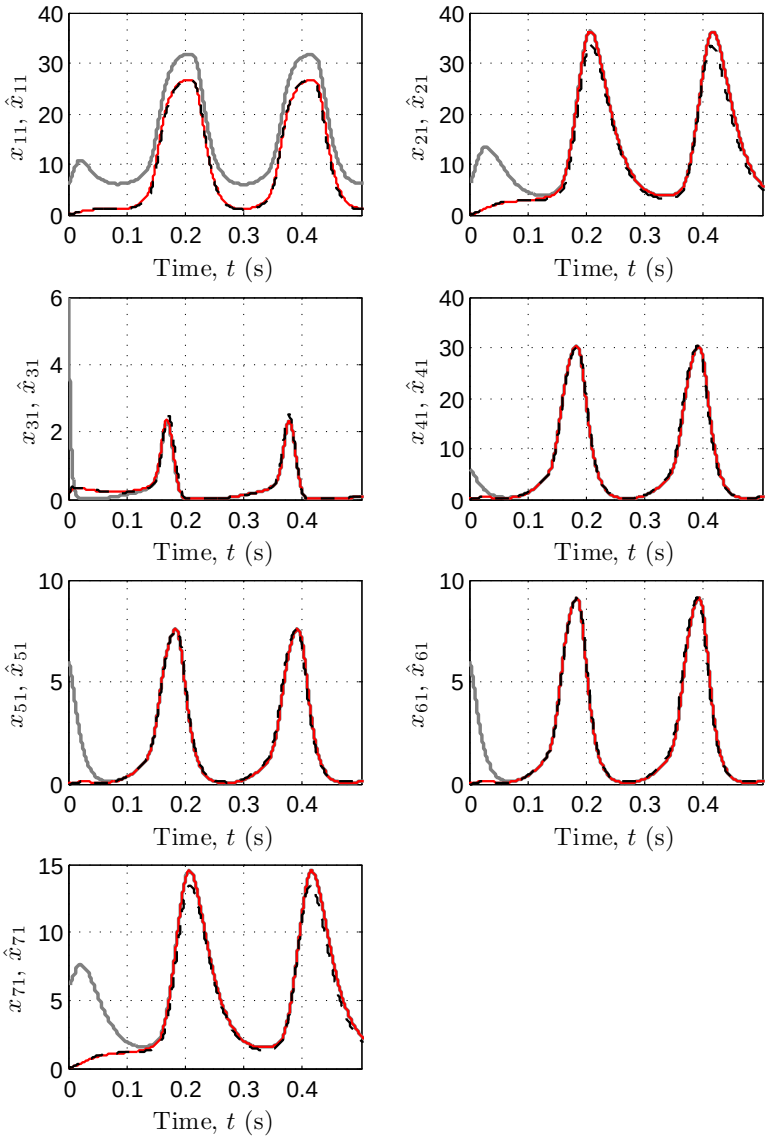

Figure 14. True membrane potential contribution x_{i1} (grey solid line) and estimated membrane potential contribution $\hat{x}_{i1}$ (red solid line: $k = l = 0$, black dashed line: $k = 0.1$, $l = -0.5$) under Assumptions 1, 3, 4 and 6.

8. Discussion and conclusion

Using a single channel EEG signal as the measurement of brain activity, we have successfully developed a class of nonlinear observers that can reconstruct the mean membrane potential of each neuronal population for several classes of neural mass models. We rewrote the neural mass models considered in suitable state coordinates and identified the essential model features that allowed us to prove that the proposed class of observers gives estimates that converge exponentially to the true states of the model, as stated in Theorem 1. Moreover, the global convergence property of our proposed observers ensures that any initial estimated state will converge to the true state of the model for any input. This result is desirable because we do not know the initial state of the brain. Limitations of our proposed class of observers lie in the requirement that parameters and input to the observers need to be known, both of which are not known in a clinical or experimental setting. However, we have shown in Theorem 2 that our observer is robust to uncertainties in parameters as well as disturbances in the model, input and EEG measurement, in the sense that bounded uncertainty will result in bounded estimation error.

Given that it is difficult to measure the actual mean membrane potential of each neuronal population, the estimates obtained serve to provide some understanding of the underlying dynamics during a specific brain activity. This is beneficial for the general EEG-based studies of sensory, motor and cognitive processing and can lead to the development of clinical treatment methods for brain disorders. One such example is in the treatment of epilepsy, where more positive outcomes of therapy may be achieved by targeting specific seizure-causing mechanisms.

So far, we have validated the efficiency of our observer analytically and further illustrated the results in simulation. The applicability of our observer to real EEG data will depend heavily on how good an abstraction the models considered are of the brain region concerned. Provided that the models are a good description of the brain region of interest, the estimates of the unmeasured brain activity provided by the observers from real EEG measurement are then a good depiction of real activity in the brain. Given that the model by Wendling et. al. [3] considered in this paper has been validated against EEG recordings obtained from depth electrodes implanted in the hippocampus of a temporal lobe epilepsy patient, we expect the estimates provided by our observer to be a good depiction of the underlying dynamics of a patient's epileptic brain. This is achieved under the assumptions stated in Section 5 and to a satisfactory degree under the more practical assumptions stated in Section 6.2.

It is important to note that neural mass models are 'lumped parameter' models that describe the properties of populations of neurons, as opposed to a single neuron. In this construct, neurons are assumed to be spatially close, so that their interconnections are dense enough such that the group of neurons can be considered as one population with an aggregated activity [41, Chapter 23]. Based on these assumptions, the model captures the temporal evolution of neuronal population dynamics and neglects the spatial interactions. As such, the interpretation of the estimates for the model describing epileptic seizures originating from the hippocampus [3] for example, applies only to the genesis of seizures in a local sense and not to the propagation of seizures from a focus. Understanding the seizure propagating mechanisms will require distributed models, also known as neural field models such as those in [42, 43, 44] (see [32] for a review). These distributed models are infinite-dimensional models that would require infinite-dimensional observers for the

estimation of states. Turning an infinite-dimensional model to a lower order model that is finite-dimensional would allow for the development of suboptimal finite-dimensional estimation methods, which are far easier to develop and practical to implement. Ultimately, the choice of the model rests in the aim of the endeavour.

An alternative to using spatially distributed or continuous models is to interconnect neural mass models, like those considered here, into spatial networks. Such an approach has been employed in dynamic causal modelling for EEG by coupling Jansen and Rit neural mass models [21]. Our deterministic observers apply to a single neural mass model, but in future research could be extended to apply to interconnected neural mass models, and therefore be applicable to the popular dynamic causal modelling framework.

One of the motivations behind this work is to provide early detection of focal seizures in epilepsy. It has been identified in [3] that varying the synaptic gains of each neuronal population θ_A, θ_B and θ_G leads to different EEG patterns. Our work is a starting point in tracking the transition of the brain from non-seizure to seizure activity. To meet this objective, the simultaneous estimation of both states and seizure-causing parameters will need to be performed. While estimating the parameters alone should achieve the aim of detecting seizure onsets, the additional estimation of states provides the added advantage of peeking into the hidden dynamics of each neuronal population's activity. The success in designing a class of nonlinear observers for a selection of neural mass models raises hope in tackling the challenging problem of detecting and predicting epileptic seizures. The approach presented in this paper is also applicable in general to neuroscientific studies, where estimating the unmeasured, or in some cases unmeasurable, activities of brain regions, from an EEG measurement is beneficial.

Appendix A. Proof of Theorem 1

We will do all proofs for the model by Wendling et. al. described in Section 3.3, which is the most general model that encompasses the models by Stam et. al. and Jansen et. al. described in Sections 3.1 and 3.2 respectively. The proof for all other models can be done by first identifying the cascade structure of the observation error system, then showing that the subsystems satisfy certain properties to conclude the global exponential stability of the whole observation error system. This is performed in a similar fashion in the proof for the Wendling et. al. model that follows.

From (5), (23), the dynamics of the state estimation error $e := x - \hat{x}$ is:

$$\dot{e} = Ae + G(\gamma(Hx) - \gamma(H\hat{x} + KCe)) + LCe. \tag{A.1}$$

The main idea is to consider the estimation error system (A.1) as the nominal error system (A.1) with $L = 0$ and $K = 0$ perturbed by the terms $G\gamma(H\hat{x}) - G\gamma(H\hat{x} + K(C\hat{x} - y)) + LCe$. As such, in Lemma 1, we build a Lyapunov function W for the nominal error system. Next, using W as a candidate Lyapunov function for (A.1), we obtain a bound for observer matrices K and L such that $e = 0$ is global exponential stable (GES) [35, Definition 4.5].

Lemma 1. *There exists a continuously differentiable $W : \mathbb{R}^n \to \mathbb{R}$ such that the following holds, for all $e \in \mathbb{R}^n$:*

$$\begin{array}{rcl} k_1|e|^2 & \leq & W(e) \quad \leq \quad k_2|e|^2, \\ \left|\frac{\partial W(e)}{\partial e}\right| & \leq & k_4|e|, \end{array}$$

and along solutions to (A.1) with $L = 0$ and $K = 0$:

$$\dot{W}(e) \quad \leq \quad -k_3|e|^2, \tag{A.2}$$

where k_1, k_2, k_3 and k_4 are strictly positive constants.

Proof. The nominal estimation error system (A.1) with $L = 0$ and $K = 0$ is

$$\dot{e} = Ae + G(\gamma(Hx) - \gamma(H\hat{x})). \tag{A.3}$$

First note that system (5) has solutions that are defined for all time due to the global Lipschitz property of function S and the fact that input $u \in L_\infty$ [35, Theorem 3.2]. In view of the same arguments, the solutions of system (23) (with y and u from system (5)) are also well-defined and exist for all time. Consequently, we have that (A.3) have solutions that are defined for all time.

We can decompose system (A.3) into 7 subsystems Σ_{e1} to Σ_{e7}:

$$\begin{array}{lrcl}
\Sigma_{e1}: & \dot{e}_1 & = & A_1e_1 + \phi_1(x_{41}) - \phi_1(\hat{x}_{41}) \\
\Sigma_{e2}: & \dot{e}_2 & = & A_2e_2 + \phi_2(x_{51}) - \phi_2(\hat{x}_{51}) \\
\Sigma_{e3}: & \dot{e}_3 & = & A_3e_3 + \phi_3(x_{61}, x_{71}) - \phi_3(\hat{x}_{61}, \hat{x}_{71}) \\
\Sigma_{e4}: & \dot{e}_4 & = & A_4e_4 \\
\Sigma_{e5}: & \dot{e}_5 & = & A_5e_5 \\
\Sigma_{e6}: & \dot{e}_6 & = & A_6e_6 \\
\Sigma_{e7}: & \dot{e}_7 & = & A_7e_7 + \phi_7(x_{51}) - \phi_7(\hat{x}_{51}),
\end{array} \tag{A.4}$$

where $e_i := (\ e_{i1}, e_{i2}\ ) = (\ x_{i1} - \hat{x}_{i1},\, x_{i2} - \hat{x}_{i2}\ ) \in \mathbb{R}^2$ for $i = \{1, \ldots, 7\}$. $\phi_1(x_{41}) = \big(0, \theta_A a C_2 S_2(x_{41})\big)$, $\phi_2(x_{51}) = \big(0, \theta_B b C_4 S_2(x_{51})\big)$, $\phi_3(x_{61}, x_{71}) = \big(0, \theta_G g C_7 S_2(x_{61} - x_{71})\big)$, $\phi_7(x_{51}) = (0, \theta_B b C_6 S_2(x_{51}))$. Matrices $A_1, \ldots A_7$ are as defined in Section 4 and have eigenvalues with strictly negative real parts.

We note that subsystem Σ_{e1} is in cascade with Σ_{e4}, subsystems Σ_{e2} and Σ_{e7} are in cascade with subsystem Σ_{e5}, and subsystem Σ_{e3} is in cascade with Σ_{e6} and Σ_{e7}. Using this cascade structure, we show that the overall system (A.3) is GES by constructing the desired Lyapunov function W from the Lyapunov functions of each subsystems V_i.

We consider Lyapunov functions $V_1, \ldots, V_7$ for each subsystem Σ_{e1} to Σ_{e7} of the form:

$$V_i = e_i^T P_i e_i \qquad \text{for } i \in \{1, \ldots, 7\},$$

where $P_i^T = P_i > 0$ satisfies the Lyapunov equation $P_iA_i + A_i^T P_i = -I$. This is always possible as the eigenvalues of A_i have strictly negative real parts, in view of [35, Theorem 4.6]. We will show that each subsystem Σ_{ei} is input-to-state stable (ISS) using V_i [45].

For subsystems Σ_{e4}, Σ_{e5} and Σ_{e6}, taking the derivative of $V_i = e_i^T P_i e_i$ along solutions of Σ_{ei} for $i \in \{4, 5, 6\}$, we obtain:

$$\dot{V}_i \leq \ -|e_i|^2 \leq -\frac{1}{2}|e_i|^2.$$

Next, we show that V_1 is an ISS-Lyapunov function [45] for subsystem Σ_{e1} w.r.t e_4. Taking the derivative of $V_1 = e_1^T P_1 e_1$ along the solutions of Σ_{e1}, we obtain:

$$\dot{V}_1 = e_1^T(P_1A_1 + A_1^T P_1)e_1 + 2e_1^T P_1(\phi_1(x_{41}) - \phi_1(\hat{x}_{41})).$$

Since the function S_2 is globally Lipschitz with constant ρ_2 from (4), we have that:

$$|\phi_1(x_{41}) - \phi_1(\hat{x}_{41})| \leq \theta_A a C_2 \rho_2 |x_{41} - \hat{x}_{41}| \leq \rho_{e1}|e_4|,$$

where $\rho_{e1} = \theta_A a C_2 \rho_2$. Therefore,

$$\dot{V}_1 \leq -|e_1|^2 + 2|e_1||P_1|\rho_{e1}|e_4|.$$

Recalling that $\xi\chi \leq \frac{1}{2}\xi^2 + \frac{1}{2}\chi^2$, for any ξ, $\chi \in \mathbb{R}$ and letting $\xi = 2|P_1|\rho_{e1}|e_4|$ and $\chi = |e_1|$, we obtain the following:

$$\begin{aligned}\dot{V}_1 \leq & -|e_1|^2 + \frac{1}{2}|e_1|^2 + 2|P_1|^2\rho_{e1}^2|e_4|^2 \\ \leq & -\frac{1}{2}|e_1|^2 + 2|P_1|^2\rho_{e1}^2|e_4|^2.\end{aligned}$$

With similar arguments for the remaining systems, along the solutions of Σ_{ei}, for $i \in \{2, 3, 7\}$:

$$\dot{V}_i \leq -\frac{1}{2}|e_i|^2 + \gamma_i|\mu_i|^2,$$

where $\mu_1 = e_4$, $\mu_2 = e_5$, $\mu_3 = (e_6, e_7)$, $\mu_7 = e_5$ and $\gamma_i = 2|P_i|^2\rho_{ei}^2$ where $\rho_{e2} = \theta_B b C_4 \rho_2$, $\rho_{e3} = \theta_G g C_7 \rho_2$ and $\rho_{e7} = \theta_B b C_6 \rho_2$.

The composite Lyapunov function for the overall system (A.4) W is constructed using the Lyapunov function for each subsystem V_i [46]. We consider the candidate Lyapunov function that is positive definite and radially unbounded:

$$W = a_1V_1 + a_2V_2 + a_3V_3 + V_4 + V_5 + V_6 + a_7V_7, \tag{A.5}$$

for a_1, a_2, a_3, $a_7 > 0$, which are determined below.

The derivative of (A.5) along the solution of the overall error system (A.4) is

$$\begin{aligned}\dot{W} \leq & -a_1\frac{1}{2}|e_1|^2 + a_1\gamma_1|e_4|^2 - a_2\frac{1}{2}|e_2|^2 \\ & + a_2\gamma_2|e_5|^2 - a_3\frac{1}{2}|e_3|^2 + a_3\gamma_3(|e_6|^2 + |e_7|^2) \\ & - \frac{1}{2}|e_4|^2 - \frac{1}{2}|e_5|^2 - \frac{1}{2}|e_6|^2 - a_7\frac{1}{2}|e_7|^2 + a_7\gamma_7|e_5|^2.\end{aligned}$$

By taking $0 < a_1 < \frac{1}{2\gamma_1}$, $0 < a_2 < \frac{1}{\gamma_2}\left(\frac{1}{2} - a_7\gamma_7\right)$, $0 < a_3 < \frac{1}{2\gamma_3}$ and $a_7 > 2a_3\gamma_3$, we obtain

$$\begin{aligned}\dot{W} \leq & -\frac{1}{8\gamma_1}|e_1|^2 - \frac{1}{2\gamma_2}\left(\frac{1}{2} - \gamma_7\right)|e_2|^2 - \frac{1}{8\gamma_3}|e_3|^2 \\ & - \frac{1}{2}|e_4|^2 - \frac{1}{2}|e_5|^2 - \frac{1}{2}|e_6|^2 - \frac{1}{4}|e_7|^2.\end{aligned} \tag{A.6}$$

By letting $\tilde{P} = \text{diag}(a_1P_1, a_2P_2, a_3P_3, P_4, P_5, P_6, a_7P_7)$ as well as denoting $\lambda_{min}(\tilde{P})$ and $\lambda_{max}(\tilde{P})$ as the maximum and minimum eigenvalues of $\tilde{P}$ respectively, we have shown that (A.2) is fulfilled with $k_1 = \lambda_{min}(\tilde{P}) > 0$, $k_2 = \lambda_{max}(\tilde{P}) > 0$, $k_3 = \min\{\frac{1}{8\gamma_1}, \frac{1}{2\gamma_2}\left(\frac{1}{2} - \gamma_7\right), \frac{1}{8\gamma_3}, \frac{1}{4}\} > 0$ and $k_4 = 2|\tilde{P}| > 0$. □

Continuing the proof of Theorem 1:

The derivative of W along the solutions to (A.1) is:

$$\begin{aligned}\dot{W} &= \frac{\partial W}{\partial e}\Big(Ae + G\big(\gamma(Hx) - \gamma(H\hat{x} + KCe)\big) + LCe\Big),\\ &= \frac{\partial W}{\partial e}\Big(\underbrace{Ae + G\gamma(Hx) - G\gamma(H\hat{x})}_{\text{nominal estimation error system (A.3)}}\\ &\qquad + \underbrace{G\gamma(H\hat{x}) - G\gamma(H\hat{x} + KCe) + LCe}_{\text{perturbation terms}}\Big)\end{aligned}$$

From Lemma 1:

$$\dot{W} \leq -k_3|e|^2 + \frac{\partial W}{\partial e}G(\gamma(H\hat{x}) - \gamma(H\hat{x} + KCe)) + \frac{\partial W}{\partial e}LCe.$$

As γ is globally Lipschitz, $|\gamma(H\hat{x}) - \gamma(H\hat{x} + KCe)| \leq \rho|KCe| \leq \rho|K||C||e|$. From (A.2):

$$\begin{aligned}\dot{W} &\leq -k_3|e|^2 + \left|\frac{\partial W}{\partial e}\right||G|\rho|K||C||e| + \left|\frac{\partial W}{\partial e}\right||L||C||e|\\ &\leq -k_3|e|^2 + k_4|C||e|^2(\rho|G||K| + |L|),\end{aligned}$$

where k_3 and k_4 are constructed in Lemma 1.

Therefore, if K and L satisfy the following condition:

$$\rho|K||G| + |L| < \frac{k_3}{k_4|C|},$$

then

$$\dot{W} \leq -\tilde{k}_3|e|^2,$$

where $\tilde{k}_3 = k_3 - k_4|C|(\rho|G||K| + |L|) > 0$.

Therefore, the origin of the estimation error system (A.1) is GES according to [35, Definition 4.5], i.e. for all $t \geq 0$,

$$|e(t)| \leq k\exp(-\lambda t)|e(0)| \qquad \forall e(0) \in \mathbb{R}^n,$$

for $k, \lambda > 0$. □

Appendix B. Proof of Theorem 2

The proof that follows is performed for the Wendling at. al. model, where the models by Stam et. al. as well as Jansen and Rit can be derived from. The proof for these models can be performed in a similar fashion.

From (26) and (27), the perturbed error system is:

$$\begin{aligned}\dot{e} &= Ae + G(\theta + \epsilon_\theta)\gamma(Hx) - G(\theta)\gamma\Big(H\hat{x} + K\big(C\hat{x} - (y + \epsilon_y)\big)\Big)\\ &\quad - L(C\hat{x} - (y + \epsilon_y)) + \sigma(u,\, y,\, \theta + \epsilon_\theta) - \sigma(u + \epsilon_u,\, y + \epsilon_y,\, \theta) + \epsilon_{sys}\\ &= \underbrace{(A + LC)e + \Psi(x,\, \hat{x})}_{\text{nominal system (A.1) from Theorem 1}} + \underbrace{\Psi_\epsilon(x,\, \hat{x},\, \epsilon_y,\, \epsilon_u,\, \epsilon_\theta,\, \epsilon_{sys})}_{\text{perturbation terms}},\end{aligned} \tag{B.1}$$

where $K = (\kappa_1, \ldots, \kappa_m)$,

$\Psi = \Big(0,\, \theta_A a C_2\big(S_2(x_{41}) - S_2(\hat{x}_{41} - \kappa_1 Ce)\big),\, 0,\, \theta_B b C_4\big(S_2(x_{51}) - S_2(\hat{x}_{51} -$

$\kappa_2 Ce\big)$, 0, $\theta_G g C_7\big(S_2(x_{61}-x_{71})-S_2(\hat{x}_{61}-\hat{x}_{71}-\kappa_3 Ce)\big)$, 0, 0, 0, 0, 0, 0, 0, $\theta_B b C_6\big(S_2(x_{51})-S_2(\hat{x}_{51}-\kappa_2 Ce)\big)\Big)$ and

$\Psi_\epsilon = \Big(0,\, \theta_A a C_2\big(S_2(\hat{x}_{41}-\kappa_1 Ce)-S_2(\hat{x}_{41}-\kappa_1 Ce-\epsilon_y)\big)+aC_2S_2(x_{41})\epsilon_\theta+au\epsilon_\theta+\theta_A a\epsilon_u,\, 0,\, \theta_B bC_4\big(S_2(\hat{x}_{51}-\kappa_2 Ce)-S_2(\hat{x}_{51}-\kappa_2 Ce-\epsilon_y)\big)+bC_4S_2(x_{51})\epsilon_\theta,\, 0,\, \theta_G gC_7\big(S_2(\hat{x}_{61}-\hat{x}_{71}-\kappa_3 Ce)-S_2(\hat{x}_{61}-\hat{x}_{71}-\kappa_3 Ce-\epsilon_y)\big)+gC_7S_2(x_{61}-x_{71})\epsilon_\theta,\, 0,\, \theta_A aC_1\big(S_2(y)-S_2(y+\epsilon_y)\big)+aC_1S_2(y)\epsilon_\theta,\, \theta_A aC_3\big(S_2(y)-S_2(y+\epsilon_y)\big)+aC_3S_2(y)\epsilon_\theta,\, 0,\, \theta_A aC_5\big(S_2(y)-S_2(y+\epsilon_y)\big)+aC_5S_2(y)\epsilon_\theta,\, 0,\, \theta_B bC_6\big(S_2(\hat{x}_{51}-\kappa_2 Ce)-S_2(\hat{x}_{51}-\kappa_2 Ce-\epsilon_y)\big)+bC_6S_2(x_{51})\epsilon_\theta\Big)+\epsilon_{sys}$.

We will show that the solutions of the error system (B.1) is input-to-state stable (ISS) [47] with respect to the uncertainties ϵ_y, ϵ_u, ϵ_θ and ϵ_{sys}. For this purpose, we use function W as defined in (A.5).

The derivative of W along the solutions of (B.1) is:

$$\dot{W} = \frac{\partial W}{\partial e}\Big((A+LC)e+\Psi(x,\,\hat{x})\Big)+\frac{\partial W}{\partial e}\Psi_\epsilon(x,\,\hat{x},\,\epsilon_y,\,\epsilon_u,\,\epsilon_\theta,\,\epsilon_{sys}).$$

From Theorem 1, there exists $\tilde{k}_3, \tilde{k}_4 > 0$ such that:

$$\dot{W} \leq -\tilde{k}_3|e|^2+\tilde{k}_4|e||\Psi_\epsilon|.$$

Using the fact that the function S_2 is globally Lipschitz with Lipschitz constant ρ_2 and $S_2(z) \leq \alpha_2$ for any $z \in \mathbb{R}$ as defined in (4), we obtain:

$$\dot{W} \leq -\tilde{k}_3|e|^2+\tilde{k}_4|e|\big(\sigma_y(|\epsilon_y|)+\sigma_\theta(|\epsilon_\theta|)+\sigma_u(|\epsilon_u|)+|\epsilon_{sys}|\big)$$

where $\sigma_y(|\epsilon_y|) = \big(|(\theta_A aC_2\rho,\, \theta_B bC_4\rho,\, \theta_G gC_7\rho,\, \theta_A aC_1\rho,\, \theta_A aC_3\rho,\, \theta_A aC_5\rho,\, \theta_B bC_6\rho)|+|L|\big)|\epsilon_y|$, $\sigma_u(|\epsilon_u|) = \theta_A a|\epsilon_u|$ and $\sigma_\theta(|\epsilon_\theta|) = |\big((aC_2\alpha+a\|u\|_{[0,t]})|\epsilon_\theta|,\, bC_4\alpha|\epsilon_\theta|,\, gC_7\alpha|\epsilon_\theta|,\, aC_1\alpha|\epsilon_\theta|,\, aC_3\alpha|\epsilon_\theta|,\, aC_5\alpha|\epsilon_\theta|,\, bC_6\alpha|\epsilon_\theta|\big)|$.

Therefore, if

$$|e| > \frac{2\tilde{k}_4}{\tilde{k}_3}\big(\sigma_y(|\epsilon_y|)+\sigma_u(|\epsilon_u|)+\sigma_\theta(|\epsilon_\theta|)+|\epsilon_{sys}|\big),$$

then

$$\dot{W} \leq -\frac{1}{2}\tilde{k}_3|e|^2. \tag{B.2}$$

From (B.2), [35, (4.49) of Theorem 4.19] is fulfilled and [35, (4.48) of Theorem 4.19] is satisfied with $\alpha_1 = k_1$ and $\alpha_2 = k_2$, where k_1 and k_2 are from Lemma 1 in Appendix A. Therefore, we can conclude that the error system (B.1) is ISS with respect to ϵ_y, ϵ_u, ϵ_θ and ϵ_{sys} with gains $\gamma_y(s) = \frac{k_2}{k_1}\frac{2\tilde{k}_4}{\tilde{k}_3}\sigma_y(s)$, $\gamma_\theta(s) = \frac{k_2}{k_1}\frac{2\tilde{k}_4}{\tilde{k}_3}\sigma_\theta(s)$, $\gamma_u(s) = \frac{k_2}{k_1}\frac{2\tilde{k}_4}{\tilde{k}_3}\sigma_u(s)$ and $\gamma_{sys}(s) = \frac{k_2}{k_1}\frac{2\tilde{k}_4}{\tilde{k}_3}.s$, for $s \geq 0$ and $\gamma_y(0) = \gamma_\theta(0) = \gamma_u(0) = \gamma_{sys}(0) = 0$. Here, we have shown it for the model by Wendling et. al. [3], but similar arguments apply to all other models considered in Section 3. □

Appendix C. Values and description of the constants

Parameter	Description	Standard value
$\frac{1}{a_1}$, $\frac{1}{a_2}$	Average time constant in the excitatory feedback loop	$a_1 = 55$, $a_2 = 605\,s^{-1}$
$\frac{1}{b_1}$, $\frac{1}{b_2}$	Average time constant in the inhibitory feedback loop	$b_1 = 27.5$, $b_2 = 55\,s^{-1}$
V_1, α_1, r_1	Parameters for the sigmoid function. α_1 is the maximum firing rate. r_1 is the slope of the sigmoid and V_1 is the threshold of the population's mean membrane potential.	$V_1 = 6\,\mathrm{mV}$, $\alpha_1 = 5\,\mathrm{s}^{-1}$, $r_1 = 0.56\,\mathrm{mV}^{-1}$
C_3, C_4	Average number of synaptic contacts in the inhibitory feedback loop	$C_3 = 32$, $C_4 = 3$
θ_A, θ_B	Synaptic gain of the excitatory and inhibitory populations respectively	$\theta_A = 1.65$, $\theta_B = 32$

Table C1. Standard constants used in the model by Stam et. al. in [1].

Parameter	Description	Standard value
$\frac{1}{a}$	Average time constant in the excitatory feedback loop	$a = 100\,s^{-1}$
$\frac{1}{b}$	Average time constant in the slow inhibitory feedback loop	$b = 50\,s^{-1}$
V_2, α_2, r_2	Parameters for the sigmoid function. α_2 is the maximum firing rate. r_2 is the slope of the sigmoid and V_2 is the threshold of the population's mean membrane potential.	$V_2 = 6\,\mathrm{mV}$, $\alpha_2 = 5\,\mathrm{s}^{-1}$, $r_2 = 0.56\,\mathrm{mV}^{-1}$
C_1, C_2	Average number of synaptic contacts in the excitatory feedback loop	With $C = 135$, $C_1 = C$ and $C_2 = 0.8\,C$
C_3, C_4	Average number of synaptic contacts in the slow inhibitory feedback loop	$C_3 = C_4 = 0.25\,C$
θ_A and θ_B	Synaptic gain of the excitatory and inhibitory populations respectively	$\theta_A = 3.25$ and $\theta_B = 22$

Table C2. Standard constants used in the model by Jansen and Rit in [2].

Parameter	Description	Standard value
$\frac{1}{a}$	Average time constant in the excitatory feedback loop	$a = 100\,s^{-1}$
$\frac{1}{b}$	Average time constant in the slow inhibitory feedback loop	$b = 50\,s^{-1}$
$\frac{1}{g}$	Average time constant in the fast inhibitory feedback loop	$g = 500\,s^{-1}$
V_2, α_2, r_2	Parameters for the sigmoid function. α_2 is the maximum firing rate. r_2 is the slope of the sigmoid and V_2 is the threshold of the population's mean membrane potential.	$V_2 = 6\,\mathrm{mV}$, $\alpha_2 = 5\,\mathrm{s}^{-1}$, $r_2 = 0.56\,\mathrm{mV}^{-1}$
C_1, C_2	Average number of synaptic contacts in the excitatory feedback loop	With $C = 135$, $C_1 = C$ and $C_2 = 0.8\,C$
C_3, C_4	Average number of synaptic contacts in the slow inhibitory feedback loop	$C_3 = C_4 = 0.25\,C$
C_5, C_6	Average number of synaptic contacts between the fast and slow inhibitory feedback loop	$C_5 = 0.3\,C$ and $C_6 = 0.1\,C$
C_7	Average number of synaptic contacts in the fast inhibitory feedback loop	$C_7 = 0.8\,C$
θ_A, θ_B and θ_G	Synaptic gain of the excitatory, fast inhibitory and slow inhibitory populations respectively	See [3] for values corresponding to different brain activity

Table C3. Standard constants used in the model by Wendling et. al. in [3]

Acknowledgements

This work was carried out with support from the St. Vincent's Hospital (Melbourne) Research Endowment Fund.

References

[1] C.J. Stam, J.P.M. Pijn, P. Suffczynski, and F.H. Lopes da Silva. Dynamics of the human alpha rhythm: evidence for non-linearity? *Clinical Neurophysiology*, 110(10):1801–1813, 1999.

[2] B.H. Jansen and V.G. Rit. Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biological Cybernetics*, 73:357–366, 1995.

[3] F. Wendling, A. Hernandez, J.J. Bellanger, P. Chauvel, and F. Bartolomei. Interictal to ictal transition in human temporal lobe epilepsy: insights from a computational model of intracerebral EEG. *Journal of Clinical Neurophysiology*, 22(5):343, 2005.

[4] I. Osorio, M.G. Frei, and S.B. Wilkinson. Real-Time Automated Detection and Quantitative Analysis of Seizures and Short-Term Prediction of Clinical Onset. *Epilepsia*, 39(6):615–627, 1998.

[5] J. Gotman. Automatic detection of seizures and spikes. *Journal of Clinical Neurophysiology*, 16(2):130, 1999.

[6] P. Mirowski, D. Madhavan, Y. LeCun, and R. Kuzniecky. Classification of patterns of EEG synchronization for seizure prediction. *Clinical neurophysiology*, 120(11):1927–1940, 2009.

[7] L. Kuhlmann, D. Freestone, A. Lai, A.N. Burkitt, K. Fuller, D.B. Grayden, L. Seiderer, S. Vogrin, I.M.Y. Mareels, and M.J. Cook. Patient-specific bivariate-synchrony-based seizure prediction for short prediction horizons. *Epilepsy Research*, 2010.

[8] F. Mormann, R.G. Andrzejak, C.E. Elger, and K. Lehnertz. Seizure prediction: the long and winding road. *Brain*, 130(2):314, 2007.

[9] B.H. Jansen, G. Zouridakis, and M.E. Brandt. A neurophysiologically-based mathematical model of flash visual evoked potentials. *Biological cybernetics*, 68(3):275–283, 1993.

[10] P.A. Robinson, C.J. Rennie, D.L. Rowe, S.C. O'Connor, J.J. Wright, E. Gordon, and R.W. Whitehouse. Neurophysical modeling of brain dynamics. *Neuropsychopharmacology*, 28, 2003. Neurophysical Modeling of Brain Dynamics.

[11] O. David and K.J. Friston. A neural mass model for MEG/EEG: coupling and neuronal dynamics. *NeuroImage*, 20(3):1743–1755, 2003.

[12] A. Babajani-Feremi and H. Soltanian-Zadeh. Multi-area neural mass modelling of EEG and MEG signals. *Neuroimage*, 52:793–811, 2010.

[13] B.D.O. Anderson, J.B. Moore, and J. Barratt. *Optimal filtering*. Prentice-Hall Englewood Cliffs, NJ, 1979.

[14] T. Kailath, A.H. Sayed, and B. Hassibi. *Linear estimation*. Prentice Hall Upper Saddle River, NJ, 2000.

[15] M.I. Garrido, J.M. Kilner, S.J. Kiebel, K.E. Stephan, and K.J. Friston. Dynamic causal modelling of evoked potentials: A reproducibility study. *Neuroimage*, 36:701–714, 2007.

[16] A.C. Marreiros, J. Daunizeau, S.J. Kiebel, and K.J. Friston. Population dynamics: Variance and the sigmoid activation function. *Neuroimage*, 42:701–714, 2008.

[17] A.C. Marreiros, S.J. Kiebel, J. Daunizeau, L.M. Harrison, and K.J. Friston. Population dynamics under the Laplace assumption. *Neuroimage*, pages 701–714, 2009.

[18] R.J. Moran, K.E. Stephan, T. Seidenbechner, H.-C. Pape, R.J. Dolan, and K.J. Friston. Dynamic causal models of steady-state responses. *Neuroimage*, 44:796–811, 2009.

[19] J. Daunizeau, O. David, and KE Stephan. Dynamic causal modelling: a critical review of the biophysical and statistical foundations. *Neuroimage*, 2009.

[20] A.C. Marreiros, S.J. Kiebel, and K.J. Friston. A dynamic causal model study of neuronal population dynamics. *Neuroimage*, 51:91–101, 2010.

[21] O. David, S.J. Kiebel, L.M. Harrison, J. Mattout, J.M. Kilner, and K.J. Friston. Dynamic causal modeling of evoked responses in eeg and meg. *NeuroImage*, 30(4):1255–1272, 2006.

[22] S. Schiff and T. Sauer. Kalman filter control of a model of spatiotemporal cortical dynamics. *Journal of Neural Engineering*, 5:1–8, 2008.

[23] P. Frogerais. *Model and identification in epilepsy: from neuronal population dynamics to EEG signals (in French)*. PhD thesis, University of Rennes 1, 2008.

[24] G. Ullah and S. Schiff. Assimilating seizure dynamics. *PLoS computational biology*, 6(5), 2010.

[25] T. Lu and J. Lee. Polytopic linear parameter-varying model of epileptiform activity. In *American Control Conference, Baltimore, U.S.A.*, 2010.

[26] DR Freestone, P. Aram, M. Dewar, K. Scerri, DB Grayden, and V. Kadirkamanathan. A data-driven framework for neural field modeling. *NeuroImage*, 2011.

[27] I. Tokuda, U. Parlitz, L. Illing, M. Kennel, and H. Abarbanel. Parameter estimation for neuron models. pages 251–256, 2003.

[28] Y. Totoki, K. Mitsunaga, H. Suemitsu, and T. Matsuo. Firing pattern estimation of synaptically coupled hindmarsh-rose neurons by adaptive observer. 5164:338–347, 2008.

[29] I. Tyukin, E. Steur, H. Nijmeijer, D. Fairhurst, I. Song, A. Semyanov, and C. Van Leeuwen. State and parameter estimation for canonic models of neural oscillators. *International journal of neural systems*, 20(3):193, 2010.

[30] Y. Mao, W. Tang, Ying Liu, and L. Kocarev. Identification of biological neurons using adaptive observers. *Cognitive Processing*, 10:41–53, 2009.

[31] P.L. Nunez and R. Srinivasan. *Electric fields of the brain: the neurophysics of EEG*. Oxford University Press New York, 2006.

[32] G. Deco, V. Jirsa, P.A. Robinson, M. Breakspear, and K. Friston. The dynamic brain: From spiking neurons to neural masses and cortical fields. *Cerebral Cortex*, 4(8):1–35, 2008.

[33] H.R. Wilson and J.D. Cowan. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal*, 12:1–24, 1972.

[34] S.F.H. Lopes, A. Van Rotterdam, P. Barts, E. Van Heusden, and W. Burr. Models of neuronal populations: the basic mechanisms of rhythmicity. *Progress in brain research*, 45:281, 1976.

[35] H.K. Khalil. *Nonlinear systems*. Prentice Hall, 3rd edition, 2000.

[36] L. Ljung. *System identification: theory for the user*. Prentice Hall, Upper Saddle River, NJ, 1999.

[37] H. Hammouri and Z. Tmar. Unknown input observer for state affine systems: A necessary and sufficient condition. *Automatica*, 46(2):271–278, 2010.

[38] M.S. Chong, R. Postoyan, D. Nešić, L. Kuhlmann, and A. Varsavsky. A nonlinear estimator for the activity of neuronal populations in the hippocampus. In *Proceedings of the 18th IFAC World Congress*, 2011.

[39] S. Raghavan and J.K. Hedrick. Observer design for a class of nonlinear systems. *International Journal of Control*, 59(2):515–528, 1994.

[40] M. Arcak and P. Kokotović. Nonlinear observers: a circle criterion design and robustness analysis. *Automatica*, 37(12):1923–1930, 2001.

[41] I. Soltesz and K. Staley. *Computational neuroscience in epilepsy*. Academic Press, 2008.

[42] H.R. Wilson and J.D. Cowan. A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Biological Cybernetics*, 13(2):55–80, 1973.

[43] P.A. Robinson, C.J. Rennie, and D.L. Rowe. Dynamics of large-scale brain activity in normal arousal states and epileptic seizures. *Physical Review E*, 65, 2002. Dynamics of large-scale brain activity in normal arousal states and epileptic seizures.

[44] D.T.J. Liley, P.J. Cadusch, and M.P. Dafilis. A spatially continuous mean field theory of electrocortical activity. *Network: Computation in Neural Systems*, 13(1):67–113, 2002.

[45] E.D. Sontag and Y. Wang. On characterizations of the input-to-state stability property. *Systems & Control Letters*, 24(5):351–359, 1995.

[46] E.D. Sontag and A. Teel. Changing supply functions in input/state stable systems. *IEEE Transactions on Automatic Control*, 40(8):1476–1478, 1995.

[47] E.D. Sontag. Input to state stability: Basic concepts and results. *Nonlinear and Optimal Control Theory*, pages 163–220, 2008.