Brain Topography (2019) 32:28–65
https://doi.org/10.1007/s10548-018-0666-3

ORIGINAL PAPER

A Blind Module Identification Approach for Predicting Effective Connectivity Within Brain Dynamical Subnetworks

Fadi N. Karameh[1] · **Ziad Nahas**[2,3]

Received: 8 November 2017 / Accepted: 28 July 2018 / Published online: 4 August 2018

Abstract
Model-based network discovery measures, such as the brain effective connectivity, require fitting of generative process models to measurements obtained from key areas across the network. For distributed dynamic phenomena, such as generalized seizures and slow-wave sleep, studying effective connectivity from real-time recordings is significantly complicated since (i) outputs from only a subnetwork can be practically measured, and (ii) exogenous subnetwork inputs are unobservable. Model fitting, therefore, constitutes a challenging blind module identification or model inversion problem for finding both the parameters and the many unknown inputs of the subnetwork. We herein propose a novel estimation framework for identifying nonlinear dynamic subnetworks in the case of slowly-varying, otherwise unknown local inputs. Starting with approximate predictions obtained using Cubature Kalman filtering, residuals of local output predictions are utilized to improve upon local input estimates. The algorithm performance is tested on both simulated and clinical EEG of induced seizures under electroconvulsive therapy (ECT). For the simulated network, the algorithm significantly boosted the estimation accuracy for inputs and connections from noisy EEG. For the clinical data, the algorithm predicted increased subnetwork inputs during the pre-stimulus anesthesia condition. Importantly, it predicted an increased frontocentral connectivity during the generalized seizure that is commensurate with electrode placement and that corroborates the clinical hypothesis of increased frontal focality of therapeutic ECT seizures. The proposed framework can be extended to account for several input configurations and can in principle be applied to study effective connectivity within brain subnetworks defined at the microscale (cortical lamina interaction) or at the macroscale (sensory integration).

Keywords Model inversion · Effective connectivity · Kalman filtering · Neuronal modeling · Blind deconvolution · Brain subnetworks

Introduction

Dynamical system modeling is an essential component in understanding, predicting and subsequently manipulating the behavior of physical processes in a wide range of scientific disciplines. Since they primarily describe the system response as it interacts with its active environment, successful models hinge on the collection of rich sets of stimuli and responses (inputs and outputs) of the system in operation. For many physical and particularly biological systems, however, a variety of driving, modulatory and environmental inputs contribute to a given output. This makes the requirement of accessing all the driving inputs in vivo not possible or impractical. In cortical networks, for example, the response of a given local neuronal dynamical process is the result of interactions with other local neuronal processes across the cortical network and with a variety of deeper brain structures (Park and Friston 2013). Particularly when the network is distributed (e.g. generalized seizures) or when deep brain observations are not available (e.g. scalp EEG), only modules or subnetworks within the true network can practically be modeled and assimilated to available measurements.

In fact this is a hard blind model inversion problem since both the subnetwork model and the inaccessible inputs are to be estimated. The problem is ill-posed (many unknowns, few

Handling Editor: Fabrice Wendling.

✉ Fadi N. Karameh
fadi.karameh@aub.edu.lb

[1] Department of Electrical and Computer Engineering, American University of Beirut, Beirut, Lebanon

[2] Department of Psychiatry, American University of Beirut, Beirut, Lebanon

[3] Present Address: Department of Psychiatry, University of Minnesota, Minneapolis, MN, USA

measurements), and the accuracy of the inversion depends critically on the assumptions regarding the role of those inaccessible inputs, or network interactions, in shaping the local process response (Roebroeck et al. 2011). Our current effort is specifically concerned with techniques that could improve the accuracy of model inversion and, hence, of causal inference in subnetworks.

A Myriad of Inputs

From a modeling standpoint, one can generally classify inputs in terms of their impact on the output as either disturbance inputs or driving inputs. Disturbance inputs have, by definition, an undesirable impact on the system performance that is often limited by the natural robustness of the physical process. They are of a random nature, are often inaccessible and their effect can be ameliorated in the averaged output [e.g. evoked potentials in EEG (Mouraux and Iannetti 2008)]. On the other hand, driving input are structured and can be broadly divided into either (1) modulating inputs that specify a "setpoint" or operating regime of the system over a long time scale [e.g. sleep mechanisms (McCormick and Bal 1997), or conductance fluctuations (Destexhe et al. 2003)], or (2) controlling inputs that dictate the time profile of the response over a fast time scale commensurate with the system bandwidth (e.g. cortical interlaminar processing (Plomp et al. 2014)] and could possibly lead to network-level phenomena [e.g oscillatory synchronization (Palva and Palva 2012)). In brain sciences, driving inputs are often not amenable to recording. Inaccessible modulating inputs are commonly represented implicitly as global parameters that tune the input/output dynamic model to a given operating condition [e.g. homeostatic sleep drive in Sedigh-Sarvestani et al. (2012), although some system models do incorporate a block corresponding to the modulating subsystem]. Inaccessible control inputs, on the other hand, are explicitly attended for, albeit with varying levels of detail. In simpler cases, the control inputs for some processes are can be approximated with stereotypical waveform shapes. This is a popular undertaking across brain effective connectivity approaches [e.g. evoked brain responses caused by short-lived localized pulses (Rennie et al. 2002) or gamma density function (David et al. 2006), box-car inputs in fMRI modeling (Dijkstra et al. 2017)]. In more complex cases, control inputs are persistent and possibly represent feedback drive from connected structures, and are thus largely unknown unless experimentally observed. Notwithstanding the paucity of input recordings, and to proceed with a useful form on neural process modeling, modelers strive to include all brain structures that are believed to be relevant for reproducing the phenomenon under investigation (Roebroeck et al. 2011). The timing and profile of the inaccessible input is constrained to acquire simple parameterized forms within the experimental design, and can hence is estimated along with the model parameters. This procedure is commonly termed model blind input deconvolution for linear systems (Bell and Sejnowski 1995), more generally model inversion or blind identification.

Depending on the dynamic model under study, the input recovery problem based solely on output recordings can generally be burdened by a host of challenges including: the presence of nonlinearity in the system dynamics, the uncertainty in the process model, the interference of process random fluctuations and environmental noise, the changes in the dynamical system itself, and the complexity of the driving input.

In brain sciences, these challenges are particularly abundant since (i) the available recordings of the brain nonlinear dynamics are often far removed from the neural activity sources of interest and (ii) the modeled processes at almost any scale are subjected to a myriad of exogenous sources constituting random, modulating or controlling inputs (above). In functional brain imaging, for example, recordings of blood oxygenation levels (or of metabolic processes) are auxiliary observations of the ongoing neural activity. Here the hemodynamic phenomena (the system) mediating fMRI output signal generation is nonlinearly related to the input neural firing, is subjected to endogenous noise in the metabolic and the vascular responses, and varies between brain areas as well as across subjects. In electromagnetic recordings, field potentials, including depth, subdural and scalp EEG, are manifestations of aggregated extracellular currents over increasingly larger spatial scales and hence are auxiliary output measurements of the underlying neural activity. In turn, the local neural circuitry mediating output field recordings is subjected to environmental noise and synaptic fluctuations, and its dynamics are modified by the activity of neuromodulating systems. The circuitry is moreover (nonlinearly) affected by a multitude of inputs arriving from distant cortical and subcortical structures. Since it is clearly not possible or impractical to obtain recordings from all these connected structures, the neuronal circuit model is realistically that of a subnetwork, or a module within the total network, and the contributions of the remaining connected yet unmodeled structures constitute unknown exogenous inputs (possibly feedback signals) arriving into that module. Hence, the quality of modeling in brain networks at any spatial scale is intimately associated with the amount of knowledge that can be gathered (or assumed) about the inputs to the subnetwork.

Brain Directed Connectivity

Over the last decade, brain connectivity have received an ever-increasing attention due to its role in improving our understanding normal and pathological brain functionality

and is being driven by advances in both technological and analytical tools. In Friston (2011), Friston defines functional and effective connectivity to classify methods that are descriptive or generative of the observed signals, respectively. Functional connectivity methods study the statistical dependence between recorded brain signals with little or no attention to the underlying system. Effective connectivity, on the other hand, explains causal inference in generative models of the recorded signals. The complexity of these generative models ranges from generic (black-box) to detailed neurophysiological (grey-box). First, autoregressive (AR) or "black-box" models furnish predictions of multiple signals based on their history; by assessing increments in predictability of a given time series (or signal) upon incorporating past knowledge from other series, causal directional relationships can be retrieved. Here, Granger causality is a main metric to explore directed brain connections particularly in electromagnetic recordings (EEG-field potentials and MEG) (Goebel et al. 2003). Second, biophysically-informed or "grey-box" models incorporate experimental evidence to construct elements of the underlying hidden process. Here, Dynamic Causal Modeling is here a mainstream approach to explore alterative models and assess directed connectivity using the estimated parameters (Friston et al. 2013).

Dynamic Causal modeling (DCM) principally induces the probability distributions of the various parameters and the selection among the candidate models, with applications based on hemodynamic responses (fMRI) or observed electromagnetic (EEG/MEG). To compute such probabilities, a large set of observations are needed. This is true for application of deterministic DCM, and particularly so for stochastic DCM since the hidden events are subjected to random fluctuations or system noise (Li et al. 2011). Practically, therefore, probability estimation necessitates some stationarity assumptions on the process itself or on the observed process output used to fit the model. In fMRI experiments, the BOLD signal is an observable that is considerably slower than the underlying neuronal processes. In EEG/MEG experiments, DCM often uses models of spectral densities (Moran et al. 2007) of evoked (David et al. 2006) or steady state responses Moran et al. (2009) since a spectrum varies much slower ($\sim$ s) than the underlying neural event unfolding in time ($\sim$ ms) Kiebel et al. (2009).

More recently, improvements on DCM-based effective connectivity methods continue to emerge; for bivariate studies, to account for noise and other confounding effects (Bielczyk et al. 2017), to adaptively enhance linear model estimation in DCM (Xiang et al. 2016); for multivariate studies, to improve on flexibility and tractability of DCM (Ambrogioni et al. 2017) and large scale network DCM in fMRI (Frässle et al. 2018). Since brain networks are dynamically reorganizing, imparting time-adaptation onto effective connectivity is highly desirable; here DCM approaches based on sliding windows have recently been applied (Park et al. 2017). Furthermore, adaptive approaches are used to improve convergence and accuracy of inducing probability distributions (Sengupta et al. 2015). However, we herein seek to circumvent key limitations of similar techniques including low dimensionality, linearity, system noise, and particularly stationarity of the model, as discussed below.

In tracking time variations of the neural events (as reflected in local field potential, EEG or MEG), several studies have demonstrated that Kalman filtering is a method of choice since (i) it furnishes estimates of the mean and variance (sufficient statistics) of the various unknowns *in time* (ii) it accommodate for internal system noise, and (iii) is able to closely track nonlinear state dynamics (as would be expected, for example, in seizure activity). Indeed, a monograph by Havlicek et al. (2011) have demonstrated advantages of using Kalman-based filtering (specifically the Cubature Kalman filter) against other Bayesian approaches including dynamic expectation maximization(DEM), to provide model inversion. Our group also further established the superiority of mixed discrete-continuous CKF in estimating unknown driving inputs in fMRI simulations (Madi and Karameh 2017), in tracking induced seizure models (Karameh et al. 2014), and in designing informative experiments for nonlinear neuronal model identification by virtue of accurate and efficient model estimation (Madi and Karameh 2018).

Input Deconvolution in Effective Connectivity

When fitting models for effective connectivity, process inputs that produce observed outputs clearly need to be defined. For black-box models, inputs are commonly assumed modulatory and represented as unstructured random fluctuations (white-noise character). For grey-box models, driving inputs often correspond to hypothesized external influences (e.g. thalamic inputs in evoked EEG models or neural activation in fMRI models). Here, the fitting (or model inversion) is often rendered tractable (regularized) by constraining many of the unknown parameters to the biophysical ranges and by enforcing a parameterize of partially known form of exogenous control inputs (Garrido et al. 2008; Havlicek et al. 2011).

Determining effective connectivity in a grey-box model is clearly a problem of fitting a dynamic network model, particularly because several electrodes or imaged subregions are involved, and because each node (brain area) is composed of a model of the underlying (hidden) neural activity.

The inputs to a node can be intra-areal (recurrent, intrinsic), inter-areal (long-range, endogenous), or exogenous to the network. Since intra- and inter-areal inputs are part of the network model, they are implicitly determined upon estimating the model parameters (S1–S3 connections Fig. 1a). However, determining the exogenous inputs (D1–D3 in Fig. 1a) constitutes a blind deconvolution problem that is

carried out under specific justifiable assumptions on the nature of these inputs,as presented earlier. Specifically, exogenous disturbances are commonly described as random noise processes parameterized with unknown summary statistics (e.g variance). Exogenous modulating inputs are described as parameters that tune the overall excitation of the system (e.g. under attention). Exogenous control inputs, on the other hand, can be only estimated for specific pathways (e.g. thalamic) which are experimentally known to follow stereotypical waveform shapes, such as oscillatory or pulse like, and are parameterized by few unknown variables to be determined within the estimation framework (Bastos et al. 2015; David et al. 2006).

The problem of model inversion becomes significantly more challenging for *distributed phenomena*, primarily because the model itself does not include all the brain structures that contribute to the dynamics of the emerging phenomenon being recorded, and particularly since the modeled neuronal generating networks are subject to unparameterized forms of exogenous control inputs arriving to many nodes of the network. An example phenomenon is the generalized seizure activity that are known to encompass wide areas of the brain. Here, excitatory drives to any given area could potentially arrive through any of its afferent synaptic connections from distant areas. Other distributed mechanisms include altered vigilance states including slow-wave sleep (Crunelli et al. 2015),

Therefore, in studying model-based effective connectivity for a set of output recordings, it is unclear how to specify the exogenous drive to the underlying neural subnetwork since such drive might be arriving in a distributed fashion from areas outside that subnetwork. Clearly, considering an overall network to represent all the EEG channels (or fMRI imaged regions) is unrealistic or, at best, constitutes a formidable computational challenge for model inversion.

Summary of Problem

Based on the previous introduction, we are here focused on the following problem: for the set of phenomena that are distributed across wide brain regions and subsystems (including deeper unobserved structures such as the brainstem and thalamus) or are under-represented in terms of measuring and modeling contributing subsystems, a) can we apply effective connectivity approaches to recordings obtained from a subnetwork or a module of the system (Fig. 1b), and under what conditions? And b) what are the algorithmic tools needed to perform a reasonably accurate model inversion?

Summary of Approach

The contribution of this paper is two fold. First, we develop an increased accuracy, efficient model inversion algorithm for a class of nonlinear, possibly time-varying multivariate processes. The form of the input/output function is assumed known yet dependent on a set of unknown parameters and the system is driven by several unknown, independent inputs. The inputs are also assumed to evolve at a slower time-scale in comparison with the system dynamics. To develop this algorithm, we employ a two-stage procedure that first utilizes square-root cubature Kalman filter (SCKF) for performing one-step ahead prediction to jointly estimate the input and parameters. In the second stage, the residual prediction error is decomposed to output-specific terms using independent component analysis (ICA) and these terms are then utilized to boost the accuracy of Square root Kalman Smoother (SCKS)

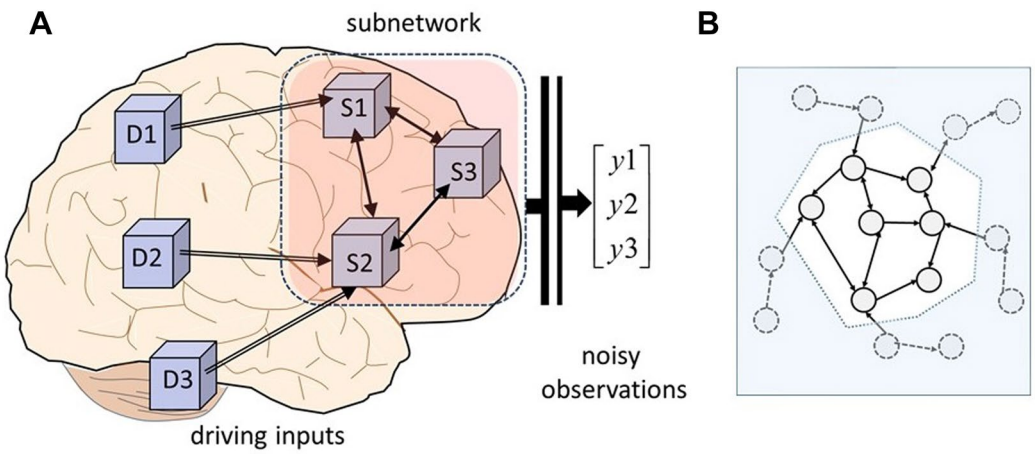

Fig. 1 Modules within a large dynamical network. **a** Schematic of subnetwork of local neuronal process (S1–S3) that interact by way of intra- an inter-areal connections (solid lines with arrows), that receive inputs from areas exogenous to the subnetwork under consideration (D1–D3, double line) and that produce noisy observations to be used in modeling. **b** A schematic of an general interconnected module or subnetwork that is to be identified within an overall network (including shaded regions)

in input prediction. The mathematical detail of the algorithm is presented in detail in "Methods", and a summary of its operation is presented in "Discussion". The presented approach is shown to push the accuracy of earlier blind model inversion using Kalman filtering (SCKS) (Havlicek et al. 2011).

Second, we apply the developed model inversion algorithm to find connections strengths on a network of neural population generator underlying a limited set of scalp EEG recordings. By virtue of the increased accuracy in input estimation to individual neural areas, we separate the inter-areal connections within the modeled network from contributions from sources external to this network, therefore leading to a proposed measure of module effective connectivity that describes physiologically justifiable connections between areas within the modeled subnetwork. We demonstrate the utility of this measure on a set simulated and empirical seizure EEG data collected during electroconvulsive therapy. To further highlight the challenges addressed, we benchmark the obtained results against a popular time-varying Granger causality measure [GPDC Havlicek et al. (2010)], and a novel extension of sliding window model reduction under DCM for seizure studies Rosch et al. (2017). We show that both methods give very accuracy for the problem at hand (nonlinear, time varying, with unknown inputs).

Electrically-Induced Seizures

Electroconvulsive therapy (ECT) is a long standing, highly effective clinical method for the treatment of severe, medication-resistant depression (Mankad et al. 2010; Merkl et al. 2009). It involves the delivery of large current stimulation to scalp electrodes leading to seizure induction. Despite the introduction of several new brain stimulation therapies, ECT remains the most effective, and often lifesaving, acute treatment for severe psychiatric disorders. Each year, over 2 million treatments are administered world wide, but not without significant risks of memory loss (Lisanby 2007). Historically, the genesis of a generalized seizure was though to be necessary and sufficient for efficacy, while the intensity of the electrical stimulus determined the magnitude of adverse cognitive effects. More recently, multiple studies demonstrated that the electrode placement, the type of electric stimulus (sine wave, wide or ultrabrief pulse) and the dosage relative to the subject-dependent seizure threshold (ST) all play critical roles in both efficacy and side effects (Sackeim et al. 1996, 2008; Spellman et al. 2009). It is now believed that seizure initiation in prefrontal cortex (PFC) is fundamental to efficacy, whereas seizure expression in medial temporal lobe (MTL) regions contributes to the loss of memory associated with the treatment. Accordingly, it is highly desirable that an ECT protocol induces prefrontal area therapeutic seizures at lower delivered electric charges while reducing the seizure propagation and impact on medial temporal areas, thereby achieving both focality and efficiency.

Since EEG provides a clear advantage in studying the temporal dynamics of seizures, we have collected multichannel EEG recordings during ECT sessions conducted using asymmetric delivery electrodes (FEAST configuration, see "Methods" for detail). We aimed to study the impact of electrode placement on seizure focality. It is thus of primary interest to identify cortical sites that are maximally involved in initiating the seizures and to determine the directionality of the ictal drive under a specific electrode configuration. Note that this problem is particularly challenging due to the occlusion of the seizure initiation data. Since seizures are induced during the stimulus delivery period, artifacts of enormous magnitude make the EEG largely unavailable (0.8 A currents). Accordingly, it is required to address connectivity and directionality of seizure activity in the time interval post-stimulus delivery—after the ictal event has been generalized and has already spread to other brain areas. The proposed method is used to assimilate four channel EEG recordings to a four-area subnetwork model of induced seizures. The combined fitting and input estimation in the algorithm separates the contributions of the inter-areal connections from that of the external sources to the seizure recordings. Unlike effective connectivity estimated using the benchmark GPDC and DCM approaches, the predicted effective connectivity by our proposed approach is shown to be sensitive to electrode configuration utilized and to concur with current hypothesis on focality of the induced seizures.

Material and Methods

Neural Models for Induced Seizures

The ECT-induced seizure model is based on modifying a popular neural population (firing rate) model for spontaneous seizures introduce by Wendling et al. (2002). Recently, our group has demonstrated that the Wendling model can be readily modified to account for seizure induction in models of normal tissue by external stimuli such as those applied during ECT (Karameh et al. 2014). By introducing a fast-to-slow inhibition, the firing dynamics for a single are demonstrated to readily switch from regular activity into seizures upon external stimulation. Details of the modified Wendling model can be found in "Appendix".

Induced Seizure in a Network

A network is built using multiple areas with long range excitatory connections from pyramidal populations in source area arriving at both pyramidal and local inhibitory

populations in the target area. A schematic of the connections in a two-area subnetwork model is shown in Fig. 2. For a total of *M* distinct EEG channels recordings, a network with M distinct blocks is constructed. The EEG recordings are treated as a multivariable output of an underlying nonlinear dynamical system as it responds to external stimulation. A single EEG channel is thus a noisy observation of the electric activity in the underlying neural tissue. Note that while the exact network size, its topology, and its constituent neural subtypes that contribute to the EEG recording are unknown, we adopt the Modified Wendling model as a lumped representation , or nonlinear building block, that exhibit many of the hallmarks of seizure EEG patterns since it (i) can switch its output from normal to ictal activity under excessive inputs (external stimuli), and (ii) can augment its activity, via recurrent positive feedback with other areas, to initiate a wide spread seizures, and (iii) can propagate ictal activity to distant areas.

We assume that the spatial distance between the considered EEG electrodes is sufficiently large so that no two electrodes record from common neural sources, thereby requiring volume conduction problems and spatial mixing. Hence, each building block contributes directly to only one output channel (for example, electrodes FP1 and FP2 in Fig. 6c). Details of the network model can be found in "Appendix".

Blind Module Identification

The main engine for identification herein is the Kalman filter (KF). Kalman filtering is a Bayesian estimation technique that allows to infer hidden dynamical states based on a set of observations. A key advantage of KF is the ability to track, recursively in time variations of the hidden states as well as adaptive behavior of the system model. Originally developed as an optimal estimator for linear estimation under gaussian noise, recent extensions allow to provide accurate estimates for nonlinear processes. Specifically for the brain sciences, the ability to track variations in nonlinear models is highly desirable since it allows tracking variability in the inputs to the system and in the model parameters a, such as synaptic plasticity and neuro-modulation, and other factors contributing to dynamic effective connectivity. In what follows we will outline the mathematical development of the algorithm. Figure 3 shows a schematic of the various components that will be presented in this section. The tutorial style detail is intended to introduce several concepts that might be novel to the interested reader. A working summary of the algorithm is available in "Discussion".

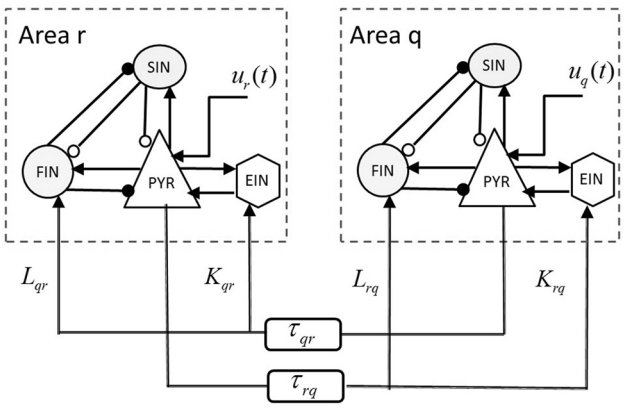

Fig. 2 Multi-area Model. Long range excitatory connections have connection strengths K and delay τ. Each area can receive a distinct time-varying exogenous input $u(t)$

The SCKS Block

This entails using a Kalman filtering procedure (the SCKS, described in "Appendix") to estimate, from noisy observations of a stochastic discrete-time dynamical system (here discretized neuronal systems), the time-domain variations of (a) the hidden states $\mathbf{x}_k$ whose dynamics are dictated by describing equations of motion (firing rates) (b) the unknown model parameters $\boldsymbol{\theta}$ which dictate the characteristics functions of the process **f** and the observation function **g** and (c) the unknown inputs **u** which drive the overall system. The dynamical system, or equations of motion, describing variations is summarized as follows.

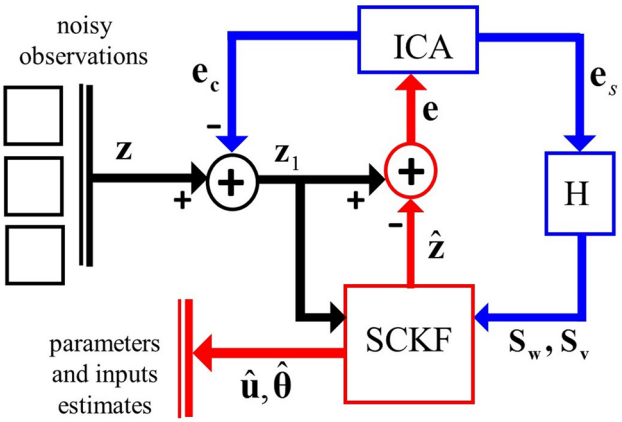

Fig. 3 Block diagram showing the main component of the proposed algorithm. SCKF is the Kalman filter (and smoother) that is applied in two stages. The components with blue lines are applied at the second and third stages. Starting with noisy output observations **z**, The SCKF produces a first-stage estimates and output predictions $\hat{\mathbf{z}}$. In the second stage, ICA is the independent component algorithm that will produce form the prediction error (stage 1: Eq. 9, stage 2: 22), a decomposition (Eq. 21) into a common error component e_c (Eq. 19) and channel-specific error components e_s (Eq. 20). The block H is applied in stage 3 and estimates the new Process covariances for the input S_v and the parameters S_w based on the channel specific error and the first stage estimates (Eqs.23–40). These estimates are used in the fourth stage SCKF and smoother to produce final estimates of the input $\hat{\mathbf{u}}$ and model parameters $\hat{\boldsymbol{\theta}}$)

$$\mathbf{u}_k = \mathbf{u}_{k-1} + \mathbf{w}_{k-1} \tag{1}$$

$$\boldsymbol{\theta}_k = \boldsymbol{\theta}_{k-1} + \boldsymbol{v}_{k-1} \tag{2}$$

$$\mathbf{x}_k = f\left(\mathbf{x}_{k-1}, \boldsymbol{\theta}_{k-1}, \mathbf{u}_{k-1}\right) + \boldsymbol{\mu}_{k-1} \tag{3}$$

$$\mathbf{z}_k = h(\mathbf{x}_k, \mathbf{u}_k) + \boldsymbol{\eta}_k \tag{4}$$

where $\mathbf{w}_{k-1}$ and $\boldsymbol{v}_{k-1}$ are independent Gaussian-distributed sequences. Recent developments in Kalman filtering such as the square root cubature Kalman filter (SCKF) and its smoother (SCKS) allow to estimate the states, parameters and inputs of the above system based on a set of observations $\mathbf{z}$ where by the covariances of the corresponding prediction errors are minimized.

We here consider the specific problem of fitting a neuronal network of n_z interconnected areas, each represented by a Modified Wendling models (see "Appendix"), using n_z channel EEG recordings. An augmented state vector contains, in addition to the hidden states, the unknown connection parameters and unknown driving inputs $\mathbf{u}_k$. We assume that $\mathbf{u}_k$ has one driving input per area ($n_u = n_z$) and we note that the state vector $\mathbf{x}_k$ consists of 10 discretized states in each area, of Eqs. (46–51), for a total of $n = 10n_z$ states. Finally, the parameter vector $\boldsymbol{\theta}_k$ consists of the inter-areal connections K_{rq} in Eq. (55). An area r with all-to-all connections has $n_\theta^r = n_z - 1$ for a total of $n_\theta = \sum_r n_\theta^r = n_z(n_z - 1)$ parameters. Denote the parameter vector for all connections emanating from area r by $\boldsymbol{\theta}_k^r = \left[\, K_{r1} \;\; K_{r2} \;\; \ldots \;\; K_{rn_z} \,\right]^\intercal$, $r = 1 \ldots n_z$ then $\boldsymbol{\theta}_k$ is organized as

$$\boldsymbol{\theta}_k^\intercal = \left[\, \boldsymbol{\theta}_k^{1\intercal} \;\ldots\; \boldsymbol{\theta}_k^{r\intercal} \;\ldots\; \boldsymbol{\theta}_k^{n_z\intercal} \,\right] \tag{5}$$

with the corresponding process noise covariance $\boldsymbol{\Lambda}_{V,k}$ (assumed diagonal) having the square root$\boldsymbol{\Lambda}_{V,k} = \boldsymbol{S}_{V,k}\boldsymbol{S}_{V,k}^\intercal$ and organized as

$$diag(\boldsymbol{S}_{V,k}) = \left[\, diag(\boldsymbol{S}_{V,k}^1) \;\ldots\; diag(\boldsymbol{S}_{V,k}^r) \;\ldots\; diag(\boldsymbol{S}_{V,k}^{n_z}) \,\right] \tag{6}$$

where $\boldsymbol{S}_{V,k}^r \in \mathbb{R}^{n_\theta^r \times n_\theta^r}$ is the square root of covariance matrix of the process noise for $\boldsymbol{\theta}_k^r$, and $diag(M)$ is a row vector containing the diagonal elements of the matrix M. Similarly, the square root covariance matrix of the input is organized as

$$diag(\boldsymbol{S}_{W,k}) = \left[\, S_{W,k}^1 \;\ldots\; S_{W,k}^r \;\ldots\; S_{W,k}^{n_z} \,\right] \tag{7}$$

Finally the output vector $\mathbf{z}_k = \left[\, z_k^1 \;\; z_k^2 \;\; \ldots \;\; z_k^{n_z} \,\right]$ where the scalar z_k^r is the noisy discrete time observation of block r in Eq. (62).

The SCKF algorithm is applied to estimate the augment state vector based on a single set or n_z channel recordings. With the filter performance dependent on the process noise square root covariances matrices $\mathbf{S}_{V,k}, S_{V,k}$, these are updated adaptively in time using the Robins–Monroe procedure as was adopted in (Havlicek et al. 2011). Furthermore, the observation noise covariance is assumed diagonal, with the square root terms update recursively in time with innovations dependent on the prediction error and a forgeting factor λ_r

$$diag\left(\mathbf{S}_{R,k}\right) = \left[\left(1 - \lambda_r\right)(diag(\mathbf{S}_{R,k-1}))^2 + \lambda_r diag\left(\mathbf{e}_k\mathbf{e}_k^\intercal\right)\right]^{\frac{1}{2}} \tag{8}$$

The SCKS block thus far has been readily implemented elsewhere (Havlicek et al. 2011; Karameh et al. 2014; Madi and Karameh 2017), with its output being estimates of the parameters and the inputs. However, and because the (square root) covariances $\mathbf{S}_{V,k}$ and $\mathbf{S}_{W,k}$ are the main regulators of its accuracy, we will introduce herein a novel approach to "learn" optimal adjustments of these covariances based on the prediction error.

Role of the Prediction Error For the state space description given in Eqs. (59–62), assume that Kalman filtering algorithm provided an initial estimate of the hidden states $\hat{\mathbf{u}}_k$, $\hat{\boldsymbol{\theta}}_k$, and $\hat{\mathbf{x}}_k$ as well as of the square root covariance matrices of $\mathbf{S}_{V,k}$ and $\mathbf{S}_{W,k}$ (Eqs. 6 and 7). Correspondingly, the output prediction is $\hat{\mathbf{z}}_k = h(\hat{\mathbf{x}}_k, \hat{\mathbf{u}}_k)$ with a prediction error $\mathbf{e}_k$ given by

$$\mathbf{e}_k = \mathbf{z}_k - \hat{\mathbf{z}}_k = h(\mathbf{x}_k, \mathbf{u}_k) - h(\hat{\mathbf{x}}_k, \hat{\mathbf{u}}_k) + \boldsymbol{\eta}_k \tag{9}$$

$$\mathbf{e}_k = \tilde{\mathbf{z}}_k + \boldsymbol{\eta}_k. \tag{10}$$

As is common for prediction error methods, a good model fit requires that the residual term $\tilde{\mathbf{z}}_k$ in the prediction error is made negligible or $h(\hat{\mathbf{x}}_k, \hat{\mathbf{u}}_k) \approx h(\mathbf{x}_k, \mathbf{u}_k)$, leaving only the noise term $\boldsymbol{\eta}_k$ which consists of independent, identically Gaussian distributed samples across time (i.i.d.).

On the other hand, inaccurate estimation of the states and/or inputs could result in a non-negligible structured residual $\tilde{\mathbf{z}}_k$ that consists of non-independent samples in time and across dimensions. That is, $E(\tilde{\mathbf{z}}_k\tilde{\mathbf{z}}_k^T)$ can have nonzero off-diagonal elements, and $E(\tilde{\mathbf{z}}_k\tilde{\mathbf{z}}_l^T)$ can be nonzero for $k \neq l$. Furthermore, this error might actually contain information that is related to the input of the system [In a linear system setting, we say that the residual error is correlated with the input—a problem that is commonly addressed using instrumental variable methods or two-step least squares which leads to reduction in this correlation.]

We herein address correcting for errors that result from input estimation inaccuracies, or exploit the structure of the error to improve on the input estimates. This procedure is conducted under the following three assumptions (Fig. 4):

A1 Some of the external input components arriving to different areas are specific to these areas. This is reasonable since the outputs (channel recordings) are collected from neural areas that are located sufficiently apart and can thus receive disparate inputs from outside the recorded network.

A2 An input component that is specific to area i, denoted by s^i, exhibits the highest correlation with the output z_k^i of that area, and smaller correlations with other outputs within the network z_k^j, $j \neq i$. This is expected since the effect of an input on neural population output is generally decremented as the signals (firing activity) moves along a processing path across multiple synaptic connections (one or more hops in the network).

A3 The residual error $\tilde{z}_k^i$ in a given area, being directly related to the area output z_k^i, is also highly correlated to the specific input s^i to that area. Furthermore, and since $E(\tilde{\mathbf{z}}_k \tilde{\mathbf{z}}_k^T)$ is not diagonal, $\tilde{z}_k^i$ is also correlated to at least one of the inputs s^j arriving to the other areas.

A4 The unknown external area-specific inputs vary at a slower time scale in relation to the local dynamics. That is, in the frequency domain, the bandwidth of the unknown inputs to a given area is smaller than the bandwidth of the local signals produced within that area. Note that this constraint acts to reduce the ill-posedness of nonlinear blind identification, and that its physiological justification for brain networks, namely that local processing occurs at high frequency while global communication is generally at lower frequencies, is addressed in "Discussion".

We desire to update the initial estimate of the input s^j based on the residual term $\tilde{z}_k^i, i, j = 1 \ldots n_z$. This relationship, however, is in general nonlinear and time varying, particularly since (i) the system itself is nonlinear and (ii) the Kalman filter algorithm is adaptive in time ($\tilde{\mathbf{z}}_k$ is also related to $\hat{\mathbf{z}}_k$). Therefore, while it is difficult to directly find this relationship, we seek to devise an indirect update procedure of the input estimate $\mathbf{u}_k$ to account for $\mathbf{s}_k$. In what follows, we will find independent factors in the residual signals using ICA decomposition which are then mapped to unique channels as unexplained, channel specific errors. These errors are then mapped back to an additional input levels to their respective channels. From this input, we devise a measure to update the process noise covariance of the various inputs. Finally, the updated covariances are used in a subsequent run of the SCKF algorithm to improve on the input estimates.

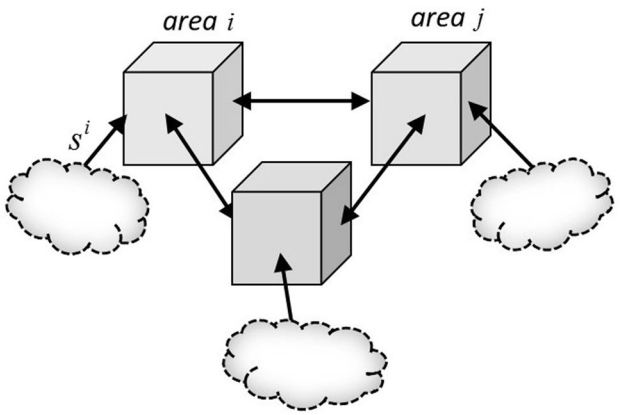

Fig. 4 Schematic of interconnected areas within a subnetwork. Inputs s^i are unknown and inter-areal connections are to be estimated

The ICA Block

We will here map output error components to the various areas, as follows.

1 *Decompose the residual error using ICA* Following an initial prediction of the state and parameter vectors using cubature Kalman filter algorithm, the prediction error signal vector is buffered over non-overlapping time windows of length L as follows:

$$\mathbf{E}_m = \left[\mathbf{e}^1 \ \ldots \ \mathbf{e}^{n_z} \right]^\intercal \tag{11}$$

where $\mathbf{e}^i = \left[e^i{}_{mL+k-L+1} \ldots e^i{}_{mL+k-1} e^i{}_{mL+k} \right]^\intercal$, for $i = 1 \ldots n_z, m = 0, 1, \ldots, m_o$ and m_o being the total number of blocks. For a given time window m, and suppressing the index m for notational simplicity, $\mathbf{E} \triangleq \mathbf{E}_m$ is an n_z-dimensional signal of length L that is decomposed [using the fast ICA algorithm (Hyvarinen 1999)] into n_z independent components $\mathbf{c}_j \in \mathbb{R}^L$, $j = 1 \ldots n_z$ which linearly mix to give the original signal $\mathbf{E}$ as

$$\mathbf{E} = \mathbf{AC} \tag{12}$$

where

$$C = \begin{bmatrix} \mathbf{c}_1^\intercal \\ \vdots \\ \mathbf{c}_{n_z}^\intercal \end{bmatrix} \text{ and } \mathbf{A} = \left[\mathbf{a}_1 \ \mathbf{a}_2 \ \ldots \ \mathbf{a}_{n_z} \right] = \begin{bmatrix} a_{1,1} & a_{1,2} & \ldots & a_{1,n_z} \\ a_{2,1} & a_{2,2} & \ldots & a_{2,n_z} \\ \vdots & \ddots & & \vdots \\ a_{1,n_z} & a_{2,n_z} & \ldots & a_{n_z,n_z} \end{bmatrix}$$

where $\mathbf{a}_j = \left[a_{1,j} \ a_{2,j} \ \ldots \ a_{n_z,j} \right]^\intercal$ is the contribution of independent component c_j to various channels i.

2 *map the ICA decomposed error to specific nodes* Considering the contribution of independent component c_j to the prediction error signal in a given node or channel i_o, we make a distinction between the following cases: (a) the contribution is almost the same across all channels or c_j is a common component that results in an error $\mathbf{e}_c$; (b) the contribution is clearly skewed towards a single channel i or c_j results in a channel-specific component $\mathbf{e}_s^i$, and (c) the contribution is neither common or specific, or c_j results in a mixture component $\mathbf{e}_m^i$.

$$\mathbf{e}^i = \mathbf{e}_c + \mathbf{e}_s^i + \mathbf{e}_m^i \quad i = 1 \ldots n_z \tag{13}$$

Referring to the weight matrix $\mathbf{A}$, common errors (case a) occur when $|a_{i_o j_o}| \approx |a_{ij_o}|, \forall i$, while channel specific errors (case b) occur when $|a_{i_o j_o}| >> |a_{ij_o}|, i \neq i_o$. One way to quantify the above distinctions is by defining a

uniformity factor for component c_j as the ratio of the mean of vector $\mathbf{a_j}$ to its standard deviation

$$\tilde{a}_j = \frac{\mu_j^a}{\sigma^{a_j}}, \tag{14}$$

then $\tilde{a}_j$ is generally large whenever the standard deviation is small, that is when the component contributes uniformly or roughly equally to various channels. Conversely, $\tilde{a}_j$ will decrease for large standard deviation σ^{a_j} which occurs in cases of skewed c_j contributions. For comparison among components, we define the relative contribution as

$$\overline{a}_j = \frac{|\widetilde{a}_j|}{\sum_{j=1}^{n_z} |\widetilde{a}_j|} \tag{15}$$

We are particularly interested in identifying those channels where $\mathbf{e}_s^i$ is nonzero and ultimately to devise ways to reduce it. We therefore seek to find, whenever possible, a set of one-to-one associations between the components c^j and the channels r^i. We suggest the following procedure:

a) Starting with the sets of indices for all channels $\mathcal{I} = \{ 1, 2, i, \ldots, n_z \}$, and for all independent components $\mathcal{J} = \{ 1, 2, j, \ldots, n_z \}$, we identify the c_j indices that are common noise terms as $\mathcal{J}_c$ and the set of candidate indices for channel-specific noise terms as $\mathcal{J}_s$ as those components that satisfy the following uniform or skewed relative contribution criteria, respectively:

$$\mathcal{J}_c = \{ j \in \mathcal{J} \mid \overline{a}_j \geq \nu_c \} \tag{16}$$

$$\mathcal{J}_s = \{ j \in \mathcal{J} - \mathcal{J}_c \mid \overline{a}_j < \nu_s \} \tag{17}$$

for some thresholds ν_s, ν_c ($\nu_c \geq \nu_s$).

b) Assuming that $\mathcal{J}_s$ has p_s elements (satisfying Eq. 17), $p_s \leq n_z$), and starting with candidate channels $\mathcal{I}_s = \mathcal{I}$, then a total of p_s one-to-one associations or pairings (i_l, j_l) are made as

$$(i_l, j_l) = \arg\max_{i \in \mathcal{I}, j \in \mathcal{J}_s} \{a_{ij}\} \quad l = 1 \ldots p_s \tag{18}$$

where the sets of candidate channel-specific indices are decremented after each pairing assignment, that is $\mathcal{I}_s = \mathcal{I}_s - \{ i_l \}$, $\mathcal{J}_s = \mathcal{J}_s - \{ j_l \}$.

c) Finally, the residual terms can now be determined. For the common residual terms $\mathbf{e}_c^i \in \mathbb{R}^L$, this follows

$$\mathbf{e}_c^i = a_{i,j}\mathbf{c}_{j_c} \quad i \in \mathcal{I},\ j_c \in \mathcal{J}_c \tag{19}$$

For the channel-specific residuals $\mathbf{e}_s^l \in \mathbb{R}^L$, $l \in \mathcal{I}_s$, this follows

$$\mathbf{e}_s^i = a_{i,j}\mathbf{c}_j \quad (i,j) \in \mathcal{P} \tag{20}$$

where $\mathcal{P}$ is the set of pairings found before

$$\mathcal{P} = \{ (i_o, j_o), (i_1, j_1), \ldots, (i_{p_s}, j_{p_s}) \}$$

The H Block

We now seek to update the process covariances for the input and parameters ($\mathbf{S}_{V,k}$ and $\mathbf{S}_{W,k}$) based on the knowledge of the common and channel-specific error components computed in Eqs. (19) and (20). Note that the trajectories of the inputs and parameters in the augmented state-space of Eqs. (59) and (60) are governed by zero mean processes with generally unknown time-dependent covariances $\mathbf{\Lambda}_{v,k}$ and $\mathbf{\Lambda}_{W,k}$, or alternatively by the square root matrices $\mathbf{S}_{V,k}$ and $\mathbf{S}_{W,k}$ (Eqs. 6, 7). Accordingly, we propose to update $\mathbf{S}_{W,k}$ and $\mathbf{S}_{V,k}$ based on an area-specific desired signal which is derived from the computed errors $\mathbf{e}_s^i$, $\mathbf{e}_c^i$ and then perform another Kalman iteration to re-estimate the inputs. In this way, we are not using the channel-specific errors to directly estimate the channel inputs, but rather to measure how far the estimate of that input is from achieving a smaller error at the output over a specified time frame. This measure is translated to changes in the time-dependent gains that the Kalman filter utilizes in re-estimating the inputs and the parameters over the same frame.

Consider the prediction error for a given channel i

$$\mathbf{e}^i = \mathbf{z}^i - \hat{\mathbf{z}}^i = \mathbf{e}_c + \mathbf{e}_s^i + \mathbf{e}_m^i \tag{21}$$

We will rewrite $\mathbf{e}^i \in \mathbb{R}^L$ as follows

$$\begin{aligned} \mathbf{e}_m^i &= \underbrace{\left(\mathbf{z}^i - \mathbf{e}_c^i\right)}_{\triangleq\ \mathbf{z}_1^i} - \underbrace{\left(\hat{\mathbf{z}}^i + \mathbf{e}_s^i\right)}_{\hat{\mathbf{z}}_s^i} \end{aligned} \tag{22}$$

Here, the signal $\mathbf{z}_1^i$ can be thought of as an artificially synthesized, cleaned version of the actual output $\mathbf{z}^i$ since the common error term across all channels has been removed. Also, the signal $\hat{\mathbf{z}}_s^i$ is a desired corrected estimate of the output and is composed of the original Kalman prediction and a channel-specific deviation term $\mathbf{e}_s^i$ as found in Eq. (20).

The desired channel-specific input update, therefore, is one which reduces $\mathbf{e}_m^i$ by effectively allowing the updated Kalman estimate $\hat{\mathbf{z}}_s^i$ to closely match the cleaned signal $\mathbf{z}_1^i$. A geometric interpretation of this idea is given in Fig. 5

In general, an accurate fit of the desired output estimate $\hat{\mathbf{z}}_s^i$ results from an update of both the states and inputs to the system, that is, $\hat{\mathbf{z}}_s^i = \mathbf{h}^i(\hat{\mathbf{x}}_s, \hat{\mathbf{u}}_s)$.

However, and since the input to channel i has highest correlation to the corresponding output (assumption A2), we seek to approximate the desired corrected output $\hat{\mathbf{z}}_s^i$ based on updating only the input signal to channel i, that is

$$\hat{\mathbf{z}}_s^i \approx \mathbf{h}^i\left(\hat{\mathbf{x}}, \tilde{\mathbf{u}}_{s,i}\right) \tag{23}$$

where

$$\begin{aligned} \tilde{\mathbf{u}}_{s,i} &= \left[\,\hat{\mathbf{u}}^1\ \hat{\mathbf{u}}^2\ \ldots\ \tilde{\mathbf{u}}_s^i\ \ldots\ \hat{\mathbf{u}}^{n_z}\,\right], \\ \tilde{\mathbf{u}}_s^i &= \hat{\mathbf{u}}^i + \boldsymbol{\delta}_u^i \end{aligned} \tag{24}$$

Upon restricting the input update $\boldsymbol{\delta}_u^i$ to be small enough (see later sections), Eq. (23) can be approximated as a small-signal deviation from the original operating point:

$$\mathbf{h}^i\left(\hat{\mathbf{x}}, \tilde{\mathbf{u}}_{s,i}\right) \approx \mathbf{h}^i(\hat{\mathbf{x}}, \hat{\mathbf{u}}) + \mathbf{h}_u^i(\boldsymbol{\delta}_u^i) = \hat{\mathbf{z}}^i + \mathbf{h}_u^i(\boldsymbol{\delta}_u^i) \tag{25}$$

where $\mathbf{h}_u^i(.)$ is the direct map from input to channel i to its output with the contribution of all other inputs and states held fixed. Finally, and comparing with Eq. (22) results in requiring the input update to approximate the channel-specific error term:

$$\mathbf{e}_s^i \approx \mathbf{h_u}^i(\boldsymbol{\delta}_u^i) \tag{26}$$

To find the input from the channel-specific error in Eq. 26 requires the inversion of the map $\mathbf{h}_u^i(.)$ and hence depends on the properties of the latter. In our case, $\mathbf{h}_u^i(.)$ is a discrete-time dynamical map derived from the continuous-time model of the synapse that relates the input firing rate $u(t)$ to the output field potential $y(t)$ from pyramidal cells. The synapse is modeled here using a second-order ordinary differential equation

$$\ddot{y}(t) = Aau(t) - 2a\dot{y}(t) - a^2 y(t) \tag{27}$$

which corresponds to a Linear Time Invariant (LTI) system with an impulse response $Aate^{-at}, t \geq 0, A, a > 0$. Since this LTI system is low-pass in nature, details of the input at high frequencies are lost at the output. Hence, finding a unique input for a recorded output is in general not possible.

We therefore devised a modified inverse function $\mathbf{g}_u^i$ that (i) estimates low frequency deviations in δu and (ii) limits the deviation so that the output remains in the vicinity of its current operating point

3 *Approximate the input inverse map* For a given output

$$\boldsymbol{\delta}\mathbf{u}^i = \mathbf{g}_u^i\left(\mathbf{e}_s^i\right) \tag{28}$$

corresponding vector $\hat{\boldsymbol{\delta}}_u^i$ is estimated in two steps: First, a discrete time inverse dynamic map $\mathbf{u}^o = \mathbf{h}_{inv}^i(\mathbf{y}^o)$ is derived from Eq. (27) (here $\mathbf{y}^o = \mathbf{e}_s^i$)using Euler approximation of the derivative operations with a time step Δ_t, resulting in:

$$u_k^o = \frac{1}{Aa\Delta_t^2}\left[y_{k+1}^o + \left(a^2\Delta_t^2 + 2a\Delta_t - 2\right)y_k^o + \left(1 - 2a\Delta_t\right)y_{k-1}^o\right] \tag{29}$$

which is reordered to give the inverse map:

$$\begin{aligned} \mathbf{u}^o &= \mathbf{h}_{inv}^i(\mathbf{y}^o) \\ u_k^o &= \frac{1}{Aa\Delta_t^2}\left[\left(y_{k+1}^o - y_k^o\right) + \left(1 - 2a\Delta_t\right)\left(y_{k-1}^o - y_k^o\right)\right. \\ &\quad \left. + a^2\Delta_t^2 y_k^o\right] \end{aligned} \tag{30}$$

Second, and since the difference operations cause amplification of high frequency noise in $\mathbf{y}^o$, a smoothed estimate of $\mathbf{u}^o$ is obtained by passing u_k through a low-pass FIR filter (denoted by g_{lpf}, with a cutoff = 100 Hz).

$$\hat{\boldsymbol{\delta}}_u^i = \mathbf{g}_{lpf}(\mathbf{u^o}) \tag{31}$$

4 *Limit the plausible range of input* the smoothed estimate $\hat{\boldsymbol{\delta}}_i^u$ was computed from $\mathbf{e}_s^i$, the channel-specific prediction error obtained upon fitting the raw (uncleaned) output signal $\mathbf{z}^i$ (Eq. 21). With the ICA decomposition, however, it was possible earlier to derive a common noise component $\mathbf{e}_c^i$ which is a contributing to $\hat{\mathbf{z}}^i$ and hence indirectly affected the Kalman-derived prediction error $\mathbf{e}_s^i$ and subsequently $\hat{\boldsymbol{\delta}}_k^u$ computed above. In other words, we intend for our new estimate $\mathbf{z}_s^i$ in Eq. (22) to approximate the cleaned signal $\mathbf{z}_1^i$, thereby reducing $\mathbf{e}_m^i$. This can be achieved upon reapplying the Kalman estimation to the clean signal $\mathbf{z}_1^i$ and finding $\mathbf{e}_s^i$ once more, a procedure that can be time consuming. We here propose a simpler ad-hoc approach as follows:

i. Estimate the total input signal $\mathbf{u}_1^i$ that could result in the clean signal $\mathbf{z}_1^i$ using a procedure similar to the inverse map and smoothing calculations in Eqs. (28–31). Specifically, let $\mathbf{u}_c^i = \mathbf{h}_{inv}^i(\mathbf{z}_1^i)$, then

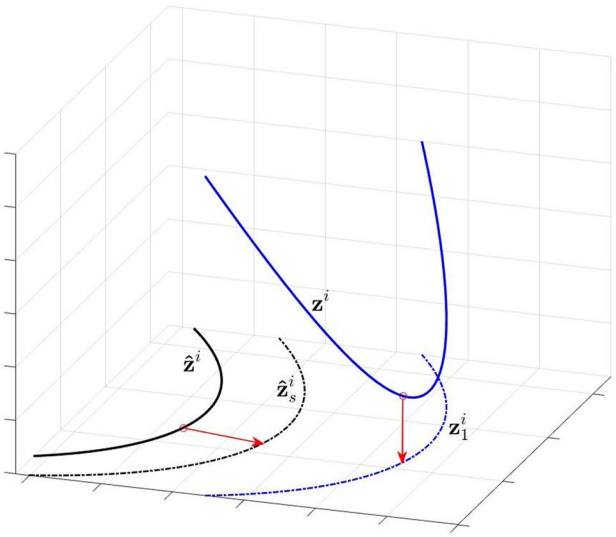

Fig. 5 Geometric depiction of the residual-based update of the output estimate. The original Kalman output prediction $\hat{\mathbf{z}}_k^i$ is modified to track a cleaned observation signal $\mathbf{z}_1^i$. See text for details

$$\mathbf{u}_1^i = \mathbf{g}_{lpf}(\mathbf{u}_c^i) \tag{32}$$

ii. Limit the deviation $\boldsymbol{\delta}_u^i$ to a a small fraction ρ of that obtained total signal $\mathbf{u}_1^i$, that is

$$\boldsymbol{\delta}_u^i = \hat{\boldsymbol{\delta}}_u^i \times \max\left(1, \frac{\|\hat{\boldsymbol{\delta}}_u^i\|}{\rho\|\mathbf{u}_1^i\|}\right) \tag{33}$$

where $\|\mathbf{x}\|$ is the vector 2-norm of $\mathbf{x} \in \mathbb{R}^L$ and $\rho = 0.02$. Note that this step also insures that the small signal approximation conducted earlier is not violated.

With the signal $\boldsymbol{\delta}_u^i$ collected for the various data windows of size L , we finally reconstruct a desired deviation signal $\mathbf{d}^i \in \mathbb{R}^N$ as a concatenation of the window-based $\boldsymbol{\delta}_u^i$ calculations performed in steps 1 and 2. That is, and denoting $\boldsymbol{\delta}^i$ in window m as $\boldsymbol{d}_m^i$, then

$$\mathbf{d}^\mathbf{i} = \left[\begin{array}{ccc} \boldsymbol{\delta}_{u,1}^{i\ \mathsf{T}} & \boldsymbol{\delta}_{u,2}^{i\ \mathsf{T}} & \boldsymbol{\delta}_{u,m_o}^{i\ \mathsf{T}} \end{array}\right]^\mathsf{T} \tag{34}$$

Subsequent steps are done over the whole length N of the data.

5 *Finally, update the process noise covariances* We here utilize the local character of the desired signal d^i, in the vicinity of a time sample k for a given channel i ($i = 1 \ldots n_z$, $k = 1 \ldots N$) to scale the diagonal elements of $\mathbf{S}_{W,k}$ and $\mathbf{S}_{V,k}$ that were initially computed adaptively in the SCKF algorithm.

Let $\mathbf{D} \in \mathbb{R}^{N \times n_z}$ be the matrix of desired signals of all the channels over all time samples. After segmenting the data into nonoverlapping time windows of size L_1, let $\mathbf{d}_l^i \in \mathbb{R}^{L_1}$ be the desired data signal of channel i in time window l, or $\mathbf{d}_l^i = \left[d_{(l-1)L_1+1}^i \ \ldots \ d_{lL_1}^i\right]^\mathsf{T}$. Similarly, let $\mathbf{D}_l \in \mathbb{R}^{L_1 \times n_z}$ be the desired data across all channels over window l. We proceed as follows:

i. *Determine significant deviations:* we here find the set of indices $\mathscr{L}_o$ for time windows in which larger deviation signals are noted using some thresholding scheme (for example, $l \in \mathscr{L}_o$ when the relative value of the induced two-norm of $\mathbf{D}_l$ is non negligible $\|\mathbf{D}_l\|_2 > 2\rho_1 \frac{L_1}{N}\|\mathbf{D}\|_2$).

Within all such windows, determine the subset of channels whose deviation signal $\mathbf{d}_l^i$ is significantly different from zero—as also can be verified using a threshold on the relative signal energy,

$$\mathscr{D}_c = \{(i,l) \in \mathscr{I} \times \mathscr{L}_o \mid \|\mathbf{d}_l^i\| > 2\rho_2 \frac{L_1}{N}\|\mathbf{d}^i\|\} \tag{35}$$

(here $\rho_1 = \rho_2 = 0.02$).

ii. *Determine the amplification factors*: Since the deviations are assumed to keep the system close to current operating point, we seek to find scaling factors based on measuring the linearity in mapping the Kalman-predicted inputs $\hat{\mathbf{u}}_l^i$ to the desired deviation signal $\mathbf{d}_l^i$ in $\mathscr{D}_c$. Specifically we compute the Weiner estimate of the transfer function

$$\tilde{H}(\Omega) = \frac{P_{ud}(\Omega)}{P_u(\Omega)}$$

where $P_u(\Omega)$ and $P_{ud}(\Omega)$ are the auto-spectrum of $\hat{\mathbf{u}}_l^i$ and the cross spectrum of $\hat{\mathbf{u}}_l^i$ and $\mathbf{d}_l^i$ respectively

Since $\tilde{H}(\Omega)$ is used to linearly predict the desired signal from the input. It can therefore be used to find the auto-spectrum $\tilde{P}_d(\Omega)$ of this predicted desired signal as

$$\tilde{P}_d(\Omega) = |\tilde{H}(\Omega)|^2 P_u(\Omega) \tag{36}$$

Finally, we devise a metric to measure the approximative power of this linear prediction as

$$C(\Omega) = \frac{\sum_{\Omega \in \mathbf{\Omega}_r} \tilde{P}_d}{\sum_{\Omega \in \mathbf{\Omega}_r} P_d} \tag{37}$$

where the power contributions in both actual and predicted signals are summed over a given range of frequencies $\mathbf{\Omega}_r$. This range is selected to contain physically relevant information, here as $\mathbf{\Omega}_r = [0 - 0.3\,\text{Hz}]$. With $0 \le C(\Omega) \le 1$, the amplification factor is then determined as

$$\alpha_l^i = 1 + K_o \sum_{\Omega} C(\Omega) \quad (i,l) \in \mathscr{D}_c \tag{38}$$

with $K_o \ge 1$, $\alpha_l^i = 1 + \epsilon$, $\epsilon << 1$ for $(i,l) \notin \mathscr{D}_c$ (here $K_o = 10, \epsilon = 0.001$)

iii. *Scale the process noise covariances:*For all samples k located within the time window l (that is, $(l-1)L_1 + 1 \le k \le lL_1$) we now adjust the process noise covariance of the inputs to all the channels $i = 1 \ldots n_z$ as follows

$$S_{W,k}^i = \alpha_l^i S_{W,k}^i \tag{39}$$

We also attenuate the process noise covariance of all the connections emanating from channel i as follows:

$$diag(\mathbf{S}_{V,k}^i) = \left(\alpha_l^i\right)^{-1} diag(\mathbf{S}_{V,k}^i) \tag{40}$$

Overall Flow of the Algorithm

With the covariances modified as specified in Eqs. (39, 40). We now outline the overall procedure of estimating the augmented state vector.

1. *Iterative Kalman Filtering* For a given data set consisting of n_z total channels, the SCKF algorithm presented earlier is applied iteratively as follows: Starting with a random initial priors on the values on the augmented states $\hat{\mathbf{x}}_0$ and on the process noise covariance for the augmented states, an initial estimate of the state vector $\hat{\mathbf{x}}_k, k = 1 \ldots N$, is determined for iteration 1, denoted by $\hat{\mathbf{x}}_k^{(1)}$. Denote similarly the corresponding square root covariance matrices for the states $\mathbf{S}_k^{(1)}$. Since the process is assumed non stationary, new time-locked trials (or realizations) of the process are generally required to correctly estimate the state vector. For a single trial we proceed by repeating presentation of the same data (or passes in the time data) to the filter, akin to iterative presentation of data while training a neural network. For a new pass, estimates for various quantities at $k = 0$ are taken as the final estimates of the previous pass or $\hat{\mathbf{x}}_0^{(j+1)} = \hat{\mathbf{x}}_N^{(j)}\, \mathbf{S}_0^{(j+1)} = \mathbf{S}_{\mathbf{x}N}^{(j)}, j \geq 1$. For a given pass, the log likelihood of the estimates L is computed as a performance measure.

   Note that, and since the process covariance $\mathbf{S}_q$ depended on the error, very fast and frequent variations in the parameter vector are not physically justifiable but can still reduce prediction error thereby incorrectly driving the log likelihood of the estimate larger. Therefore, one would like to enforce a limit on rate of change in the parameter vector with subsequent presentations of the data to a range that is justifiable experimental evidence. This is practically done by reducing the prior on the process covariance and the corresponding forgetting factor λ for iterations $j \geq 2$. The iteration process can be terminated after reaching a maximum value of L or reaching a preset number of iterations where by changes in the variations in the parameters between two successive data presentations is minimal. Denote the final estimate of the state vector by $\hat{\mathbf{x}}_k^{(j_o)}$ and its corresponding square root covariance by $\mathbf{S}_k^{(j_o)}$.
2. *ICA adjustment* For the final estimate obtained above ($\hat{\mathbf{x}}_k^{(j_o)}$, $\mathbf{S}_k^{(j_o)}$ and $\mathbf{S}_{\mathbf{q}k}^{(j_o)}$) the ICA adjusting is applied to adjust the covariances $\mathbf{S}_{\mathbf{q}k}^{a}$. A final forward presentation of the data is performed with the updated $S_{v,k}$ and $S_{w,k}$ (Eqs. 39, 40).
3. *Kalman Smoothing* The ICA-adjusted forward run is finally subjected to a (backward) pass of a Square root Kalman smoother (SCKS) to obtain time-smoothed estimates for the augmented state vector.

Algorithm 1 ICA-based residual-adjusted Kalman filtering

Initialization: for $k = 1 \ldots N$ and for n_z channels, obtain SCKF estimate of parameters $\hat{\theta}_k$, inputs $\hat{\mathbf{u}}_k$, their square-root process noise covariances ($\mathbf{S}_{V,k}$, $\mathbf{S}_{W,k}$ respectively), outputs $\hat{z}_k$, and prediction errors e_k.

function FIND DESIRED INPUT($\mathbf{z}_k$,$\mathbf{e}_k$)
  for $m = 0, 1, \ldots$ **do**
    Buffer $\mathbf{e}_k$ into $\mathbf{E}$, for time $k = (m-1)L+1 : mL$ (windows of size L, eq. 11).
    Apply fast ICA decomposition of $\mathbf{E}$ to compute n_z independent components $\mathbf{c}^i$
      such that $e_k^i = \sum_{j=1}^{n_z} a_{i,j} c_k^j$ (eq 12)
    for each component c_j **do**
      Find the relative contribution factor $\bar{a}_j$ (eq. 15).
      if $\bar{a}_j > \nu_c$ (eq.16) **then** ▷ *c_j a common component.*
        assign j to common-component set $\mathscr{J}_c$.
        reconstruct common error $e_c^i = \sum_{j \in \mathscr{J}_c} a_{i,j} c_j$ (eq. 19)
      else if $\bar{a}_j < \nu_s$ (eq. 17) **then** ▷ *c_j is channel-specific*
        map c_j to single channel i_o (eq. 18).
        reconstruct channel-specific error $e_s^{i_o} = a_{i_o,j} c_j$ (eq. 20).
        assign i_o to the set with channel-specific inputs $\mathscr{I}_s$
      end if
    end for
    for $i \in \mathscr{I}_s$ **do** ▷ *Propagate the error $\mathbf{e}_s^i$ back to deviations δ_u^i in the input to channel i*
      find $\hat{\delta}_u^i$ as a smoothed inverse map from e_s^i (eq. 30, 31).
      find u_1^i as a smoothed inverse map from cleaned observation signal $z_1^i = z^i - e_c^i$ (eq. 32).
      find δ_u^i by limiting $\hat{\delta}_u^i$ to small deviations from u_1^i (eq. 33).
    end for
    concatenate δ_u^i, $i \in \mathscr{I}_s$, in a desired input vector $\mathbf{d}^i$ (eq. 34).
  end for
end function

function UPDATE COVARIANCES($\mathbf{d}_k$,$\mathbf{u}_k$, $\mathbf{S}_{V,k}$,$\mathbf{S}_{W,k}$)
  for $l = 0, 1, \ldots$ **do**
    Buffer $\mathbf{d}_k$ and $\mathbf{u}_k$ into $\mathbf{d}_\mathbf{l}$ and $\mathbf{u}_\mathbf{l}$ for time $k = (l-1)L_1 + 1 : lL_1$,(windows of size L_1).
    if d_l^i is significant (eq. 35) **then**
      Find the best linear prediction $\hat{\mathbf{d}}_l^i$ of $\mathbf{d}_l^i$ from $\mathbf{u}_l^i$ (eq. 36)
      Determine the approximative power $C(\Omega)$ of $\hat{\mathbf{d}}_l^i$ over the relevant frequency domain (eq. 37).
      Find the amplification factor $\alpha_l^i = 1 + K_o \sum_{\Omega} C(\Omega)$, $K_o \geq 1$ (eq. 38).
    else
      Set $\alpha_l^i = 1$
    end if
    Amplify covariances of inputs to channel i (eq. 39): $S_{W,k}^i = \alpha_l^i S_{W,k}^i \qquad i = 1 \ldots n_z$
    Attenuate covariances of all connections from channel i (eq. 40): $diag(\mathbf{S}_{V,k}^i) = (\alpha_l^i)^{-1}\, diag(\mathbf{S}_{V,k}^i)$
  end for
end function

Apply SCKF algorithm starting from initial state vector ($\hat{\mathbf{x}}_0 = \hat{\mathbf{x}}_N$) and with the updated $\mathbf{S}_{V,k}$, $\mathbf{S}_{W,k}$

Simulated Network Model

An induced-seizure network model was setup to simulate four interconnected areas with corresponding EEG observations (Fig. 6a, c). Each area is composed of a modified Wendling model (46–54), with the following parameters: $c_{pe} = C = 135$; $c_{ep} = 0.8C$, $c_{ps} = 0.5C$; $d_{sp} = 0.25C$; $c_{pf} = 0.1C$; $d_{sf} = 0.45C$; $d_{fp} = 0.85C$; Synaptic dynamics were as given in original paper: $e = 2.5$; $x_o = 6$; $k = 0.560$, $A = 3.5$;$B = 35$;$G = 20$; $a = 100$; $b = 50$; $g = 500$ and sigmoidal constants (see "Appendix") $e = 2.5$; $x_o = 6$; $k = 0.560$; In the modified model, $d_{fs} = 0.99$; The delay of interconnections between the areas (Fig. 2) were $\tau_{12} = \tau_{21} = 1$ ms, $\tau_{14} = \tau_{41} = 2$ms,$\tau_{23} = \tau_{32} = 2$ ms, and $\tau_{43} = \tau_{34} = 1$ ms. These delay values are chosen to be within acceptable physical ranges and reflect a distance dependent absolute value (areas 1 and 2 are frontal while areas 3 and 4 are central, fronto-central connections are presumed longer than callosal connections). Note that we also tested the model for larger delays (4 and 8 ms, respectively) without finding any significant changes in the estimates of the parameter vector.

The connections $(K_{41}, K_{12}, K_{23}, K_{43}, K_{14}, K_{34})$ were set to (48, 60, 74.4, 48, 84, 64), respectively, for $0 \le t < 10$ s, with $K_{21} = 168$, $K_{32} = 136$ for $0 \le t < 5$ switching to $K_{21} = 120$ and $K_{32} = 200$ for $5 \le t < 10$ s with a sampling step of 1 ms. Finally, the long-range pyramidal to inhibitory connections L_{rq} were set to the same magnitude as pyramidal to pyramidal connections K_{rq}. Input firing rates were assumed of sinusoidal nature to two of the areas while they were constant to the other two (Fig. 6b). In all multi-trial simulations, additive randomized components were added to the input signals so that a range of firing behaviors were generated (signal to noise ratio between 6 and 22 dB). At the recording level, two scenarios were considered: a noise-free scenario, whereby the output $z(t)$ from the simulations were used for estimation, and a noisy observation scenario, where a common noise signal was added to all the channels. Here, the noise signal was scaled relative to the power of each channel such that the signal to noise ratio of 20 dB is maintained across all the electrodes.

For estimation purposes, and since the model is known, a single forward run of the SCKF (one iteration) was utilized and was noted to produce satisfactory results. Subsequent to this forward run, three variations were introduced: (a) a smoothing backward run using the standard Square-root Kalman Smoother algorithm (SCKS) (Havlicek et al. 2011); (b) an ICA-adjuster run followed by the SCKS and is termed Independent Component adjusting SCKS or RA-SCKS; (c) application of ICA algorithm to remove only common noise components in the error without applying covariance adjustment. This later step was separate the impact of adjusting from naive ICA application.

Implications of Connectivity Strength

The connection strength K_{rq} is a summary figure that represents the total amount of excitatory synaptic drive from area a ro area r. Hence it is a function of the proportion of neurons that are contributing an active synapse at any given time.

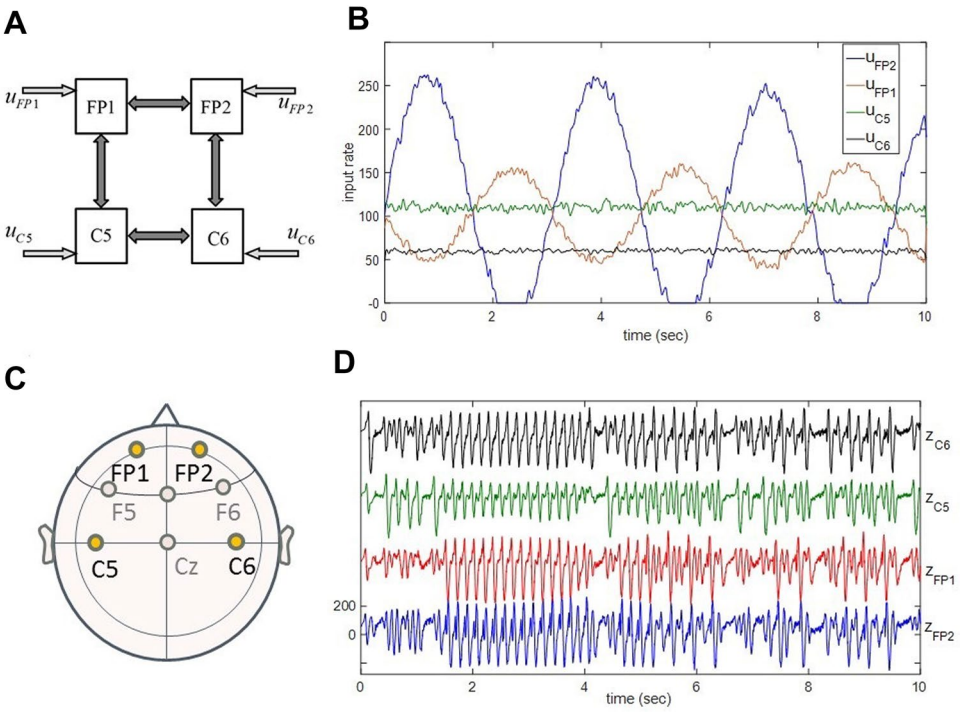

Fig. 6 Simulation example for induced seizures. **a** Schematic of four area network. **b** Simulated input firing rates. **c** Representative locations of the network in terms of the EEG channel locations. **d** Simulated output EEG of the four areas based on inputs given in **b**

For example, a low connectivity strength $K^o_{rq}(t)$ at a rate $r^o_q(t)$ (such as the one observed during anesthesia) implies that the a small proportion of neurons from the originating area q are firing at a rate r^o_q. Conversely, high connectivity $K^1_{rq}(t)$ at a rate $r^1_q(t)$ implies a large proportion of the total neural population in area q are firing at rate r^1_q. During a seizure, this proportion is very large since the ictal activity dominates the behavior of affected neurons (which in turn are firing at high burst-like rates).

Nonlinear Partial Directed Influence

Consider a network of size M where area-specific inputs u_r in Eq. (55), the excitatory drive from excitatory populations $x^{(r)}_p$, and the inter-areal connection (excitatory synaptic) strength between these populations K_{qr}, are either known or are estimated (as will be done using Kalman filters). We now proceed to compute a measure of the partial functional drive that a single area receives from other areas as follows.

Let $P_{qr}(t)$ be the average directed power of the excitatory drive that the local excitatory population of area r receives from area q computed over a time window of size N_w (here 0.5 s), given by

$$P_{qr}(t) = \frac{1}{N_w}\int_{t-N_w}^{t}\left(K_{qr}(\lambda)x_e^{(q)}(\lambda-\tau_{qr})\right)^2 d\lambda \tag{41}$$

We similarly define an average local power P_{rr} as the area-specific drive from local pyramidal cells $x_e^{(r)}$ to excitatory population of area r over the same interval N_w. Accordingly, define the partial directed influence from area q to area r as

$$\mathrm{PDI}_{qr}(t) = \frac{P_{qr}(t)}{P_{1r}(t)+\ldots+P_{rr}(t)+\ldots+P_{qr}(t)+\ldots+P_{Mr}(t)} \quad q = 1\ldots M \tag{42}$$

Accordingly, a PDI_{qr} describes fraction of excitatory power that an area q contributes out of its own total output to another channel r, taken into consideration all the other contributions. It is hence a directionality of information flow from area q to area r. Large PDI_{qr} values indicates an efficient, strong pathway for this transfer, that is, most of the excitatory drive of q is delivered to r; hence the influence of q is directed toward r . On the other hand, low PDI values indicate a weakness in this directed flow.

Granger Causality Testing

To accommodate time variations in the model, the generalized partial directed coherence measure (GPDC) was computed for both simulated and clinical data. GPDC is a Granger-causality tool derived from the fitted time varying parameters of a vector auto-regressive (VAR) model of the multichannel data Havlicek et al. (2010). Briefly, for a vector of m recorded signals at time k, $\mathbf{x}_k = \begin{bmatrix} x_{1,k} & \ldots & x_{i,k} & x_{j,k} & \ldots & x_{m,k}\end{bmatrix}$, $\mathbf{x}_k \in \mathbb{R}^m$, the multivariate autoregressive model (MAR) is given by

$$\mathbf{x}_k = \sum_{r=1}^{p}\mathbf{A}_{r,k}\mathbf{x}_{k-r} + \mathbf{e}_k \text{ for } k = 1\ldots N \tag{43}$$

where N is the total number of observations, p is the model order, and $\mathbf{x}_k$ is the m dimensional multivariate vector, and $\mathbf{A}_{r,k} \in \mathbb{R}^{m\times m}$ is the AR coefficient matrix at delay r and time k. In the frequency domain, denoting $\mathbf{A}(f,k) = \sum_{r=1}^{p}\mathbf{A}_{r,k}e^{-j2\pi fr}$, $\mathbf{A}^s(f,k) = I - \mathbf{A}(f,k)$ (the inverse of the cross spectral matrix), and $a^s_{ij}(f,k)$ as the $(i,j)^{th}$ component of $\mathbf{A}^s(f,k)$, then the resultant two dimensional generalized partial directed coherence function is defined as $GPDC_{i,j}$

$$GPDC_{ij}(f,k) = \frac{a^s_{ij}(f,k)\frac{1}{\sigma_i(k)}}{\sqrt{\sum_{i=1}^{k}|a^s_{ij}(f,k)|^2\frac{1}{\sigma_i(k)^2}}} \tag{44}$$

where $GPDC_{ij}(f_o,k)$ represents the contribution of the frequency f_o component of the time series j to the f_o component of time series i, at time instant k, taken to account the contribution of all other $m-2$ time series at the same frequency. We follow closely the procedure outlined by Havlicek where a linear Kalman filter is applied to find the parameter vector (Havlicek et al. 2010). We utilize a time step of $dt = 5$ ms, a third order model ($p = 3$) whereby the results show most accuracy for the simulated data. To achieve convergence, three iterations of the Linear Kalman filter were conducted (time profile of the parameters remains largely unchanged). We finally smooth the resultant GPDC over 200 ms windows. We establish statistical significance, we compare against GPDC values obtained under the null hypothesis of no causality. These GPDC values were obtained from 10 independent surrogate data sets where each set is generated by independent random shuffling of the phase of the Fourier transform of each signal while keeping the corresponding magnitude fixed. To facilitate comparison with other reported PDI measures, we form the *average GPDC* value as follows: the significant GPDC value $GPDC_{ij}(f_o,k)$ at is first weighted by the spectral magnitude at f_o and k of the sending channel (with the spectrum normalized to have a maximum magnitude of one over all time); the sum over all frequencies of the weighted GPDC value is then computed, and finally the average value is reported over the whole time segment

FEAST Induced Seizures

As mentioned in the introduction, the therapeutic outcome of ECTtreatment session is dependent on several parameters in the stimulation protocol including the location of the stimulation electrodes, the polarity of the stimulation, and the duration of the delivered pulse train. Among various ECT protocols, electrode placement is believed to be a more effective determinant of short-term and long-term cognitive deficits than stimulus dosage (Lisanby 2007). Specifically, the bilateral stimulation (BL) protocol (two symmetric electrodes of equal size, placed frontally, with bidirectional currents alternating across them) results in more profound amnesia independent of dosage. Stimulation with a sine wave current is inherently inefficient and results in greater cognitive deficits. Different protocols have thought to introduce various adjustments to the stimulation parameters to achieve better outcomes. For example, the right-unilateral (RUL) ECT (electrodes of equal size) applied ultrabrief rectangular pulses (0.3 ms) at high dosage (6 × Seizure Threshold) to attain significant advantages in acute, short-and long-term cognitive effects. A recent modification of such protocol is the Focal Electrically Administered Seizure Therapy (FEAST) which uses unidirectional current and asymmetrically sized electrodes with a narrower frontal electrode as schematically shown in Fig. 7. Note that the placement of the frontal electrode is also modified by it being over the forehead instead of the temporal area. This aimed to focus the delivery of charge/unit area to the right frontal areas and hence, to the prefrontal cortex implicated in therapeutic outcome while reducing the involvement of anterior temporal pole regions and hence alleviating cognitive side effects.

To explore the specificity of the FEAST-induced seizure right prefrontal drive, we devised a study where subjects were cross-randomized to receive FEAST in its classic configuration or with electrode placements reversed. Subjects were under general anesthesia pre classic clinical protocols (Nahas et al. 2013). All subjects signed informed consent approved by the American University of Beirut Medical Center Institutional Review Board.

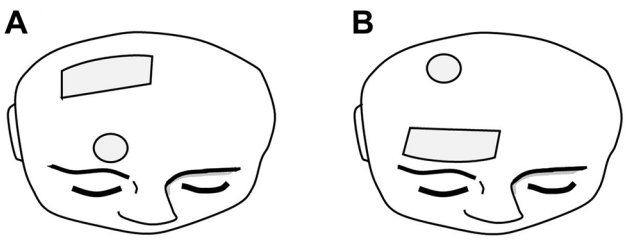

Fig. 7 Schematic of stimulation electrode placement in FEAST. **a** Normal configuration. **b** Reversed configuration

We here focus on scalp EEG recordings obtained from the frontoparietal electrodes FP2, FP1 and from the central electrodes C5 and C6 to understand interaction between underlying neural structures during the baseline anesthesia and during the wide-spread seizure time period. Note that the interaction here is modeled using generative models (effective connectivity) in the absence of knowledge about the connectivity to sources outside this subnetwork.

The four-area model was fitted in a similar manner to simulations used above. The internal model parameters and the long-range delays were kept fixed to those used for the simulated data.

Unlike the exact dynamics of the simulated data presented earlier, the model is obviously an approximation of the underlying dynamics. Hence the initial estimate of the model (before adjusting) was obtained using multiple forward passes of the SCKF through the data (here 3 iterations) to allow for a slow approach towards the solution. With the tuning procedure, we noted that few forward passes sufficient to become in the vicinity of the solution. This condition was determined quantitatively when the change in the log-likelihood of the prediction error between iterations becomes small, and this was corroborated qualitatively by small between-iterations modification in the profile of model parameters and the estimated input.

Table 1 shows the various tuning parameters of the SCKF (iterations 1–3) and the residual-adjusting iteration (denoted by RA). All the hidden firing states of the model started from zero initial conditions $\mathbf{x}_0 = 0$. The parameters and inputs started from the same initial value θ_0 and u_0, respectively. Their initial covariance matrices were all diagonal with the variances listed as $P_{x,0}$, $P_{\theta,0}$ and $P_{u,0}$. The covariance matrices of the process noise driving the states were following were $Q_{x,0}$, $Q_{\theta,0}$ and $Q_{u,0}$;the adaptation parameters were λ_x and λ_θ , and λ_u.Finally the observation noise had a diagonal covariance matrix with variance $R_{z,0} = 0.5$ and learning parameter $\lambda_z = 0.01$.

Results

The proposed algorithm was applied first to a simulated system whereby the true connections and inputs are known, and next to EEG recordings obtained under anesthesia and seizure conditions during an electroconvulsive therapy session.

Simulated Induced seizures

Externally induced seizures were simulated in a network of four interacting neural mass models subjected to different

Table 1 Simulation Parameters for the RA-SCKF

Iter.	x_0	θ_0	u_0	$P_{x,0}$	$P_{\theta,0}$	$P_{u,0}$	$Q_{x,0}$	$Q_{\theta,0}$	$Q_{u,0}$	λ_x	λ_θ	λ_u
1	0	150	50	400	1	5	30	0.8	$1e^{-3}$	0.5	0.05	$5e^{-5}$
2	p	p	p	0	0	0	30	0.05	$1e^{-3}$	0.35*	0.05	$5e^{-5}$
3	p	p	p	0	0	0	30	0.2	$1e^{-3}$	0.35	0.05	$5e^{-6}$
RA	p_1	p_1	p_1	p_1	p_1	p_1	30	0.2	$1e^{-3}$	0.35	0.05	$5e^{-6}$

Runs 1–3 are SCKF for obtaining initial estimates. RA indicates the residual adjustment run. Notation: p value is the one obtained at the end of previous run. p_1 is the value at the same time in previous run. All values are for both anesthesia and ictal conditions (except for $*$: value for ictal run was 0.4)

input waveforms to approximate area-specific activity (see "Methods" and "Appendix"). Models of distant areas were reciprocally connected via delayed synaptic pathways, some of which change in time. The observations, electrical activity approximating EEG recordings and related to neuronal firing rates, were subsequently used to estimate various hidden states in the model (firing rates) as well as assumed unknown parameters (area-to-area connections) and external inputs.

Noiseless Observations

We first studied the performance of the residual adjusted square-root cubature Kalman smoother (RA-SCKS) under an ideal situation where all the simulated observations have no noise. Multiple trials with randomized input profiles (input signals having an additive random component) were used to obtain the statistical profile of the parameter and input estimates. Figure 8 show the obtained estimates of the

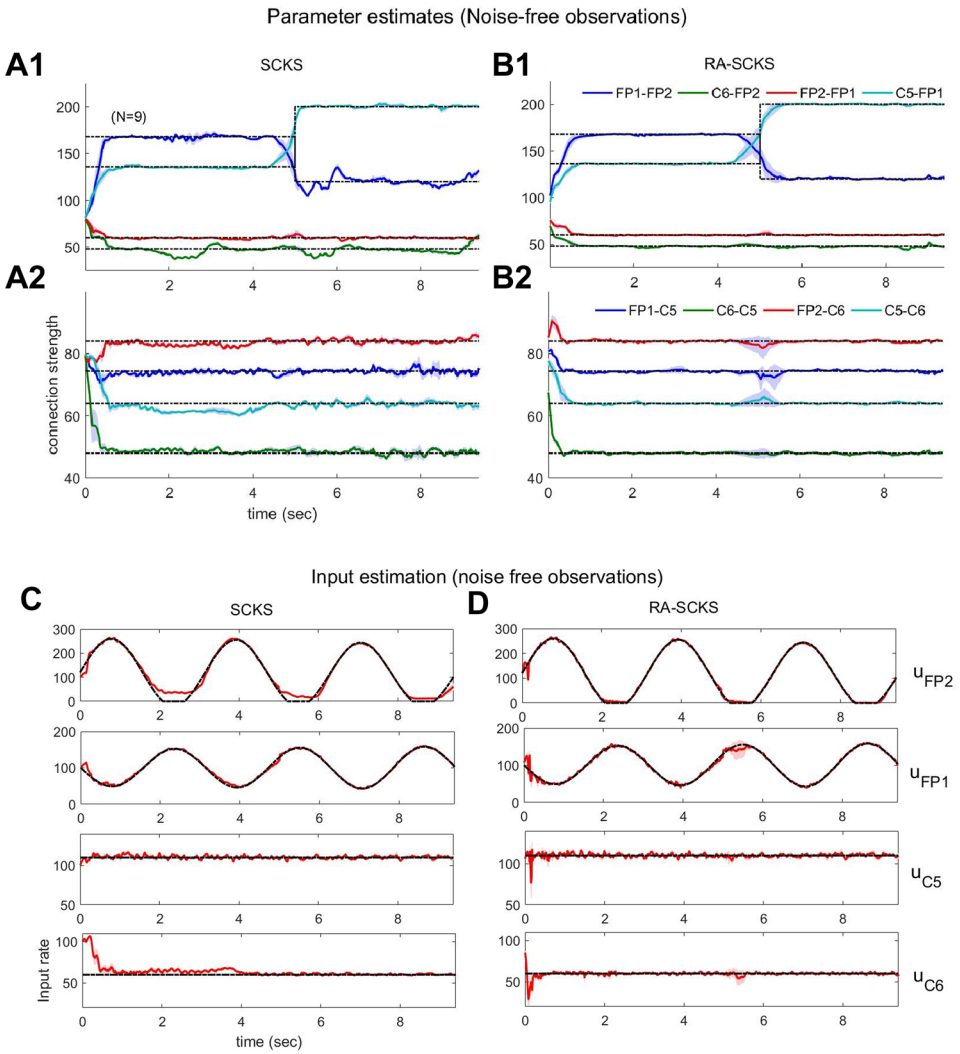

Fig. 8 Parameter and Input estimation in simulated network under noise-free output recordings. For all plots, dashed lines represent true values used in simulation. Solid lines represent the mean over N = 9 trials and shaded regions represent the corresponding 95% confidence interval (standard error). **a**, **b** Estimates of the inter-areal connection parameter under SCKS algorithm (**a1**, **a2**) and the RA-SCKS algorithm (**b1**, **b2**). Legend in **b1**, **b2** apply to **a1**, **a2** respectively. **c**, **d** estimates of the local-inputs to different areas under the SCKS algorithm (**c**) and the RA-SCKS algorithm (**d**). Trace labels in **d** apply to corresponding traces in **c**

parameters and the input for SCKS and RA-SCKS. Solid Traces show the mean values and shadings are the 95 % confidence intervals over multiple runs. For estimating the model parameters, it is noted that, while the SCKS algorithms perform fairly well (Fig. 8a1, a2), RA-SCKS exhibited significantly lower variations in the parameters, and particularly gave more accurate estimates around the jump in parameters FP1–FP2 (Fig. 8b1, b2). In estimating the area-specific inputs, SCKS again performed well but with more error than RA-SCKS in estimating u_{FP2} at low values and with slower convergence in estimating the smallest input rate u_{C6} (Fig. 8c, d). Notice that in all cases the confidence intervals were fairly narrow, which is expected since the only randomness was an additive component at the input level.

Noisy Observations

We later considered the more realistic situation where the observations are subjected to measurement noise (example trace seen in Fig. 6d). Average estimates over 10 trials are shown in Fig. 9 for both filters starting from the same initial conditions. It is noted that (i) the time-varying parameters (FP1–FP2 and C5–FP1 in Fig. 9b1, a1) were both more closely tracked (means values close to the true values shown in dashed lines) and were more consistent across trials (narrower confidence intervals) under RA-SCKS when compared to SCKS; (ii) the remaining time-invariant parameter were more closely tracked under RA-SCKS and were less affected by modifications of the time-varying parameters (compare the period $t < 5$ s in Fig. 9b2, a2, respectively); (iii) the input estimates were significantly closer to their true profiles (dashed black lines in Fig. 9d, c) under RA-SCKS despite having larger variations across trials compared to SCKS. Indeed, it is seen that the regular SCKS has a fairly consistent estimate (narrow shaded regions in c) across all trials which is indicative of the lack of discriminability of the input estimate under a regular SCKS.

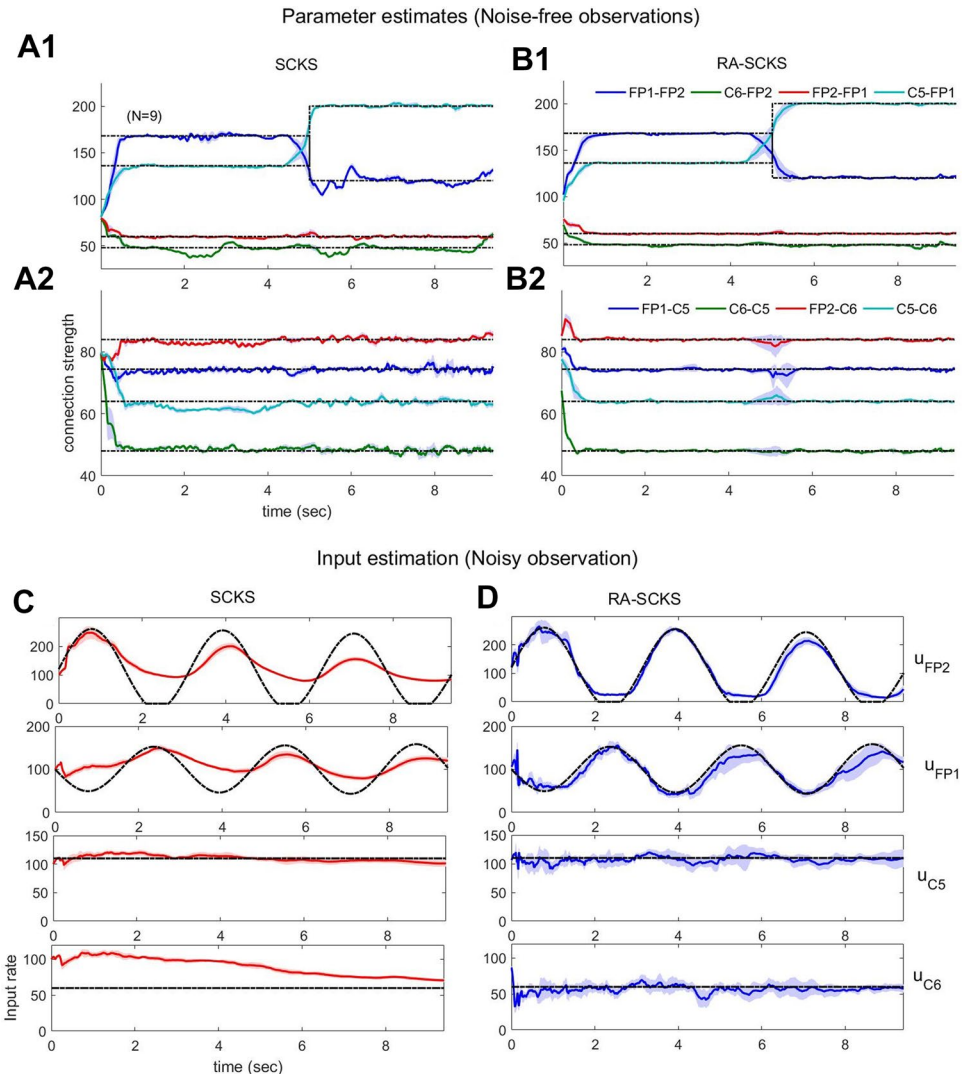

Fig. 9 Parameter and Input estimation in simulated network under noisy output recordings. For all plots, dashed lines represent true values used in simulation. Solid lines represent the mean over N = 9 trials and shaded regions represent the corresponding 95% confidence interval (standard error). **a**, **b** Estimates of the inter-areal connection parameter under SCKS algorithm (**a1**, **a2**) and the RA-SCKS algorithm (**b1**, **b2**). Legend in **b1**, **b2** apply to **a1**, **a2**) respectively. **c**, **d** estimates of the local-inputs to different areas under the SCKS algorithm (**c**) and the RA-SCKS algorithm (**d**). Trace labels in **d** apply to corresponding traces in **c**.

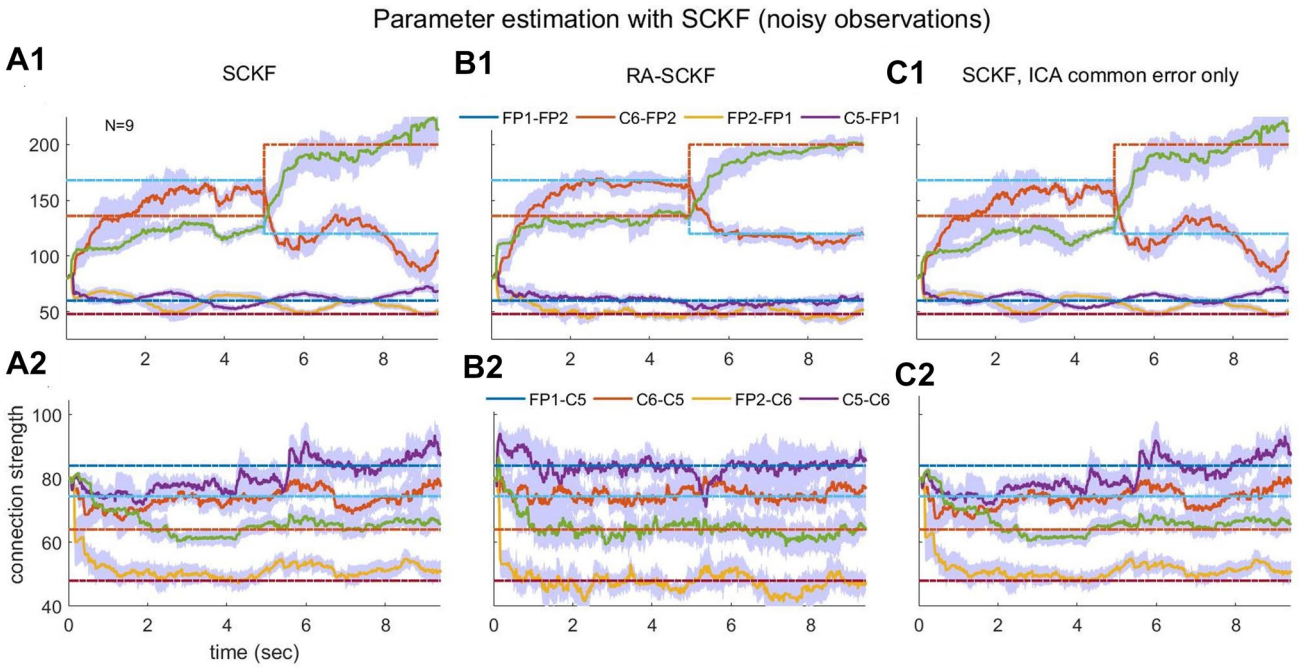

Fig. 10 Comparison of the connection parameters estimate quality under tested algorithms. **a** SCKF algorithm, **b** the RA-SCKF algorithm and **c** SCKF algorithm with ICA applied to remove common channel noise. Dashed lines represent true values used in simulation. Solid lines represent the mean over N=9 trials and shaded regions represent the 95% confidence interval (standard error). Legends in **b1**, **b2** apply to corresponding traces in **a1**, **a2** and **c1**, **c2**

We next emphasized the role of the applied residual-adjusted algorithm in the attaining the above improvement in estimation. To do so, we applied ICA algorithm alone to removing the common noise component without adjusting and we also investigate the results for only the square-root cubature Kalman filter SCKF (without the smoother). In Fig. 10 shows the results for multi-trial runs of the SCKF, RA-SCKF, and SCKF with ICA used to remove the common error. It is noted that while the absence of a smoother resulted in more fluctuations—a characteristic of the forward filter, the RA-SCKF indeed retained the salient improvements discussed earlier (compare Fig. 10a, b). Importantly, ICA removal of common noise had very little impact on the regular SCKF performance (Fig. 10c, a) and, therefore, noise attenuation was not a key factor in the attained improvement, which is indeed due to the applied adjusting algorithm.

Comparison with Granger Causality Measures

The RA-SCKF performance on the simulated model above is benchmarked against a time-varying Granger Causality measure, the Generalized Partial Directed Coherence or GPDC (see "Methods" for implementation details). Note that the results reported for the *noiseless* simulated data set (and hence are the best expected outcomes). We again assume the knowledge that diagonal connections (C5, FP2 and C6, FP1) do not exist.

The resultant GPDC image in Fig. 11 shows a drop in GPDC values from FP1 to FP2 around the correct change time (5 s) is apparent which agrees with the ground truth. The increase in C5 to FP1 connection is not detected. Instead, other changes that do not correspond to true parameter values can be seen, namely the drop in C6 to FP2 and the associated increase in FP2 to FP1 and FP2 to C6 connectivity. Again, and while abrupt time variability was well captured, the correct direction of changes was not correctly represented.

Real EEG Recordings

We next applied the RA-SCKS algorithm on sample recordings obtained during Electroconvulsive therapy sessions (FEAST protocol, see "Methods"). For a given subject in a single session, two data segments were extracted (1) shortly before application of the electrical stimulation (anesthesia condition) and (2) during the induced seizure period, after the stimulating electrodes have been removed and the seizure is wide-spread across the brain. Figure 12a, b show sample recordings obtained during an ECT session [FEAST protocol (Fig. 7a)] from subject 20 collected from the frontoparietal (FP1, FP2) and central electrodes (C5, C6) for

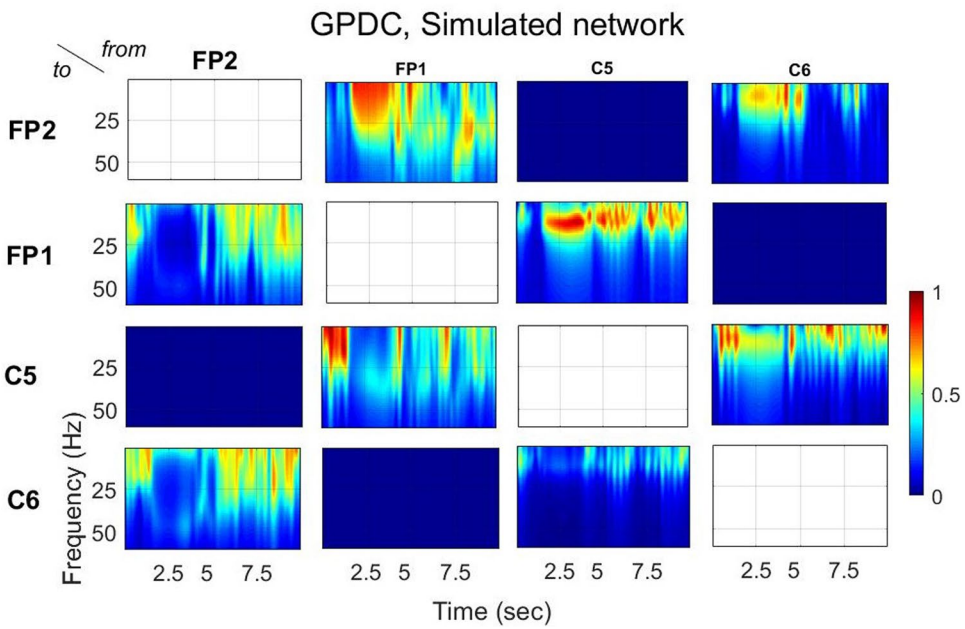

Fig. 11 GPDC values for the simulated network. Rows represent the receiving channels; columns the sending channel. Colorscale applies to all images

both anesthesia and induced seizure conditions (approximate location of these electrodes in Fig. 6c).

In terms of the area-specific driving inputs, it is noted that for the anesthesia condition (Fig. 12c), the estimated inputs are elevated for all the electrodes, with the central electrodes showing slightly larger inputs compared to the frontal areas. On the other hand, for the induced seizure condition, all the area-specific inputs were significantly lower than the same inputs under anesthesia condition (Fig. 12d) and particularly for area FP1. Being a measure of the unexplained drive to the recording activity at a specific channel, the driving input indicates how well a given channel is integrated within the subnetwork under consideration (see "Discussion"). Hence the large drive during anesthesia points to the contribution of other areas to the recorded activity. On the other hand, the decreased drive for the channels during the induced seizure points to increased contribution of the local and subnetwork to the activity of the channel, as particularly evident in channel FP1. The decomposition of this within-network contribution can be understood by the strength of the self excitation and afferent connection to that channel, described next.

In terms of the inter-areal connections, it is noted that no significant trends were seen and their strengths rather reaches nearly constant values for both of the given conditions and the relatively short ~ 10 s segments. The corresponding average values are displayed in Fig. 12e. Under the anesthesia condition (blue horizontal bars), the model shows relatively weak connections with the largest contributions in the forward direction (C5–FP1 and C6–FP2), a significantly smaller contribution in the backward direction (FP1–C5 and C6–FP2) and a symmetric profile across the hemispheres. Under the induced seizure condition, the model shows a significant increase in the connection strengths throughout the subnetwork. This was particularly true for the backward direction in the frontal-to-central (FP2–C6 and FP1–C5) and the reciprocal central connections (C6–C5 and C5–C6).

Figure 13 shows an ECT/FEAST session for the same subject with the electrode configuration reversed , that is the wider electrode is place on the forehead and the smaller electrode placed centrally (Fig. 7b). The anesthesia condition again shows elevated inputs across all the electrodes as and the absolute connections strengths have similar profiles to those obtained in the normal configuration session (compare Figs. 12c, d, 13c, d). This indicates robustness of the obtained model and points to an expected result that the anesthesia condition has similar dynamical character across sessions. However, the induced seizure for this reversed configuration shows a very different profile from that under normal configuration. Specifically, the driving inputs are now of smaller magnitudes indicating more confined activity within the network; the average connections strengths are also weaker compared with the normal FEAST, as shown in Figs. 12e, 13e and listed numerically in Table 3 of "Appendix". This is particularly true for the backward connection from FP2 (FP2–C6 dropped by 72%), the reciprocal forward connection from C6 (C6–FP2 dropped by 45%), and the interhemispheric central connections (C6–C5 dropped by 80 % and C5–C6 dropped by 57%).

To gain a closer understanding of the individual channel contributions to the overall network activity, we devised a normalized directional interaction measure which considers the strength of a directed pathway between two channels in relation to the overall directed pathways from the originating node. Specifically, we consider the transmitted power

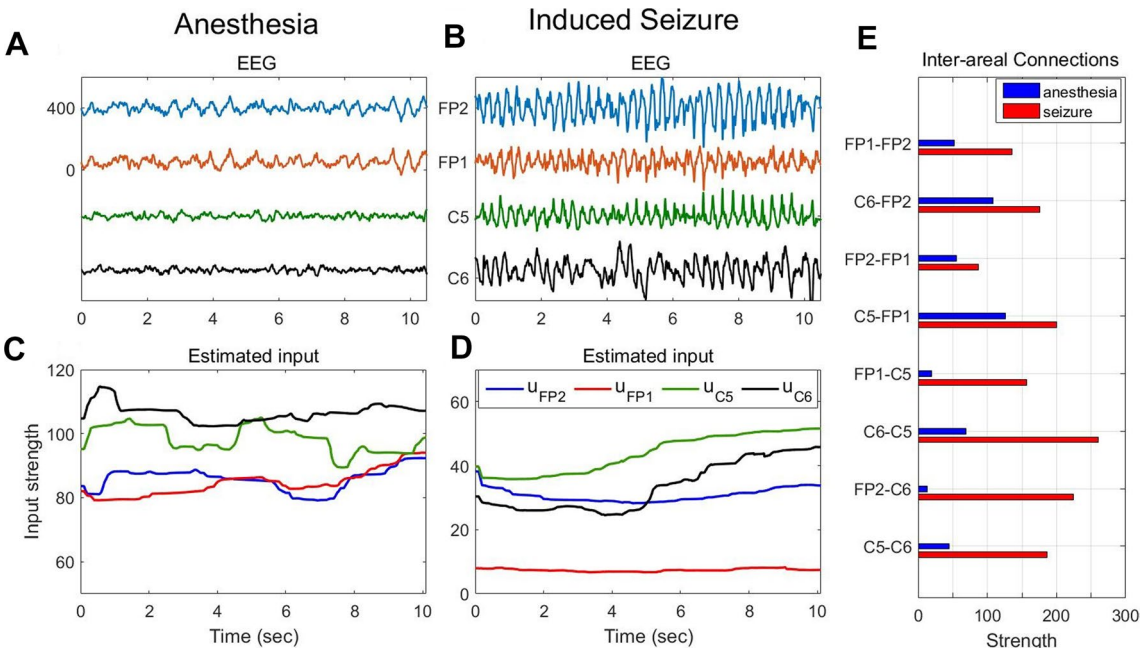

Fig. 12 Predictions of inputs and connections for clinical data, normal configuration FEAST (subject 20). **a** and **b** EEG recordings from indicated channels during anesthesia (prestimulus) and seizure conditions (post-stimulus). **c** and **d** estimated inputs to various channels for the anesthesia and seizure conditions (legend in **d** applied to both). **e** Average inter-areal connection during anesthesia (blue bars) and seizure conditions (red bars)

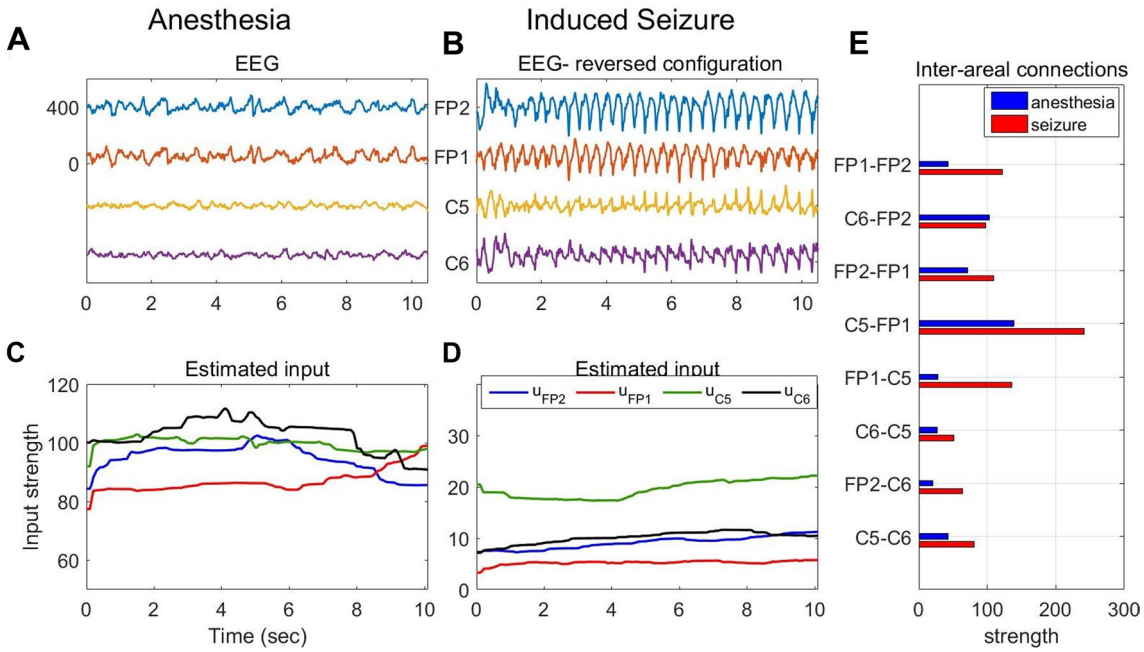

Fig. 13 Predictions of inputs and connections for clinical data, reversed configuration FEAST (subject 20). **a** and **b** EEG recordings from indicated channels collected during anesthesia (prestimulus) and seizure conditions (post-stimulus). **c** and **d** Estimated inputs to various channels for the anesthesia and seizure conditions (legend in d applied to both). **e** Average inter-areal connection during anesthesia (blue bars) and seizure conditions (red bars)

along a synaptic pathway originating from the deep layers of that emanating node and arriving at the receiving node. We then compute the *partial directed influence* (PDI) from one channel on another as the power transmitted across their

connection to the total power transmitted from the emanating channel (see "Methods" for details).

Fig. 14 shows the PDI in the considered four-channel network for the anesthesia and seizure conditions blue and red bars, respectively) for the normal and reversed configurations. In the normal configuration (Fig. 14a and Table 2), we note for example that out of the right frontal channel, the backward connection (FP2–C6) constitutes around 65.3% of the total power emanating from channel FP2, the interhemispheric (FP2–FP1) constitutes 10.6 %, and the remaining is self excitation 24.1% (FP2–FP2, not shown). On the other hand, the connections out of the FP2 in the reverse configuration drop to 13.8 % (FP2–C6), 35.4 % (FP2–FP1) and 50.8 % for self excitation (Fig. 14b) . To gain further insight across sessions, we measure the change of PDI values during seizures with respect to the baseline PDIs, obtained during anesthesia (that is, $\left(\mathrm{PDI}_{seizure} - \mathrm{PDI}_{anesth}\right)/\mathrm{PDI}_{anesth}$). This is shown in Fig. 14c for the normal and reversed configuration (green and black, respectively). Clearly, their is a substantial difference in the backward directions (FP2–C6 and FP1–C5) during the seizure between the two conditions. This points to a decrease in the centrality of the frontal areas in contributing to the overall network activity upon reversing the electrodes with the largest decrease in the contribution of the right frontal channel FP2.

Comparison with Granger Causality

The GPDC maps for the ictal conditions are shown in Fig. 15. Under normal configuration, the GPDC of seizure data activity shows a dominance at low frequencies of the right hemispheric outgoing influence (from PF2: FP2–C6, FP2–FP1,and from C6: C6–FP2, C6–C5 in Fig. 15a). This is changed under reverse configuration to a decreased influence of outgoing FP2 connections (FP2–C6 and FP2–FP1) but not of C6 connections (C6–FP2 and wider frequency influence of C6–C5). To facilitate comparison with RA-SCKF results, we plot the average GPDC value (time-averaged, weighted GPDC values; see "Appendix" for details). This is shown in Fig. 15c for both configurations. In agreement with RA-SCKF, it is noted that the backward connections (FP2–C6 and FP1–C5) have decreased for reverse in comparison with normal configuration [ratio(normal/reverse) = 2.96 and 2.15 respectively].On the other hand, the forward connection is also decreased (ratio = 1.35 and 1.18 respectively) unlike the prominent increase reported under RA-SCKF (from table 2), for C6–FP2 and C5–FP1 ratio of PDI are 0.86 and 0.62 respectively).

To compare within the same session, we plot the average PDCF values for anesthesia condition (blue) along with the seizure conditions (red) for both normal and reverse configurations in Fig. 15d and e. It is here noted that the strength for many connections is in fact significantly larger during

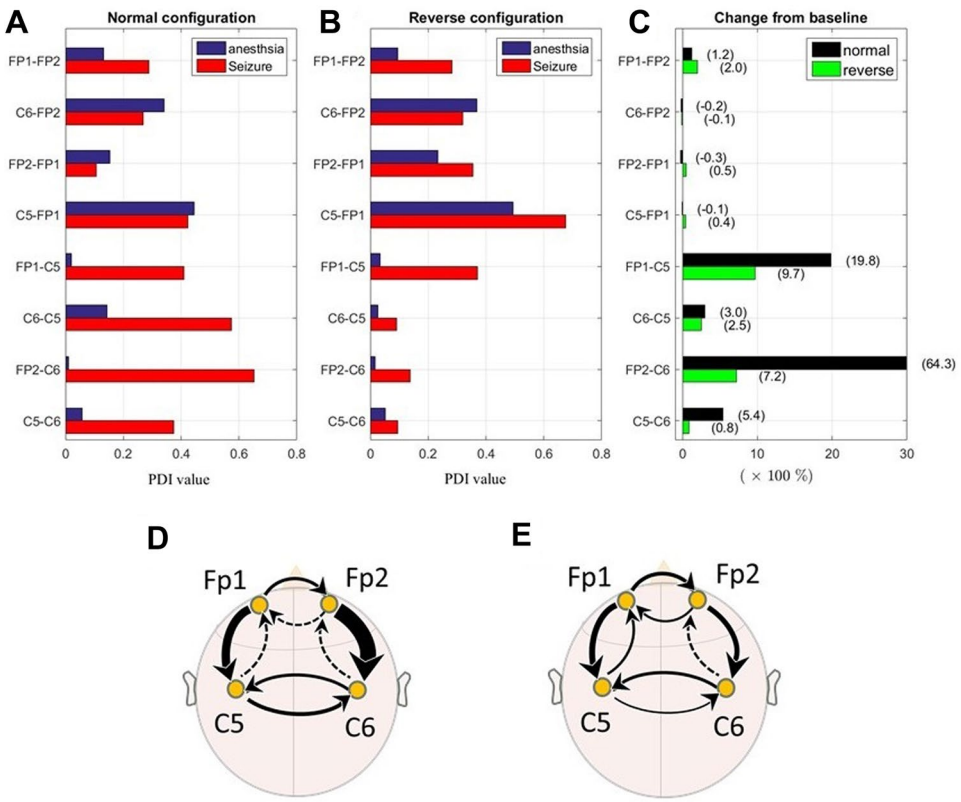

Fig. 14 Partial directed influence to detect directional connectivity. Partial directed influence (PDI) for anesthesia (blue) and seizure (red) conditions for **a** normal and **b** reverse FEAST electrode configurations for subject 20. **c** Change in PDI values of seizure activity from the anesthesia baseline for normal (black) and reverse (green) electrode configurations for subject 20. Schematic of PDI changes with normal (**d**) and reverse (**e**) configurations. Width of solid black lines reflects magnitude of increase in strength. Dashed arrows reflect marginal decrease in strength

Table 2 Estimated partial directed influence values for ictal activity under normal and reverse electrode configuration for Subject 20. (Nonexisting connection marked by –)

(a) Normal configuration FEAST					(b) Reverse configuration FEAST				
	FP2	C6	FP1	C5		FP2	C6	FP1	C5
FP2	0.24	0.27	0.29	–	FP2	0.51	0.32	0.28	–
C6	0.65	0.16	–	0.38	C6	0.14	0.59	–	0.09
FP1	0.11	–	0.30	0.42	FP1	0.35	–	0.35	0.68
C5	–	0.57	0.41	0.20	C5	–	0.09	0.37	0.23

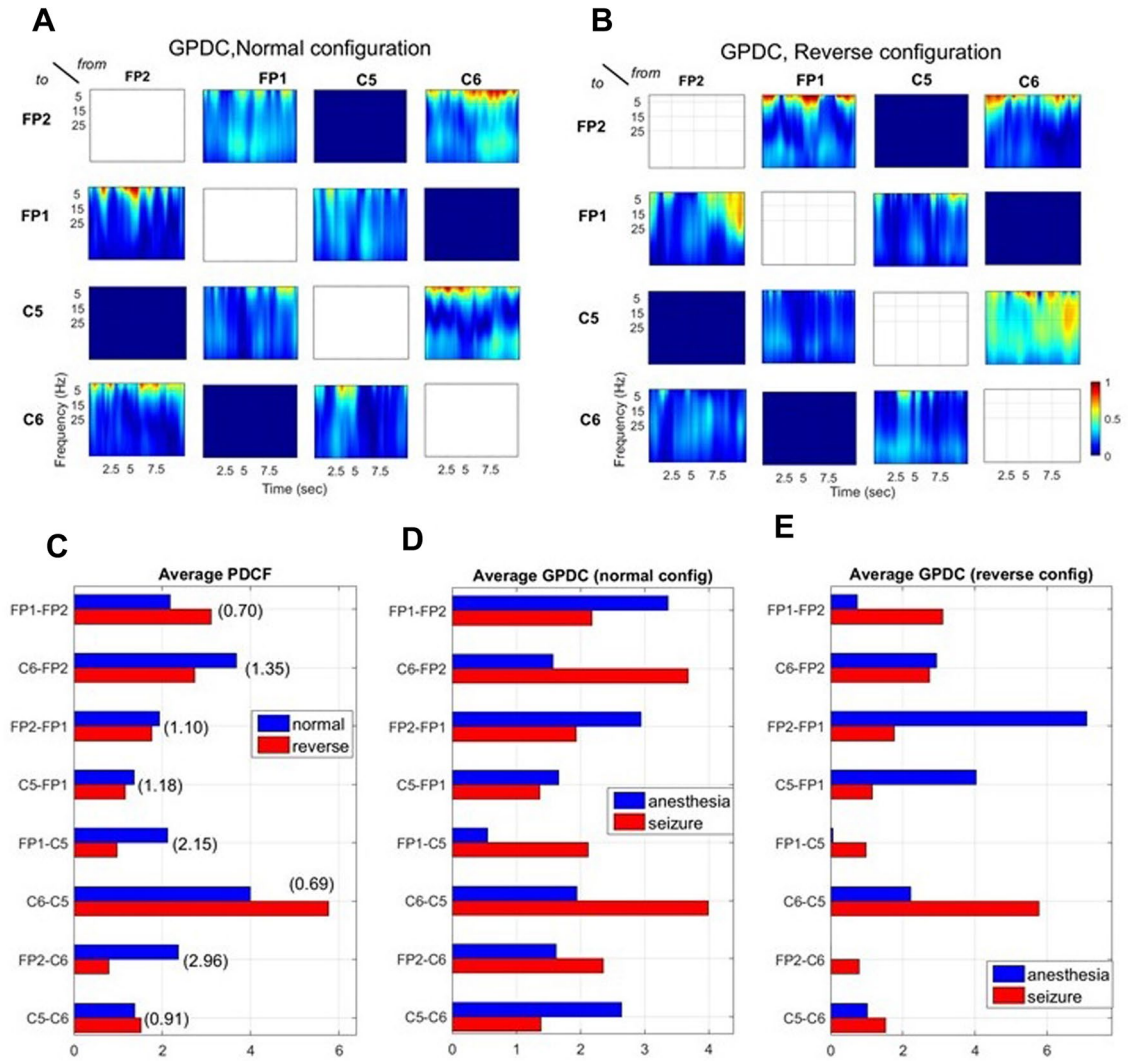

Fig. 15 GPDC maps for clinical ECT-induced seizure in subject 20. **a** under normal electrode configuration; **b** under reverse electrode configuration. Color scale in **b** applies to both maps. **c** Average GPDC values computed for the ictal condition for the two configurations. **d** and **e** Average GPDC computed for the anesthesia (red) and ictal conditions (blue) in normal and reverse electrode configuration. **e** ratio of ictal to anesthesia average GPDC values for the two configurations

anesthesia than that during ictal activity (e.g. FP2–FP1 and C5–FP1 in Fig. 15e). This can be explained by the lack of input representation in the GPDC computations; that is large input drives that are common to two areas can incorrectly be reflected as increase in causal connection between these areas.

Discussion

Functional and effective connectivity are two complimentary approaches for discovering directed information flow in dynamic networks (Friston 2011). While functional connectivity is model-free, that is, can be applied to study

causal relationships between any number of measurements in a network, effective connectivity requires observing the output from all the major areas in the network in order to obtain accurate modeling of the generative underlying process. Still, a main advantage of effective connectivity is its ability to induce causality and directed information flow as constrained by plausible underlying generative processes, and hence to provide experimentally testable predictions. At the current stage, the ability to study effective connectivity in electrical recordings of distributed phenomena is limited by the (i) the technical difficulty in obtaining measurements from distributed, large scale dynamic networks (e.g. including deep brain structures) and (ii) the high computational complexity of fitting large scale, generative interconnected models to the available data. By employing subnetwork modeling, to address effective connectivity between component areas in that subnetwork is therefore a highly desirable target since it can provide a mechanistic or biological explanation of directed information flow in the subnetwork.

Fitting subnetwork models is an ill-posed problem since both the model parameters and the driving inputs. Improved accuracy solutions were proposed using (i) state space descriptions of local dynamics that are amenable to efficient estimation procedures using Kalman filtering, and (ii) residual error decompositions that update the Kalman-induced estimates of the unknown inputs.

State space descriptions of dynamical systems lend themselves naturally to the benefits of Kalman filters for estimating hidden internal states based solely on noisy output observations. The obtained estimates are optimal for one-step ahead prediction under Gaussian uncertainty. With the introduction of cubature rule, the efficient time-adaptive nature of the framework has been extended to both nonlinear process and observation functions. A main advantage of state space descriptions is the ability to account for noise at the process level, which allows the incorporation of the uncertainty and environmental noise impacting the process at various scales.

For model fitting and input deconvolution, the Kalman procedure searches for the various unknown quantities as hidden states with simple dynamics that are driven by Gaussian noise processes. A main challenge in this procedure is the selection of the initial conditions of the corresponding states and the covariances of the associated noise processes. This a well known lengthy process where the modeler manually tunes the filter behavior and incorporates any physical knowledge or a priori information about the various unknowns.

Algorithm Summary

The current approach is an automated tuning procedure for adaptively adjusting the covariances of the input and the parameters by judiciously dissecting the unexplained residual or mismatch in fitting the output observations. The tuning procedure occurs in four sequential stages that we qualitatively summarize and critically evaluate below.

Initial Estimate

In the first stage, an SCKF searches for the model parameters and the input(s) to the system in a manner similar to the other approaches presented for nonlinear Kalman filtering in neural estimation (Havlicek et al. 2011; Karameh et al. 2014; Madi and Karameh 2017). Specifically, the process noise driving the augmented states of the inputs and the parameters were all driven by noise processes whose covariances adapted in time as proposed by Robbins and Monroe (see "Methods" for detail). The search is conducted over multiple forward passes through the experimental data and is designed to incrementally approach the solution for the parameters of the known function. This procedure assumes some knowledge of the rate of variations for the model parameters, which are here deduced from the physical nature of the problem. In the limit of constant parameters, the iterative process could include forward-backward filtering (or smoothing) thereby decreasing the error until it arrives at minimal change in the log-likelihood (as in Havlicek et al. 2011). For time varying parameters, the problem is significantly harder with single trials. Still, our simulated data showed a good convergence with the slow approach whereby multiple passes through the data act to inch closer to the correct solution. For time varying parameters where observations are locked to a particular event (e.g. evoked potentials), a more general approach could be to incorporate multiple trials as multiple observations of the underlying system,hence increasing the dimensions of the observation vector (Hu et al. 2012). Note that this output augmentation can be handled quite efficiently in SCKF since cubature points are dependent on the dimension of the state vector (which remains the same) and not that of the output.

For this stage, the incremental learning of the parameters occurs at a faster rate than that of the input. This is justifiable since we assume that the dynamics of the output are shaped by the system itself and that the driving input modifies the operating point of the nonlinear system and hence introduces impacts the output dynamics at lower frequencies [Thinking in terms of linear systems, this is akin to introducing a switching linear time invariant, autoregressive model of the output as being driven by white noise that is stationary over the time window for which the model holds].

Error Decomposition

For the state estimation resulting from the first stage, the residual mismatch between the output and the prediction in

all the channels, the multivariate output prediction error, is decomposed using fast ICA () over non-overlapping time windows. Independent component analysis has traditionally been used in decomposition of fMRI to study functional connectivity (Correa et al. 2007; Damaraju et al. 2014) with fastICA demonstrating reliable results (Correa et al. 2007). For each time window, this decomposition results in a number of independent sources as the number of channels. Depending on their relative contribution to the various residuals, these components could be used to reconstruct two error signals (i) a global or common error signal whenever an independent component has roughly a uniform contribution to the error in various nodes and (ii) channel-specific error signals to be deduced from individual independent components whenever such components are skewed towards representing only single channels.

Since the common error component cannot be attributed to any individual channel, it is thought of as an unexplained global residual that bears no resemblance, or information about, the driving inputs to single channels. Rather, it could be an environmental noise impinging at the channel outputs (e.g. muscle or blink artifact). Indeed ICA decompositions are utilized to remove such artifact components from multichannel EEG recordings.

The channel-specific error component is localized to the output of a given channel. It is based on a single independent component that contributes almost solely to the residual in predicting that channel output. This mapping is based on the premise that channel-specific driving inputs maximally exert their effect on their respective nodes of the network, although their effect could be reflected in connected nodes (see "Methods", assumptions A1–A4). Note that different reconstructions can be achieved depending on the working assumptions of the system, as will be detailed later.

Covariance Update

We now utilize the error signals computed above to update our estimates of the covariances of the noise processes for the various inputs and corresponding parameters. For each of the non-overlapping time windows used in the second stage, we first remove the common noise component. Then, for each channel with nonzero channel-specific error, we estimate a hypothetical desired additive driving input that (i) possess the same frequency range as the estimated input (based on apriori physical knowledge), and (ii) that reduces the channel-specific error or that pushes or "nodges" the predicted output closer towards the actual output of that channel—after the latter is modified to account the common noise component. A geometric interpretation of this idea is presented in Fig. 5. Note that the "nodge" effect is chosen so as to keep the system dynamics in the vicinity of the previously estimated dynamics estimated (first stage) such that local linearizability and input additivity is exploited.

Here, after finding the desired additive input deviation from the error, we compute the linear approximation of this desired input from the previously estimated input (first stage) using an optimal linear (Wiener) filter. Finally we measure the accuracy of this approximative as a ratio of energy in the linearly predicted drive to the energy of the desired additive deviation, over relevant frequencies. We devise a scaling factor for the covariances of that channel in terms of the accuracy for each of the desired signals in each of the time-windows. This adaptive scaling factor acts to augment the process noise covariance of the channel input and to reduce the process noise covariance for all the parameters of the same channel.

Adjusted Estimate

A forward run of the SCKF is now conducted with the scaled process noise covariances for all the channels to predict the cleaned output utilized in the third stage. Since the scaled covariances are applied locally over non overlapping time windows, the inputs to different channels are adjusted independently by different amounts as determined in stage 3 to decrease the error specific to that channel. The decrease is warranted by the constraints set in stage 2 that keep the system within its linearizable range and hence attain a significant linear predictability of the error performance.

Note here that the SCKF acts to improve the overall prediction error. Since the procedure acts to modify the covariance of the input rather than its magnitude, we do not impose any direction of change; rather, the SCKF autonomously chooses this direction for both the input as well as the parameters of the model. This positive outcome is noted in the simulated system where, in addition to fitting the input, an improvement in the parameters, particularly around jumps, is noted. In both the simulated and experimental data, we noted that this adjustment and improvement of the Kalman performance results in decreasing the prediction error (increasing the log likelihood).

Subsequently, the forward Kalman residual-adjusting run is followed by a backward smoothing run (SCKS) that removes any high frequency artifacts introduced by the time-window approach of covariance updates, and provides information propagation across windows.

Performance with Simulated Data

In the four-channel neuronal model, the core dynamics of a single channel (or node) was a variation of well-known neuronal population model, the Jansen model (Jansen and

Rit 1995), where sigmoidal nonlinearities transform membrane potentials to firing rates of each population. The large spikes exhibited during a simulated induced seizure are well into the nonlinear operating range of these dynamics. The abruptly changing connection strengths utilized further complicate the overall picture, with the output data representing single-trials field potential recordings of a nonlinear, time varying system driven by for which model inversion is to be performed.

The difficulty of this problem is demonstrated by the performance of the native Square Root Cubature Kalman filter in conducting the model inversion under noisy observations. It is important to note that the accuracy of the SCKF (or the filter with the backward smoother SCKS) has been bench marked previously against other state-of-the-art Bayesian methods for fMRI model inversion (Dynamic Expectation Maximization) by Havlicek where it showed superior performance (Havlicek et al. 2011). Additional experimental validation test on SCKS on deconvolution of neuronal inputs to BOLD signals were done in Dubeau et al. (2012). Our group has also tested the SCKF performance on parameter estimation for modeled field-potentials against various noise types impacting the system under operation in Madi and Karameh (2017).

For the simulated model with unknown time varying parameters and unknown sustained driving inputs, the SCKS performance was quite accurate in the ideal situation of exact, noise-free observations but it became inaccurate under more realistic situations of slightly-noisy recordings (see Fig. 6d). Specifically, starting with close estimates of the parameters (except for high frequency variations) and sustained time varying inputs (except at low input drives) for the noise-free recordings, the SCKS estimates under noisy recordings showed very slow tracking for some of the parameters and inputs while for others the estimates were quite far from the real values (left columns in Figs. 8, 9).

For the simulated scenario with noise-free observations, the RA-SCKS algorithm showed smoother parameter estimates and more accurate input estimates compared with the same estimates under the SCKS (Fig. 8). The combined effect can be understood by the simultaneous modification of the process covariances for the input and the parameters of a given channel. In particular, while a native SCKS introduces high frequency changes in the parameters to locally reduce the output prediction error, the ICA-dependent residual adjustment locally evaluates the linear predictability of the channel-specific error from the input to that channel, augments the process covariance of the input (based on the linear predictability of the error from that channel) and concurrently attenuates the process covariance for the parameters. Note that this behavior is obtained consistently for multiple Monte-carlo simulations of the system with random noise added to the driving input except at the parameter transition interval as implicated by the width of the confidence intervals (shaded region). The relatively wider interval immediately around the parameter jump (Fig. 8b, d) is caused by sudden movement of the operating point of the system (thereby reducing accuracy in the linearization approach and creating a transient uncertainly in the estimate).

For the noisy observations, the adjusted SCKS showed marked improvement in accuracy for both input and parameter estimation when compared with the SCKS performance. First, while the SCKS estimates for some parameters were quite far from their corresponding true values, the RA-SCKS parameter estimates were significantly closer to their true values; furthermore, the variations of the latter parameter estimates upon switching converged significantly faster to the neighborhood of the true values for all the parameters. Second, the input estimation under the RA-SCKS procedure was resilient to the output noise showing, on average, a close tracking of the various input profiles. This is in stark contrast to the poor input-tracking performance under native SCKS.

Indeed, it is the noise-prone error in SCKS estimation that the proposed algorithm seems to be most efficient at correcting. the RA-SCKS judiciously decomposes the residual prediction error to an input-irrelevant component (uniformly spread error) and to a set of components that are highly correlated with the channel-specific input (channel-skewed error). It is only the latter set that causes covariance scaling for the inputs and parameter states and this occurs only when non-trivial error predictability from these inputs is detected [Naturally, the underlying assumption here is that the observation noise at a given channel is not correlated with the input. This is a realistic assumption particularly since (i) driving inputs arrive from deeper brain structures or through long-distance fibers and (ii) the output recording are often related to the system dynamics through an observation function (EEG, fMRI) which is in a distinct physical domain than that of the driving input.] Note here the wide confidence interval for the estimates, and particularly the high frequency fluctuations in the input estimate, across different trials. We think that this behavior is linked to the window-based ICA decomposition over non-overlapping windows in single trials which leads to (i) high variability of the extracted common-noise signal per window and per trial, and (ii) variability in scaling factor across non-overlapping windows. This effect can be reduced by (i) averaging, for example, when imposing a smoothness constraint on covariance scaling across different windows in a trial (based on a priori knowledge of the input smoothness);

(ii) reduction in the scaling factor of the predictability (K_o in Eq. 38).

Performance with ECT Seizure Data

The neuronal dynamics of electrically induced seizures are sufficiently replicated within the mathematical description of individual areas undergoing stimulation (Karameh et al. 2014). Starting with such descriptions, the algorithm was used to jointly identify the interconnection strengths and the distinct inputs to neural generators to reproduce the observed ECT seizure recordings. It is noted that since electrodes were chosen to be spatially distant (FP1, FP2, C5 and C6), the impact of mixing in the forward model on the overall quality of the results is expected to be minimal.

The outcome of this estimation was a set of biologically meaningful predictions of the directed information flow underlying recordings from a subset of scalp EEG channels. This is particularly significant both at the technical as well as the clinical level.

From a technical perspective, the challenge here is posed by two major obstacles: the generalized nature of the recorded seizure activity, and the nontrivial impact of external influence on different nodes in the network. Principally, the induced seizure recordings are taken post-stimulus, in a time period where the seizure is well-spread across wide areas of the brain including structures located deeper in the brain stem. Therefore, the development of neuronal models that address effective connectivity in a subnetwork should account for all the key drivers into that subnetwork. Accordingly, the model inversion problem at hand includes a blind source deconvolution of a nonlinear model and a persistent input drive whose temporal profile is largely unknown.

The proposed solution dissects the observed recordings as the output of two forms of dynamics that are separated by the time scale at which they unfold: (a) the modeled fast dynamics that are explained by the interactions between local neuronal dynamics, and (b) the unexplained or unmodeled slow dynamics that are summarized by external drives into the various channels. While the fast dynamics explain the stereotypical high frequency waves (in seizures, these are the large spikes), the slow dynamics dictate the necessary additional excitation arriving from outside the subnetwork to arrive at the operating point at which local fast dynamics occur. The inherent assumption or condition here is that the faster dynamics occur within smaller subnetworks (and hence can be explained with a subnetwork of appropriate size), whereas very slow dynamics are the outcome of wide-scale network interaction (and hence are exogenous to the subnetwork). Indeed, this is in line with numerous computational efforts and experimental evidence on brain oscillations across multiple frequencies and on the relationship between temporal and spatial scales of brain function (Hyafil et al. 2015; Proix et al. 2016; Honey et al. 2007; Stein and Sarnthein 2000; Engel et al. 2001; Canolty and Knight 2010; Varela et al. 2001). Naturally, as the contribution of the unmodeled dynamics component increases, the dependence of the subnetwork model on external afferents to explain the observed recordings increases, and the ability of the interactions within the subnetwork to explain the observed phenomenon decreases, thereby rendering the conclusions about information transfer solely based on the modeled component less certain.

Accordingly, the model inversion problem here is the separation of the underlying phenomenon into modeled dynamics (the neuronal model) and summary external drives (the input). Specifically for the ECT model, the magnitude of the driving input to a given channel model is proportional to the excitation that cannot be explained as endogenous to the model neither locally (self excitation) nor regionally in the modeled fronto-central subnetwork.

Anesthesia condition The large exogenous inputs that were estimated during anesthesia are hence indicative of the nontrivial low-frequency contribution of areas outside the subnetwork itself to the overall slow-wave activity. This is an expected result due to the distributed suppressive action of anesthesia over wide cortex areas leading to dominance of low frequency delta waves in EEG (Franks 2008; Uhrig et al. 2014). Although more data is clearly needed, the estimated inter-areal connectivity results suggest a lower feedback influence originating from the frontal areas (Figs. 12e, 13e) and is in agreement to results based on phase directionality reported by (Lee et al. 2009) and across several anesthetic agents (Lee et al. 2013). It is however at odds with results obtained using functional connectivity using time-series modeling (Granger causality) showing increased bidirectional connectivity in the frontoparietal network (Barrett et al. 2012). Still, with recent multi-unit laminar recordings corroborating the reduced inter-areal functional connectivity between prefrontal cortex and visual area during anesthesia (Sellers et al. 2015), functional connectivity (time-series models) might be as a disadvantage in comparison with models that account for driving inputs such as the one proposed herein.

Seizure condition The significant decrease in exogenous inputs to all the channels indicates a larger contribution of the modeled within-subnetwork (endogenous) excitation to the observed ictal activity, and hence reflects a higher confidence in the estimated information transfer predicted by the subnetwork model. In turn, the

increased endogenous excitation can be understood by the significant increase in the inter-areal connectivity strengths, particularly in the normal FEAST electrode configuration. Since connectivity strength is an overall synaptic strength emanating from one area to another (see "Methods"), larger strengths indicate a more effective recruitment of the directed cortico-cortical synaptic pathways as an expected result from dramatic increase in both the firing rate (spiking activity) and the size of the contributing neural populations in the originating area.

Clinical evidence on Electroconvulsive therapy point to the character of a desired therapeutic ECT as one that would trigger seizures in the PFC while limiting its propagation to other brain areas (Sackeim 1999; Spellman et al. 2009). At the level of EEG, it has been noted that ECT produced sharp increases in EEG slow wave activity (Sackeim et al. 1996; Mankad et al. 2010). Strong associations have been demonstrated between the magnitude of such increases in the prefrontal cortex regions (PFC) and the antidepressant effects of the treatment. Spatial specificity has also been shown, in that functional changes in other brain regions are independent of efficacy. Instead, EEG slow wave changes in the temporal cortical areas have been linked to the magnitude of resultant amnesia (Moscrip et al. 2006).

Notwithstanding the clinical manipulations that characterize the effectiveness of a particular protocol, a functional understanding of the areas involved in (a) initiating an ictal seizure under an external stimulus, and (b) propagating this seizure to other brain areas remains limited (Sackeim 1999; Pagerols and Rojo 2009; Grover et al. 2005; Merkl et al. 2009). Imaging studies conducted shortly after an ECT session on cerebral blood flow point to increased average activity in prefrontal area (Ackermann et al. 1986; Nobler et al. 1994, 2001; Enev et al. 2007), but do not provide real clues on the dynamic of the recruitment neither do they highlight the causal influence of affected cortical areas in a given session. With EEG recordings providing a real-time access to the underlying brain activity, a detailed model-based analysis of inter-channel directional relationships as the one conducted here, could shed important light on the ongoing dynamics (Grover et al. 2005).

In the present study, and for the normal FEAST electrode configuration (Fig. 7a), the predicted connectivity showed a significant fronto-central, or feedback transfer of activity in both hemispheres (We are using the term feedback for connections since frontal areas FP1 and FP2 are of a higher order nature and contribute to attentional and planning of cognitive function including motor areas underlying the central electrodes C5 and C6.) The increase from baseline was the highest in the feedback pathway from the frontal (FP2) to the central (C6) region in the right hemisphere where the stimulation electrodes are placed. This indicates a very strong effective recruitment of this excitatory pathway from the ictal frontal areas to the central motor regions. On the other hand, the forward pathway (C6–FP2) shows slight change in strength from the anesthesia condition indicating an ineffective recruitment of this pathway despite the ictal activity in the right central region (that is, smaller proportion of the neural populations are engaged in transmitting their ictal electric activity to the frontal areas).

This indicates the focality of the seizure on the right frontal area in this patient. This is, to the best of our knowledge, the first demonstration of FEAST focality using actual EEG recordings obtained within an ECT session. The prediction is in line with the presumed activity for which the asymmetric electrode configuration was designed. That is, since the front electrode has a much smaller surface area that the back electrode, the current density in the underlying frontal cortical tissue is significantly higher thereby creating higher supra threshold stimulation for the local neural population. Although this was studied in passive current propagation finite element studies by (Deng et al. 2011; Lee et al. 2012) and independently in our group.

The reverse electrode configuration demonstrated that while a seizure can still be attained with wider electrode placed over the frontal regions, this seizure is significantly weaker in the frontal areas as supported by the following results. First, the right feedback pathway (FP2–C6) experienced a dramatic decrease in absolute connection strength (Fig. 13c), in its directionality, as well as in its change from the baseline anesthesia condition (Fig. 14b, c). Second, while the absolute connection strengths from the right central region C6 showed a significant drop along both the forward (C6–FP2) and the interhemispheric (C6–C5) pathways. Since the directional interaction of C6 showed a significant increase in self excitation, it is apparent that this area in not an originator of the ongoing ictal activity. Third, the left central area C5 experienced an increase in the forward (C5–FP1) connection, both in absolute strength and directional interaction, suggesting that this area is the most likely driver of the ongoing ictal activity. That the central regions are now the drivers agrees with physiological evidence of low threshold of seizure initiation in motor-related areas (which are here receiving significantly increased stimulation current density by the smaller electrode).

Note that the anesthesia condition for both the normal and reversed configuration sessions were analogous and showed similar connectivity profiles. Moreover, both

sessions exhibited similar exogenous driving inputs under anesthesia and ictal conditions[Note here that there is a slight reduction of these inputs in the reversed configuration (Figs. 12b, 13b) and hence could be implicated to further emphasize the confidence in respective predicted strengths.] and hence the comparison across sessions shows both consistency and specificity of the predictions.

This preliminary study, therefore, supports the conclusion that the normal FEAST configuration used by clinicians does induce an ictal activity whose focal drive occurs in the frontal and particularly right frontal regions, and that the size of the frontal electrode (rather than other protocol parameters) is a promoter for this focality as a size reversal shifted the ictal activity to central (motor) regions with significantly weaker involvement of the frontal areas.

Comparison with Benchmark Effective Connectivity methods

The nonlinear, time-varying nature of the neuronal dynamics coupled with the ambiguity of unknown distributed inputs renders the model inversion problem highly challenging. Simplifications that are commonly undertaken in common effective connectivity measures including (i) linearity, (ii) fitting of slow observables (DCM, spectral output) rather than real time neural activity, (iii) random fluctuations at both the process and observation levels, and (iv) stationarity of the model are all assumptions that have largely been relaxed in the RA-SCKF framework.

For the simulated network, effective connectivity analysis using GPDC showed that time variations in the parameters are reflected in the overall character of the GPDC maps. In terms of causal inference, some connections were correctly identified while others incorrectly showed the existence of causal interaction. That this occurred can be explained by (a) the highly nonlinear character of the underlying model and (b) the ambiguity of the influence of unobserved inputs. This is not surprising particularly since GPDC, and other Vector AR models, assume that inputs are of modulatory nature (see Introduction) and are hence represented as random fluctuations of fixed power.

For the clinical data set, GPDC results showed a lack of sensitivity to the electrode configuration; that is, unlike the directional changes in connections obtained with RA-SCKF, the resultant GPDC did not show variability that is consistent with interchanging the location of large and small electrodes. Furthermore, the dramatic increase in causal connections during anesthesia for both configurations (in comparison with the respective seizure) is at odds with the principle of lack of communication in an anesthetized brain; rather, it emphasizes the need to account for inputs that are exogenous to the subnetwork under consideration (thalamus, deeper brain stem or other cortical areas).

As limitation of the current benchmarking step is lack of comparison with Dynamic Causal models. It is to be noted here that since DCM analysis depends on slow observables (commonly spectral density for EEG), simulated scenarios need to start from the same ground truth model (e.g. canonical microcircuit) and then incrementally relax the assumptions inherent in fitting models. These include: time variations in the extrinsic connections, length of available recordings, nonlinear character of the operating point, and random process fluctuations. At this point, such a comparison would seem improper and rather unfair: DCM is designed to perform multiple Bayesian comparisons and empirical inference over probability posteriors while RA-SCKF is designed for accurate blind deconvolution of sufficient statistics (means and covariances). On the other hand, RA-SCKF showed a marked improvement over SCKF (simulated data) in blind module identification. The latter has been shown superior to Dynamic Expectation Maximization (Bayesian inversion) even in the context of fMRI which is a slow observable. With EEG dynamics unfolding significantly faster than fMRI, it is reasonable to expect that such disparity in performance will even be larger. Clearly, an integrative scheme whereby filtering approaches are used guide the selection of models within DCM could reap the benefits of both approaches.

Generalization of the algorithm We have shown that our algorithm could specify the primary drives within the nodes of a neuronal sub-network. Aside from the simulated data and in-vivo analysis of ECT-induced seizures, showing distinct changes when alternating electrodes and polarity configurations, extensions of algorithm as a method for nonlinear identification can be understood by analogy to techniques in linear identification. Details of such extensions can be found in "Appendix".

Implications and Future Directions

The significant influx of data from novel methods for recording and interrogating brain function continues to highlight the need for efficient techniques in dynamic brain networks analysis (Sporns 2014). Since model-based analysis is often challenged by the resultant high-dimensionality complex estimation problems, functional-based analysis continue to be the main vehicle for investigating interrelations between nodes in the networks (Stephan et al. 2000; Hilgetag et al. 2002). As methods of detecting neuronal communities or modules continue to evolve (Müller-Linow et al. 2008; Sporns and Betzel 2016), developing model-based analysis techniques become more relevant. The current model-fitting approach is an instantiation of module effective connectivity in dynamic brain networks based

separating local module-based dynamics, and external influences.

The presented algorithm points to a novel framework for blind identification for a class on nonlinear, likely time varying, processes. The setup was inspired by common problems in neural process estimation from noisy observations and the often overlooked problem of inaccessible driving input was explicitly addressed. At the technical level, the framework proposes dissecting the residual error to infer systematic improvements in the input deconvolution, and concurrently in the modeling accuracy. Our methods of choice included ICA decomposition and local linearization that worked very well for the processes at hand. Still, the residual adjusted Kalman filtering framework is generalizable to include modifications of the various subblocks (ICA, Weiner), such as information-theoretic based error decompositions that avoid linearization assumptions (Pereda et al. 2005). While the algorithm increases the accuracy of input deconvolution under observation noise in standard CKF, it is also possible to further augment the noise robustness by inclusion of hybrid CKF algorithms (Arasaratnam et al. 2010) where the continuous process-discrete observations shown to be more resilient for larger noise components (Madi and Karameh 2017). In addition, a more systematic study of the impact of various subcomponents and tuning parameters on the algorithm performance can be subsequently conducted Still, the proposed solution was simple enough and worked quite well, under the reasonable assumptions for neural system modeling, to boost the accuracy of flexible and efficient nonlinear Kalman filtering techniques.

Since the developed framework was used to identify a subnetwork or module in an interconnected dynamic neuronal network, it can be thought of as implementing an ad-hoc nonlinear module identification in the context of neuronal modeling. From a system theoretic point of view, the concept of module identification, and generally the identification of dynamic networks, is a fairly new area of research with few theoretical results solely for linear systems and known inputs (Everitt et al. 2016; Everitt 2017; Dankers et al. 2013; Hof et al. 2013; Dankers et al. 2015, 2016). In brain networks, modules are local computational elements at individual nodes in a network with intrinsic connectivity and functionality (Park and Friston 2013; Rubinov and Sporns 2010), that might well be defined by cross linking structural and functional connection maps [Structure-function modules (Diez et al. 2015)]. The empirical positive results that are obtained here were mainly possibly by assuming a natural separation of the system dynamics into fast local dynamics (here modeled as the sub-network) and slower global dynamics (here summarized as the drive), which reasonably holds for the electrical activity in neuronal networks and could well hold for related functional brain imaging (Weaver et al. 2016; Jann et al. 2010; Thompson and Fransson 2015), and other physical processes. From a brain connectivity perspective, the proposed partial directed interaction PDI is a model-based measure of causal interaction between unobserved areas, and hence can be considered as an effective connectivity measure (Friston 2011) with the following modifications (i) it considers the impact of the connection as the source-normalized directed power contribution from source area to destination area, and (ii) it is based on within-module connections that were computed concurrently with external influences onto various areas of the modeled subnetwork.

At the neuronal modeling level, the framework proposes to improve connectivity predictions by explicitly accounting for unobserved slowly varying inputs impinging onto an active neuronal module even when these inputs do not admit parametric or stereotypical forms (i.e. pulse shapes or sinusoids). To the best of our knowledge, the present study undertook a unique approach to the largely unaddressed problem of subnetwork identification. Clearly, the blind identification procedure hinged upon several assumptions on the effect of exogenous inputs to the subnetwork. Further testing on different clinical or research applications and larger networks is necessary to clearly define regions for improvement and limitations.

The residual adjustment of the input within the estimation algorithm can be additionally constrained with other problem-dependent a priori information. Examples include correlated inputs that arrive to multiple areas in a modeled subnetwork, such as top down or feedback (higher order or attentional signals) and bottom up feed forward inputs (thalamic drive) (Fontolan et al. 2014; Dijkstra et al. 2017), and structural pathway information provided by connectome studies (Heuvel and Sporns 2011). Furthermore, selection of competing candidate model structures can be performed by quantifying the goodness of fit for a given structure using common measures (e.g. Bayesian information criterion).

Finally, it remains to be seen how a similar framework can in principle be devised for a variety of neuronal network modeling problems at the local scale (laminar interaction) (Pinotsis et al. 2017; Freestone et al. 2014; Yamamura et al. 2017) and across sensory subsystems (Jones et al. 2007; Bastos et al. 2015; Holland et al. 2016) to induce partial-directed interaction between various nodes in the subnetwork.

Funding This work has been supported by the Lebanese National Council for Scientific Research (LNCSR, Award No. 102630) and the following internal grant programs at AUB (1) the F Jabre Award for biomedical research, and (2) University Research Board award.

Compliance with Ethical Standards

Conflict of Interest Fadi N. Karameh has no conflict of interest. Ziad Nahas has had research funding from MECTA Inc. in form of FEAST device loan.

Ethical Approval All procedures performed in studies involving human participants were approved by the institutional review board (IRB) at the American University of Beirut under a FDA investigation device exemption and in accordance with the 1964 Helsinki declaration and its later amendments.

Informed Consent Informed consent was obtained from all individual participants included in the study.

Research Involving Human Participants and/or Animals This article does not contain any studies with animals performed by any of the authors

Appendix

Neural Population Models

Early efforts by Wilson and Cowan showed that modeling of oscillations in electric potential recordings, such as EEG, can be reasonably linked to the average firing activity of interacting excitatory and inhibitory within a unit area of neural tissue. Later extensions showed lumped parameter representations of alpha rhythms (Rotterdam et al. 1982) and basic cortical column dynamics (Jansen and Rit 1995). The latter model was the basis for the neural population model developed by Wendling in 2001 that accounts rhythm generation as an interaction between the main excitatory pyramidal, the local fast inhibitory cells and the slow inhibitory dendritic targeting cells. In this model, different interaction strengths among the three populations can generate spontaneous normal activity, sporadic spiking and sustained large amplitude low-frequency spikes as seen in seizures (Shayegh et al. 2011).

The basic building block for a single neural population in all of the aforementioned models is a rate-to-current-to-rate transformation. In a given neural population, an intracellular current results from the summed activity of synaptic elements H_e, H_i representing excitatory and inhibitory connection of the population. This current is then mapped to an output firing rate by a sigmoidal transformation. A synapse $H_l, l \in \{e, i\}$ is thought of as a second order linear system whose impulse response follows a double-exponential (or alpha) function $h_l(t) = A_l a_l t e^{-a_l t}$, where a_l is the time constant and A_l is an amplitude constant for that synapse (Fig. 16). Accordingly, for a given input rate $r_l(t)$ arriving at this synapse, the corresponding postsynaptic current $x_l(t)$ can be computed using a two-state linear system by introducing an intermediate state $z(t)$ in

$$\begin{aligned} \dot{x}_l(t) &= z(t) \\ \dot{z}(t) &= A_l a_l r_l(t) - 2a_l z(t) - a_l^2 x_l(t) \end{aligned} \tag{45}$$

Finally, the population output is the firing rate which is produced after scaling and passing all synaptic currents $x(t) = \sum_l C_l x_l(t)$ through a static sigmoidal nonlinearity $r(t) = S\{x(t)\}, S\{x\} = \frac{2e}{1+e^{k(x_o - x)}}$.

As one of the earliest neural population models to simulate seizures, the Jansen model could reproduce basic sustained spike discharges but the fast EEG associated with low-voltage rapid discharges. To accommodate the latter, Wendling added to the slow inhibition, present in the Jansen model, a second class of inhibitory interneurons with faster dynamics, namely the somatic targeting fast interneurons (FIN) with fast GABAergic synapsis (Wendling et al. 2002).

Induced Seizure Neuronal models

The Wendling model became hence a popular tool to investigate spontaneous seizure genesis in pathological tissue. Recently, our group has demonstrated that the Wendling model can be readily modified to account for seizure induction in models of normal tissue by external stimuli such as those applied during ECT (Karameh et al. 2014).

Spontaneous Seizure Model

In the basic Wendling model (Wendling et al. 2002), a local area is represented by a primary pyramidal cell population that connects with three distinct interneuron populations that are common across cortical structures. As shown in Fig. 17, the pyramidal cell population (PYR) are reciprocally

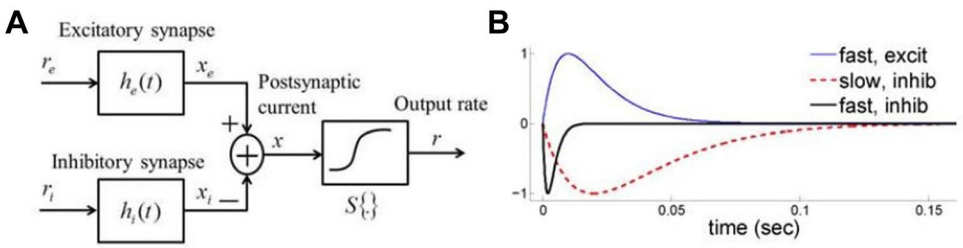

Fig. 16 Basic building blocks for a neuronal model. **a** excitatory and inhibitory blocks. **b** impulse responses of the various synapses

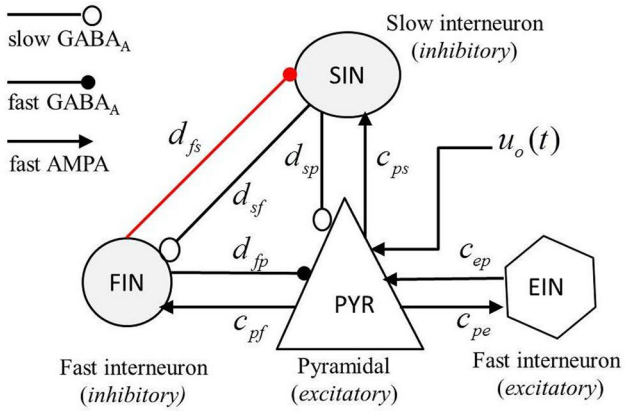

Fig. 17 Local Area model of a neuronal population that can be driven into an induced seizure by external inputs. Model is modified from (Wendling et al. 2002) to include the fast-to-slow inhibitory connection d_{fs} (in red)

connected to (i) a local population of excitatory interneurons (EIN) that provide excitatory feedback to the pyramidal cells (ii) a population of fast inhibitory interneurons (FIN) that provides fast GABA-A somatic inhibition to PYR, and finally, (iii) a population of slow inhibitory interneurons (SIN) that provide slow GABA-A dendritic inhibition to the PYR as well as inhibit the FIN population. Starting with the above formulation of single synapses, the local area is represented by the following set of non-linear state-space equations:

$$\dot{x}_p(t) = y_1(t) \tag{46}$$

$$\begin{aligned} \dot{y}_1(t) &= AaS\{x_e(t) - x_{is} - x_{if}(t)\} - 2ay_1(t) - a^2 x_p(t) \\ \dot{x}_e(t) &= y_2(t) \end{aligned} \tag{47}$$

$$\begin{aligned} \dot{y}_2(t) &= Aa\left[u_0(t) + c_{ep} S\{c_{pe} x_p(t)\}\right] - 2ay_2(t) - a^2 x_e(t) \\ \dot{x}_{is}(t) &= y_3(t) \end{aligned} \tag{48}$$

$$\begin{aligned} \dot{y}_3(t) &= Bbd_{sp} S\{c_{ps} x_p(t)\} - 2by_3(t) - b^2 x_{is}(t) \\ \dot{x}_{if}(t) &= y_4(t) \end{aligned} \tag{49}$$

$$\begin{aligned} \dot{y}_4(t) &= Ggd_{fp} S\{c_{pf} x_p(t) - d_{sf} x_{ii}(t)\}) - 2gy_4(t) - g^2 x_{if} \\ \dot{x}_{ii}(t) &= y_5(t) \end{aligned} \tag{50}$$

$$\dot{y}_5(t) = BbS\{c_{ps} x_p(t)\} - 2by_5(t) - b^2 x_{ii}(t) \tag{51}$$

$$z(t) = K_{iv}\left(x_e(t) - x_{is} - x_{if}(t)\right) \tag{52}$$

where $x_p(t)$ is the synaptic current output from pyramidal cell population to all local cells; x_e, x_{is}, and x_{if} are, respectively, the excitatory, the slow GABA-A inhibitory input and the fast GABA-A inhibitory inputs to the pyramidal cells); x_{ii} is the slow inhibition from SIN to FIN population. The constants (A, a), (B, b) and (G, g) dictate the time profiles of postsynaptic currents associated with connections to EIN, FIN, SIN populations, respectively. $u_0(t)$ is an external input assumed to arrive as a firing rate to the local pyramidal population. The output of the model is the voltage trace $z(t)$ which is proportional to the over all postsynaptic current in the pyramidal cell population and is assumed to be representative of the EEG traces of the overall local area. Finally, the constants $d_{m,n}$ and $c_{m,n}$ denote the excitatory and inhibitory connection strengths, respectively, originating from population m to population n and $(m, n) \in \{$ p: PYR, e: EIN, f: FIN, s SIN $\}$. For concise reference, the above continuous time state-space nonlinear system can written as

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \tilde{f}(\mathbf{x}(t), u(t)) \\ z(t) &= \tilde{h}(\mathbf{x}(t)) \end{aligned} \tag{53}$$

where $\mathbf{x}(t) \in \mathbb{R}^{10}$ are the firing states of this system, $\tilde{f}$ is the process nonlinear function, and $\tilde{h}$ is the measurement function. The above model produces, for a given set of fixed connections parameters, an oscillatory behavior that change little in character regardless of the applied input $u_o(t)$. That is, the dynamics of the nonlinear system exhibit single stable solutions for a given parameter set. For example, a change in the balance of slow-to-fast connection parameters can move the oscillations from normal to sporadic spikes and into sustained slow quasi-sinusoidal epileptic activity. Accordingly, the model could produce spontaneous seizures caused by pathology (functional imbalance) in the local population, but not induced seizures which are initiated by external inputs, such as those occurring under ECT.

Electrically-Induced Seizure Model

To accommodate the induction of seizures in normal neural population models, we have previously introduced a simple modification to the basic Wendling model that is physiological in nature and that allows for the normal activity to switch to seizure activity when subjected to large external stimuli without changing internal model parameters (Karameh et al. 2014), as shown in Fig. 18. The basic idea is rather simple: specifically, normal levels of dendritic inhibition act to balance the strength of excitatory afferents on the local pyramidal populations while reduce inhibition (or disinhibition) in this region will cause a runaway excitation. Several experimental

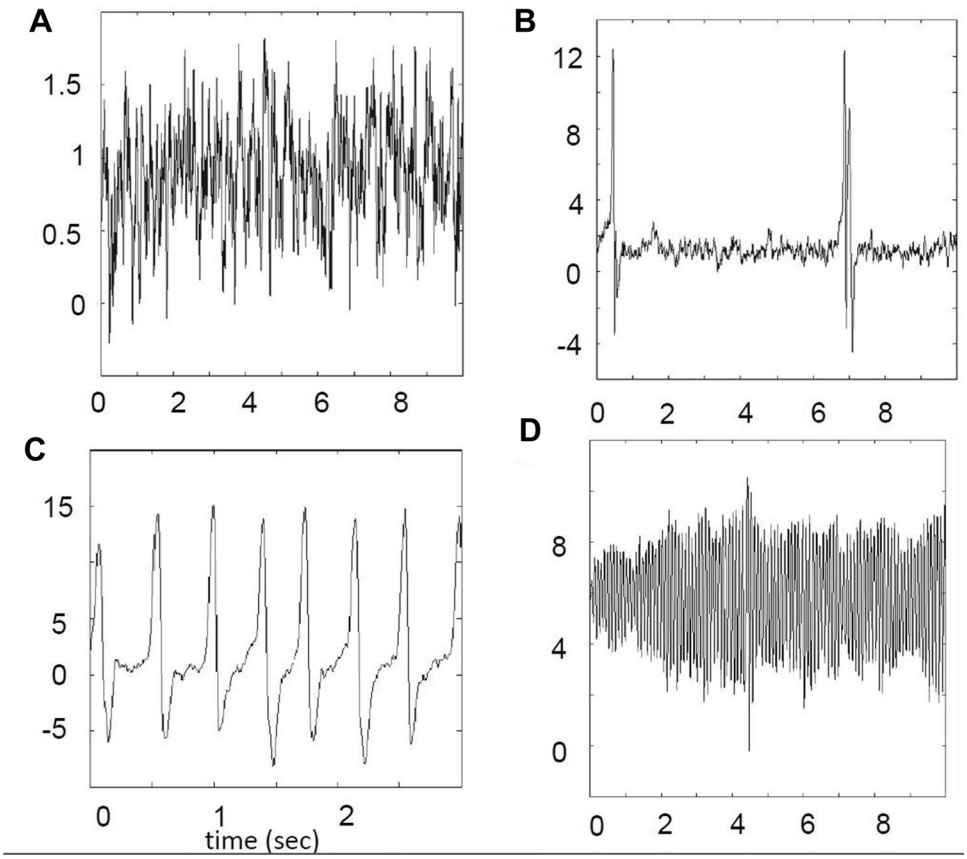

Fig. 18 Various EEG patterns produced by the multi-area model. **a** Background activity. **b** Sporadic spikes. **c** Sustained slow oscillations. **d** Seizures [for details, cf. (Karameh et al. 2014)]

lines of evidence suggest that impaired dendritic inhibition may play in important role in spontaneously occurring seizures (Wendling et al. 2005). Particularly, fast inhibition show significant firing increase (Curtis and Gnatkovsky 2009) and exhibit depolarization block (Cammarota et al. 2013) just prior to seizure onset, and the collapse of the inhibitory control lead to runaway pyramidal excitation (Trevelyan and Schevon 2013). The reciprocal connections between slow dendritic targeting interneurons and fast interneurons have also been reported in the middle (Beierlein et al. 2003) and upper cortical layers (Cruikshank et al. 2012; Markram et al. 2004), show preference in interneuron targeting (Pfeffer et al. 2013) with slow GABA-A inhibitory cells targeted by fast spiking cells (Staiger et al. 1997; Tamás et al. 2004; Palmer et al. 2012). We have therefore proposed the existence of an inhibitory pathway from the fast spiking FIN interneurons to the SIN populations that can under large external stimuli act to inhibit the SIN population, thereby creating brief dendritic disinhibition in the PYR population.

Accordingly, the Wendling model for a single area is modified so as to include a fast inhibitory to slow inhibitory neuron connection (Fig. 17, red line). The corresponding nonlinear state-space equations for the SIN input synapse becomes

$$\begin{aligned}\dot{x}_{is}(t) &= y_3(t)\\ \dot{y}_3(t) &= Bbd_{sp}S\{c_{ps}x_p(t) - d_{fs}x_{if}(t)\} - 2by_3(t) - b^2x_{is}(t)\end{aligned} \tag{54}$$

When subjected to increasing levels of external stimulation, the modified Wendling (MW) model for a single area switched from a baseline normal background activity first into sporadic spikes and then into continuous low-frequency spikes similar to those observed in ECT induced seizures. The MW model was also utilized in a multi-area simulation where the interareal connections emanate from pyramidal cells of one area and contact both the pyramidal and FIN population of the other areas, as seen in Fig. 2). For a model of two interconnected areas and with only one area subjected to external stimulus, it was noted that the threshold for seizure induction in that area is now lowered and that the ensuing seizure also propagates to the other areas even though that area was not under the influence of the stimulus. Thus the MW model accounts for both the initiation and the propagation of electrically-induced seizures in normal tissue (see Karameh et al. 2014) for details).

Induced Seizure in a Network

For a total of M distinct EEG channels recordings, a network with M distinct blocks is constructed. Between any two channels, reciprocal connections are made via long-range excitatory synapses to the local pyramidal PYR population and its associated fast interneuron population FIN (as in Fig. 2). We also assume that long range connections from area q to area r incur a fixed propagation delay τ_{qr}. This shows as an extra synaptic input in the corresponding equations (Eqs. 47, (49) of the local area Modified Wendling equations as follows

$$\dot{y}_2^{(r)}(t) = Aa\left[u_r(t) + c_{ep}S\left\{c_{pe}x_p^{(r)}(t) + \sum_{\substack{q=1\\ q\neq r}}^{M} K_{qr}x_p^{(q)}(t-\tau_{qr})\right\}\right] - 2a\dot{y}_2^{(r)}(t) - a^2 x_e^{(r)}(t) \tag{55}$$

$$\dot{y}_4^{(r)}(t) = Ggd_{fp}S\left\{c_{pf}x_p^{(r)}(t) + \sum_{\substack{q=1\\ q\neq r}}^{M} L_{qr}x_p^{(q)}(t-\tau_{qr}) - d_{sf}x_{ii}^{(r)}(t)\right\} - 2g\dot{y}_4^{(r)}(t) - g^2 x_{if}^{(r)}(t) \tag{56}$$

where $x_p^{(q)}(t)$ is the output from the pyramidal cells in area q. K_{qr} and L_{qr} are the connections from pyramidals in area q to the pyramidal and fast inhibitory cells in area r, respectively.

Kalman Filters for System Modeling

From a systems point of view, physiological recordings, such as EEG $e(t)$ in Eq. 52, can often be thought of imperfect measurements or noisy observations related to an underlying dynamical system. The dynamical system response, in turn, can be decomposed into constituent internal states that are either physical in nature (e.g. neural firing rates $x_e(t)$ in Eq. 47) or mathematical entities that allow determining the time trajectory of the system output. In what follows, we will briefly outline Kalman filtering approaches to estimate the internal states of a system that fits the temporal dynamics of the observations.

State Estimation

We here consider the discrete-time state space description of a dynamical system that has the following form

$$\begin{aligned} \text{Process equation: } \mathbf{x}_k &= f(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}) + \boldsymbol{\mu}_{k-1} \\ \text{Measurement equation: } \mathbf{z}_k &= h(\mathbf{x}_k, \mathbf{u}_k) + \boldsymbol{\eta}_k \end{aligned} \tag{57}$$

where $\mathbf{x}_k \in \mathbb{R}^n$ is the n dimensional state vector at discrete time k; $f : \mathbb{R}^n \times \mathbb{R}^{n_u} \to \mathbb{R}^n$ is a known function that describes that states as a function of their history and of a control input $\mathbf{u}_k \in \mathbb{R}^{n_u}$; $h : \mathbb{R}^n \times \mathbb{R}^{n_u} \to \mathbb{R}^{n_z}$ is an known observation function that relates the measurements $\mathbf{z}_k \in \mathbb{R}^{n_z}$ to the hidden states of the system; finally, $\boldsymbol{\mu}_k$ and $\boldsymbol{\eta}_k$ are additive independent process and observation Gaussian-distributed noise sequences with zero means and covariances $\mathbf{Q}_k$ and $\mathbf{R}_k$ respectively. In the case when both h and f are known, finding the hidden states based on the noise observations is an estimation problem which will considered next. In the case when at least one of the functions h and f is partially unknown, finding the hidden states and the model unknowns is a model learning problem which will be considered after introducing the cubature Kalman filters.

Strictly speaking, however, a majority of the practical physiological systems are nonlinear (in the problem at hand, for example, the sigmoidal functions introduce nonlinearity). This implies the existence of a nonlinearity in either (or both) of the functions f and h. and hence a nonlinear transformation of the Gaussian priors in the above integrals, thus obfuscating the existence of simple general solutions. To overcome this problem, the extended Kalman filter (EKF) provides an approximate solution by linearizing the nonlinear dynamics $f(\mathbf{x}_{k-1}, \mathbf{u}_{k_1})$ locally around an estimated state by a linearized system, which is limited by the accuracy of this linearization. The unscented Kalman filter (UKF) provides an approximate solution to the integral(s) as a weighted summations of the nonlinearly transformed set of sigma points (so called unscented transformation). The UKF had earlier been a popular choice to solve these problems but suffered from numerical instabilities, which were mitigated by the Cubature Kalman filter (CKF), described next.

Developed by Arasaratnam and Haykin in 2009 (Arasaratnam and Haykin 2009), the Cubature Kalman Filter (CKF) approximated the integrals of the form (nonlinear × Gaussian) for n-dimensional vectors $\mathbf{x} \in \mathbb{R}^n$ using third degree spherical cubature rule as

$$\int f(\mathbf{x}) \times \mathcal{N}(\mathbf{x};\mu,\Sigma)d\mathbf{x} \approx \frac{1}{2n}\sum_{i=1}^{2n} f(\mathbf{X}_i) \tag{58}$$

where $\mathbf{X}_i$ are the set of $2n$ sigma points that symmetrically sample the distribution at locations $\mathbf{X}_i = \mu + \xi_i\sqrt{\Sigma}$, and the spread factor $\xi_i = \sqrt{n}e_i$, $i = 1 \ldots n$ and $\xi_i = -\sqrt{n}e_i$, $i = n+1 \ldots 2n$ with e_i as a unit vector for the *ith* dimension. With the time and measurement update steps involve approximate computations of covariance matrices, numerical inaccuracies might still incorrectly produce matrices that are not positive semidefinite leading to algorithm divergence. This however does not occur in a variant of the CKF, the

Square root Cubature Kalman filter (SCKF), which instead computes and propagates the square root of the covariance matrices. The SCKF will be used here and a brief description of the algorithm is given in the appendix.

Joint Estimation

The standard Kalman setup in Eq. (57) estimates the hidden states for known control inputs $\mathbf{u}_k$ and for known process and observation functions $f(.)$ and $h(.)$, respectively. In the case where the input is unknown, or the functions are partially unknown and parameterizable by an unknown parameter vector $\boldsymbol{\theta} \in \mathbb{R}^{n_\theta}$ (i.e. $f\left(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}\right) = f\left(\mathbf{x}_{k-1}, \boldsymbol{\theta}_{k-1}, \mathbf{u}_{k-1}\right)$), Kalman filters allow to augment the state vector to search for $\mathbf{u_k}$ and $\boldsymbol{\theta}$ as part of the hidden states by commonly assuming these to follow random walk processes, leading to the following overall state-space system representation

$$\mathbf{u}_k = \mathbf{u}_{k-1} + \mathbf{w}_{k-1} \tag{59}$$

$$\boldsymbol{\theta}_k = \boldsymbol{\theta}_{k-1} + \boldsymbol{v}_{k-1} \tag{60}$$

$$\mathbf{x}_k = f\left(\mathbf{x}_{k-1}, \boldsymbol{\theta}_{k-1}, \mathbf{u}_{k-1}\right) + \boldsymbol{\mu}_{k-1} \tag{61}$$

$$\mathbf{z}_k = h(\mathbf{x}_k, \mathbf{u}_k) + \boldsymbol{\eta}_k \tag{62}$$

where $\mathbf{w}_{k-1}$ and $\boldsymbol{v}_{k-1}$ are independent Gaussian-distributed sequences with zero mean and covariances $\boldsymbol{\Lambda}_{W,k}$ and $\boldsymbol{\Lambda}_{V,k}$, respectively. The equations above constitute joint-estimation problem for an augmented hidden state vector $\left[\mathbf{u}_k^\top \boldsymbol{\theta}_k^\top \mathbf{x}_k^\top\right]^\top \in \mathbb{R}^{(n_u+n_\theta+n)}$ thereby allowing the SCKF to estimate time-varying control inputs and, particularly, time varying parameters which is a very desirable property in the case where the nonlinearity is not fixed in time.

The SCKF Algorithm

The main SCKF algorithm is presented next, where all of the steps can be deduced directly from the CKF except for the update of the posterior error covariance.

Time Update

1) Evaluate the cubature points ($i = 1, 2, \ldots m$)

$$\mathbf{X}_{i,k-1|k-1} = \mathbf{S}_{k-1|k-1}\boldsymbol{\xi}_i + \hat{\mathbf{x}}_{k-1} \tag{63}$$

where $m = 2n_x$

2) Evaluate the propagated cubature points ($i = 1, 2, \ldots m$)

$$\mathbf{X}^*_{i,k|k-1} = f(\mathbf{X}_{i,k-1|k-1}, \mathbf{u}_{k-1}) \tag{64}$$

3) Estimate the predicted state

$$\hat{\mathbf{x}}_{k|k-1} = \frac{1}{m}\sum_{i=1}^{m} \mathbf{X}^*_{i,k|k-1} \tag{65}$$

4) Estimate the square-root factor of the predicted error covariance

$$\mathbf{S}_{k|k-1} = \mathbf{Tria}([\mathcal{X}^*_{k|k-1} \quad \mathbf{S}_{Q,k-1}]) \tag{66}$$

where $\mathbf{S}_{Q,k-1}$ denotes a square-root factor of $\mathbf{Q}_{k-1}$ such that $\mathbf{Q}_{k-1} = \mathbf{S}_{Q,k-1}\mathbf{S}^T_{Q,k-1}$ and the weighted centered(prior mean is subtracted off) matrix

$$\mathcal{X}^*_{k|k-1} = \frac{1}{\sqrt{m}}[\mathbf{X}^*_{1,k|k-1} - \hat{\mathbf{x}}_{k|k-1} \quad \mathbf{X}^*_{2,k|k-1} - \hat{\mathbf{x}}_{k|k-1} \ldots \mathbf{X}^*_{m,k|k-1} - \hat{\mathbf{x}}_{k|k-1}] \tag{67}$$

Measurement Update

1) Evaluate the cubature points ($i = 1, 2, ..m$)

$$\mathbf{X}_{i,k|k-1} = \mathbf{S}_{k|k-1}\boldsymbol{\xi}_i + \hat{\mathbf{x}}_{k|k-1} \tag{68}$$

2) Evaluate the propagated cubature points ($i = 1, 2, ..m$)

$$\mathbf{Z}_{i,k|k-1} = h(\mathbf{X}_{i,k|k-1}, \mathbf{u}_k) \tag{69}$$

3) Estimate the predicted measurement

$$\hat{\mathbf{z}}_k = \frac{1}{m}\sum_{i=1}^{m} \mathbf{Z}_{i,k|k-1} \tag{70}$$

4) Estimate the square-root of the innovation covariance matrix

$$\mathbf{S}_{zz,k|k-1} = \mathbf{Tria}([\mathbf{Z}^*_{k|k-1} \quad \mathbf{S}_{R,k}]) \tag{71}$$

where $\mathbf{S}_{R,k}$ denotes a square-root factor of $\mathbf{R}_k$ such that $\mathbf{R}_k = \mathbf{S}_{R,k}\mathbf{S}^T_{R,k}$ and the weighted centered matrix

$$\mathcal{Z}_{k|k-1} = \frac{1}{\sqrt{m}}[\mathbf{Z}_{1,k|k-1}\hat{\mathbf{z}}_k \quad \mathbf{Z}_{2,k|k-1} - \hat{\mathbf{z}}_k \ldots \mathbf{Z}_{m,k|k-1} - \hat{\mathbf{z}}_k] \tag{72}$$

5) Estimate the cross-covariance matrix

$$\mathbf{P}_{xz,k|k-1} = \mathcal{X}_{k|k-1}\mathcal{Z}^T_{k|k-1} \tag{73}$$

where the weighted, centered matrix

$$\mathcal{X}_{k|k-1} = \frac{1}{\sqrt{m}}\left[\mathbf{X}_{1,k|k-1} - \hat{\mathbf{x}}_{k|k-1} \quad \mathbf{X}_{2,k|k-1} - \hat{\mathbf{x}}_{k|k-1} \ldots \mathbf{X}_{m,k|k-1} - \hat{\mathbf{x}}_{k|k-1}\right] \tag{74}$$

6) Estimate the Kalman gain

$$\mathbf{G}_k = (\mathbf{P}_{xz,k|k-1}/\mathbf{S}^T_{zz,k|k-1})/\mathbf{S}_{zz,k|k-1} \tag{75}$$

7) Estimate the updated state

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k|k-1} + \mathbf{G}_k(\mathbf{z}_k - \hat{\mathbf{z}}_k) \tag{76}$$

8) Estimate the square-root factor of the corresponding error covariance

$$\mathbf{S}_{k|k} = \mathbf{Tria}([\mathcal{X}_{k|k-1} - \mathbf{G}_k\mathcal{Z}_{k|k-1} \quad \mathbf{G}_k\mathbf{S}_{R,k}]) \tag{77}$$

Generalization of the RA-SCKF Algorithm

It is noted that the developed procedure shares some properties with two-stage approaches common in linear system identification, namely the instrumental variables and the two-step least squares (Ljung 1999). These approaches are utilized whenever the residual (prediction error) of a linear regression (the first stage) is correlated with the regressors (including inputs) leading to biased parameters estimates. Starting with these biased estimates, the regressors for the second stage are adjusted to remove any correlation with the residual and hence reduce biases. By analogy, upon dissecting the error of the first stage SCKF into channel-specific components, inverting the linearized inout/output map to find a desired deviation signal, and then finding its linear predictability from the system input, the algorithm incorporates a gain corresponding to the input-correlated residual into the estimation structure. This allowing the second stage SCKF (the residual adjustment) to further reduce the bias in the estimates.

The presented mapping of the independent components is based on the assumption that all the channels are driven by distinct inputs. Here, a one-to-one mapping between the ICA-extracted components and the channel-specific errors was formed. In principle, this procedure can be modified to account for any segmentation or clustering of the channel-specific inputs to common or distinct sets of inputs. To find this segmentation, it is best to utilize a priori knowledge of common pathways that drive separate channels. Still, in the absence of such knowledge, it is also plausible to create different candidate mappings of the ICA-extracted components onto channel-specific errors. For example, a channel-specific error e_c can be formed from every independent component j to each channel i (20, $\mathscr{P}$ is all combinations), along with the corresponding desired input deviation (Eq. 34) and the corresponding linear predictability (Eq. 37). A mapping from j to i is to be created only if such predictability exceeds a predefined threshold. The channel-specific error can then be constructed incrementally starting with components with the next highest predictability. A summary of the proposed algorithm is shown below. Subsequently, after all the mapping are formed, a residual-adjustment SCKF run is conducted for the candidate segmentation and its log-likelihood is computed. Out of all the segmentations, the winner is selected based on the largest log likelihood.

Algorithm 2 Generalization of the Residual-Adjusted Kalman Algorithm

Initialization: for $k = 1 \ldots N$ and for n_z channels, obtain SCKF estimate of parameters $\hat{\theta}_k$, inputs $\hat{\mathbf{u}}_k$, their square-root process noise covariances ($\mathbf{S}_{V,k}$, $\mathbf{S}_{W,k}$ respectively), outputs $\hat{z}_k$, and prediction errors e_k.
set a segmentation order g

function FIND DESIRED INPUT($\mathbf{z}_k$,$\mathbf{e}_k$)
  for $m = 0, 1, \ldots$ **do**
    Buffer $\mathbf{e}_k$, $\mathbf{u}_k$, $\mathbf{z}_k$ into $\mathbf{E}_m$, $\mathbf{u}_m$ and $\mathbf{z}_m$ (resp.), $k = (m-1)L+1 : mL$ (eq. 11).
    Apply fast ICA on $\mathbf{E}_m$ to compute n_z independent components $\mathbf{c}_m$ (eq.12)

    function FIND SEGMENTED ERROR($\mathbf{z}_m$,$\mathbf{u}_m$, $\mathbf{c}_m$,g)
      find common component error e_c^i and its index set $\mathscr{J}_c$ (eqs.15-19)
      for each channel i **do**
        for each component $j \notin \mathscr{J}_c$ **do**
          reconstruct candidate error $e_j^i = a_{i,j}c_j$.
          find corresponding desired deviation signal $\mathbf{d}_u^{i,j}$ (eqs. 30- 34).
          find best linear prediction $\hat{\mathbf{d}}_u^{i,j}$ of $\mathbf{d}_u^{i,j}$ from $\mathbf{u}_m^i$ (eq. 36)
          determine approximative power $C^{i,j}(\Omega)$ of $\hat{\mathbf{d}}_u^{i,j}$ (eq. 37).
          if $C^{i,j}(\Omega) > \eta_c$ **then**
            add j to index set Γ^i based on decreasing $C^{i,j}(\Omega)$
            assign i to the set with channel-updated inputs $\mathscr{I}_s$
          end if
        end for
        set $e_s^i = \sum_{j\in\Gamma^i,\, j=1}^{g} a_{i,j}c_j$ (largest g components).
      end for
    end function

  end for
end function

for $i \in \mathscr{I}_s$ **do** ▷ *Propagate the error* $\mathbf{e}_s^i$ *back to deviations* δ_u^i *in the input to channel i*
  (eqs. 30- 34).
end for

function UPDATE COVARIANCES($\mathbf{d_k}$,$\mathbf{u}_k$, $\mathbf{S}_{V,k}$,$\mathbf{S}_{W,k}$)
  (eqs. 35-40)
end function

Apply SCKF algorithm starting from initial state vector ($\hat{\mathbf{x}}_0 = \hat{\mathbf{x}}_N$) and with the updated $\mathbf{S}_{V,k}$, $\mathbf{S}_{W,k}$

Supplementary Connectivity Tables

See (Table 3).

Table 3 Average absolute interconnection strengths for ictal activity under (a) normal and (b) reverse configurations for Subject 20 (self connections are fixed, and numbers in parentheses are percentage change from normal)

	(a) Normal FEAST				(b) Reverse configuration FEAST			
	FP2	C6	FP1	C5	FP2	C6	FP1	C5
FP2	135.0	175.7	135.2	–	135.0 (0%)	97.5(– 45)	122.0 (10)	–
C6	224.5	135.0	–	186.7	63.4 (– 72)	135.0 (0)	–	80.4 (– 57)
FP1	86.8	–	135.0	200.6	109.2 (– 21)	–	135.0 (0)	241.8 (21)
C5	–	260.2	156.7	135.0	–	51.0 (– 80)	135.6 (13)	135.0 (0)

References

Ackermann RF, Engel J, Baxter L (1986) Positron emission tomography and autoradiographic studies of glucose utilization following electroconvulsive seizures in humans and rats. Annal N Y Acad Sci 462(1):263–269

Ambrogioni L, Hinne M, Van Gerven M, Maris E (2017) Gp cake: effective brain connectivity with causal kernels. In: Advances in Neural Information Processing Systems, pp 951–960

Arasaratnam I, Haykin S (2009) Cubature kalman filters. IEEE Trans Autom Control 54(6):1254–1269

Arasaratnam I, Haykin S, Hurd TR (2010) Cubature kalman filtering for continuous-discrete systems: theory and simulations. IEEE Trans Signal Process 58(10):4977–4993

Barrett AB, Murphy M, Bruno MA, Noirhomme Q, Boly M, Laureys S, Seth AK (2012) Granger causality analysis of steady-state electroencephalographic signals during propofol-induced anaesthesia. PLoS ONE 7(1):e29–072

Bastos AM, Litvak V, Moran R, Bosman CA, Fries P, Friston KJ (2015) A dcm study of spectral asymmetries in feedforward and feedback connections between visual areas v1 and v4 in the monkey. Neuroimage 108:460–475

Beierlein M, Gibson JR, Connors BW (2003) Two dynamically distinct inhibitory networks in layer 4 of the neocortex. J Neurophysiol 90(5):2987–3000

Bell AJ, Sejnowski TJ (1995) An information-maximization approach to blind separation and blind deconvolution. Neural Comput 7(6):1129–1159

Bielczyk NZ, Llera A, Buitelaar JK, Glennon JC, Beckmann CF (2017) Increasing robustness of pairwise methods for effective connectivity in magnetic resonance imaging by using fractional moment series of bold signal distributions. arXiv:160608724v3

Cammarota M, Losi G, Chiavegato A, Zonta M, Carmignoto G (2013) Fast spiking interneuron control of seizure propagation in a cortical slice model of focal epilepsy. J Physiol 591(4):807–822

Canolty RT, Knight RT (2010) The functional role of cross-frequency coupling. Trends Cogn Sci 14(11):506–515

Correa N, Adalı T, Calhoun VD (2007) Performance of blind source separation algorithms for fmri analysis using a group ica method. Magn Reson Imaging 25(5):684–694

Cruikshank SJ, Ahmed OJ, Stevens TR, Patrick SL, Gonzalez AN, Elmaleh M, Connors BW (2012) Thalamic control of layer 1 circuits in prefrontal cortex. J Neurosci 32(49):17

Crunelli V, David F, Lőrincz ML, Hughes SW (2015) The thalamocortical network as a single slow wave-generating unit. Curr Opin Neurobiol 31:72–80

Damaraju E, Allen E, Belger A, Ford J, McEwen S, Mathalon D, Mueller B, Pearlson G, Potkin S, Preda A et al (2014) Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia. NeuroImage 5:298–308

Dankers A, Van den Hof PM, Bombois X, Heuberger PS (2015) Errors-in-variables identification in dynamic networks-consistency results for an instrumental variable approach. Automatica 62:39–50

Dankers A, Van den Hof PM, Bombois X, Heuberger PS (2016) Identification of dynamic models in complex networks with prediction error methods: predictor input selection. IEEE Trans Autom Control 61(4):937–952

Dankers A, Van den Hof PM, Bombois X, Heuberger PS (2013) Predictor input selection for two stage identification in dynamic networks. In: Control Conference (ECC), 2013 European, IEEE, pp 1422–1427

David O, Kiebel SJ, Harrison LM, Mattout J, Kilner JM, Friston KJ (2006) Dynamic causal modeling of evoked responses in eeg and meg. NeuroImage 30(4):1255–1272

De Curtis M, Gnatkovsky V (2009) Reevaluating the mechanisms of focal ictogenesis: the role of low-voltage fast activity. Epilepsia 50(12):2514–2525

Deng ZD, Lisanby SH, Peterchev AV (2011) Electric field strength and focality in electroconvulsive therapy and magnetic seizure therapy: a finite element simulation study. J Neural Eng 8(1):016007

Destexhe A, Rudolph M, Paré D (2003) The high-conductance state of neocortical neurons in vivo. Nat Rev Neurosci 4(9):739

Diez I, Bonifazi P, Escudero I, Mateos B, Muñoz MA, Stramaglia S, Cortes JM (2015) A novel brain partition highlights the modular skeleton shared by structure and function. Sci Rep 5(srep10):532

Dijkstra N, Zeidman P, Ondobaka S, Gerven MA, Friston K (2017) Distinct top-down and bottom-up brain connectivity during visual perception and imagery. Sci Rep 7(1):5677

Dubeau S, Havlicek M, Beaumont E, Ferland G, Lesage F, Pouliot P (2012) Neurovascular deconvolution of optical signals as a proxy for the true neuronal inputs. J Neurosci Methods 210(2):247–258

Enev M, McNally KA, Varghese G, Zubal IG, Ostroff RB, Blumenfeld H (2007) Imaging onset and propagation of ect-induced seizures. Epilepsia 48(2):238–244

Engel AK, Fries P, Singer W (2001) Dynamic predictions: oscillations and synchrony in top-down processing. Nat Rev Neurosci 2(10):704

Everitt N (2017) Module identification in dynamic networks: parametric and empirical bayes methods. PhD thesis, KTH Royal Institute of Technology

Everitt N, Bottegal G, Rojas CR, Hjalmarsson H (2016) Identification of modules in dynamic networks: an empirical bayes approach. In: 2016 IEEE 55th Conference on Decision and control (CDC), IEEE, pp 4612–4617

Fontolan L, Morillon B, Liegeois-Chauvel C, Giraud AL (2014) The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex. Nat Commun 5:4694

Franks NP (2008) General anaesthesia: from molecular targets to neuronal pathways of sleep and arousal. Nat Rev Neurosci 9(5):370

Frässle S, Lomakina EI, Kasper L, Manjaly ZM, Leff A, Pruessmann KP, Buhmann JM, Stephan KE (2018) A generative model of whole-brain effective connectivity. NeuroImage 179:505–529

Freestone DR, Karoly PJ, Nešić D, Aram P, Cook MJ, Grayden DB (2014) Estimation of effective connectivity via data-driven neural modeling. Front Neurosci 8:383

Friston KJ (2011) Functional and effective connectivity: a review. Brain Connect 1(1):13–36
Friston K, Moran R, Seth AK (2013) Analysing connectivity with granger causality and dynamic causal modelling. Curr Opin Neurobiol 23(2):172–178
Garrido MI, Friston KJ, Kiebel SJ, Stephan KE, Baldeweg T, Kilner JM (2008) The functional anatomy of the mmn: a dcm study of the roving paradigm. Neuroimage 42(2):936–944
Goebel R, Roebroeck A, Kim DS, Formisano E (2003) Investigating directed cortical interactions in time-resolved fmri data using vector autoregressive modeling and granger causality mapping. Magn Reson Imaging 21(10):1251–1261
Grover S, Mattoo SK, Gupta N (2005) Theories on mechanism of action of electroconvulsive therapy. German J Psychiatry 8:70–84
Havlicek M, Jan J, Brazdil M, Calhoun VD (2010) Dynamic granger causality based on kalman filter for evaluation of functional network connectivity in fmri data. Neuroimage 53(1):65–77
Havlicek M, Friston KJ, Jan J, Brazdil M, Calhoun VD (2011) Dynamic modeling of neuronal responses in fmri using cubature kalman filtering. NeuroImage 56(4):2109–2128
Hilgetag CC, Kötter R, Stephan KE, Sporns O (2002) Computational methods for the analysis of brain connectivity. In: Ascoli G (ed) Computational neuroanatomy. Springer, New York, pp 295–335
Holland R, Leff AP, Penny WD, Rothwell JC, Crinion J (2016) Modulation of frontal effective connectivity during speech. NeuroImage 140:126–133
Honey CJ, Kötter R, Breakspear M, Sporns O (2007) Network structure of cerebral cortex shapes functional connectivity on multiple time scales. Proc Natl Acad Sci 104(24):10240–10245
Hu L, Zhang Z, Hu Y (2012) A time-varying source connectivity approach to reveal human somatosensory information processing. Neuroimage 62(1):217–228
Hyafil A, Giraud AL, Fontolan L, Gutkin B (2015) Neural cross-frequency coupling: connecting architectures, mechanisms, and functions. Trends Neurosci 38(11):725–740
Hyvarinen A (1999) Fast and robust fixed-point algorithms for independent component analysis. IEEE Trans Neural Netw 10(3):626–634
Jann K, Kottlow M, Dierks T, Boesch C, Koenig T (2010) Topographic electrophysiological signatures of fmri resting state networks. PLoS ONE 5(9):e12–945
Jansen BH, Rit VG (1995) Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. Biol Cybern 73(4):357–366
Jones SR, Pritchett DL, Stufflebeam SM, Hämäläinen M, Moore CI (2007) Neural correlates of tactile detection: a combined magnetoencephalography and biophysically based computational modeling study. J Neurosci 27(40):10751–10764
Karameh FN, Awada M, Mourad F, Zahed K, Abou-Faycal IC, Nahas Z (2014) Modeling of neuronal population activation under electroconvulsive therapy. In: Biosignals, pp 229–238
Kiebel SJ, Garrido MI, Moran R, Chen CC, Friston KJ (2009) Dynamic causal modeling for EEG and MEG. Hum Brain Mapp 30(6):1866–1876
Lee U, Kim S, Noh GJ, Choi BM, Hwang E, Mashour GA (2009) The directionality and functional organization of frontoparietal connectivity during consciousness and anesthesia in humans. Conscious Cogn 18(4):1069–1078
Lee WH, Deng ZD, Kim TS, Laine AF, Lisanby SH, Peterchev AV (2012) Regional electric field induced by electroconvulsive therapy in a realistic finite element head model: influence of white matter anisotropic conductivity. Neuroimage 59(3):2110–2123
Lee U, Ku S, Noh G, Baek S, Choi B, Mashour GA (2013) Disruption of frontal-parietal communication by ketamine, propofol, and sevoflurane. J Am Soc Anesthesiol 118(6):1264–1275
Li B, Daunizeau J, Stephan KE, Penny W, Hu D, Friston K (2011) Generalised filtering and stochastic DCM for fMRI. Neuroimage 58(2):442–457
Lisanby SH (2007) Electroconvulsive therapy for depression. N Engl J Med 357(19):1939–1945
Ljung L (1999) System identification, Wiley Encyclopedia of Electrical and Electronics Engineering
Madi MK, Karameh FN (2017) Hybrid cubature kalman filtering for identifying nonlinear models from sampled recording: estimation of neuronal dynamics. PLoS ONE 12(7):1–49
Madi MK, Karameh FN (2018) Adaptive optimal input design and parametric estimation of nonlinear dynamical systems: application to neuronal modeling. J Neural Eng 15(4):046028
Mankad MV, Beyer JL, Weiner RD, Krystal A (2010) Clinical manual of electroconvulsive therapy. American Psychiatric Pub, Washington, DC
Markram H, Toledo-Rodriguez M, Wang Y, Gupta A, Silberberg G, Wu C (2004) Interneurons of the neocortical inhibitory system. Nat Rev Neurosci 5(10):793
McCormick DA, Bal T (1997) Sleep and arousal: thalamocortical mechanisms. Annu Rev Neurosci 20(1):185–215
Merkl A, Heuser I, Bajbouj M (2009) Antidepressant electroconvulsive therapy: mechanism of action, recent advances and limitations. Experim Neurol 219(1):20–26
Moran RJ, Kiebel SJ, Stephan K, Reilly R, Daunizeau J, Friston KJ (2007) A neural mass model of spectral responses in electrophysiology. NeuroImage 37(3):706–720
Moran RJ, Stephan KE, Seidenbecher T, Pape HC, Dolan RJ, Friston KJ (2009) Dynamic causal models of steady-state responses. Neuroimage 44(3):796–811
Moscrip TD, Terrace HS, Sackeim HA, Lisanby SH (2006) Randomized controlled trial of the cognitive side-effects of magnetic seizure therapy (MST) and electroconvulsive shock (ECS). Int J Neuropsychopharmacol 9(1):1–11
Mouraux A, Iannetti GD (2008) Across-trial averaging of event-related eeg responses and beyond. Magn Reson Imaging 26(7):1041–1054
Müller-Linow M, Hilgetag CC, Hütt MT (2008) Organization of excitable dynamics in hierarchical biological networks. PLoS Comput Biol 4(9):e1000–190
Nahas Z, Short B, Burns C, Archer M, Schmidt M, Prudic J, Nobler MS, Devanand D, Fitzsimons L, Lisanby SH et al (2013) A feasibility study of a new method for electrically producing seizures in man: focal electrically administered seizure therapy. Brain Stimul 6(3):403–408
Nobler MS, Sackeim HA, Prohovnik I, Moeller JR, Mukherjee S, Schnur DB, Prudic J, Devanand D (1994) Regional cerebral blood flow in mood disorders, III: treatment and clinical response. Arch General Psychiatry 51(11):884–897
Nobler MS, Oquendo MA, Kegeles LS, Malone KM, Campbell C, Sackeim HA, Mann JJ (2001) Decreased regional brain metabolism after ECT. Am J Psychiatry 158(2):305–308
Pagerols J, Rojo J (2009) Electrophysiological mechanisms of action of electroconvulsive therapy. Actas Esp Psiquiatr 37(6):343–351
Palmer L, Murayama M, Larkum M (2012) Inhibitory regulation of dendritic activity in vivo. Front Neural Circuits 6:26
Palva S, Palva JM (2012) Discovering oscillatory interaction networks with m/EEG: challenges and breakthroughs. Trends Cogn Sci 16(4):219–230
Park HJ, Friston K (2013) Structural and functional brain networks: from connections to cognition. Science 342(6158):1238–411
Park HJ, Friston K, Pae C, Park B, Razi A (2017) Dynamic effective connectivity in resting state fMRI. NeuroImage. https://doi.org/10.1016/j.neuroimage.2017.11.033

Pereda E, Quiroga RQ, Bhattacharya J (2005) Nonlinear multivariate analysis of neurophysiological signals. Prog Neurobiol 77(1):1–37

Pfeffer CK, Xue M, He M, Huang ZJ, Scanziani M (2013) Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons. Nat Neurosci 16(8):1068–1076

Pinotsis D, Geerts J, Pinto L, FitzGerald T, Litvak V, Auksztulewicz R, Friston K (2017) Linking canonical microcircuits and neuronal activity: dynamic causal modelling of laminar recordings. NeuroImage 146:355–366

Plomp G, Quairiaux C, Kiss JZ, Astolfi L, Michel CM (2014) Dynamic connectivity among cortical layers in local and large-scale sensory processing. Eur J Neurosci 40(8):3215–3223

Proix T, Spiegler A, Schirner M, Rothmeier S, Ritter P, Jirsa VK (2016) How do parcellation size and short-range connectivity affect dynamics in large-scale brain network models? NeuroImage 142:135–149

Rennie CJ, Robinson PA, Wright JJ (2002) Unified neurophysical model of EEG spectra and evoked potentials. Biol Cybern 86(6):457–471

Roebroeck A, Formisano E, Goebel R (2011) The identification of interacting networks in the brain using fmri: model selection, causality and deconvolution. Neuroimage 58(2):296–302

Rosch R, Friston K, Tisdall M, Thornton R (2017) Patient-specific modelling of epileptogenic networks from stereotactic EEG recordings. https://doi.org/10.6084/m9.figshare.5311198.v1

Rubinov M, Sporns O (2010) Complex network measures of brain connectivity: uses and interpretations. Neuroimage 52(3):1059–1069

Sackeim HA (1999) The anticonvulsant hypothesis of the mechanisms of action of ECT: current status. J ECT 15(1):5–26

Sackeim HA, Luber B, Katzman GP, Moeller JR, Prudic J, Devanand D, Nobler MS (1996) The effects of electroconvulsive therapy on quantitative electroencephalograms: relationship to clinical outcome. Arch General Psychiatry 53(9):814–824

Sackeim HA, Prudic J, Nobler MS, Fitzsimons L, Lisanby SH, Payne N, Berman RM, Brakemeier EL, Perera T, Devanand D (2008) Effects of pulse width and electrode placement on the efficacy and cognitive effects of electroconvulsive therapy. Brain Stimul 1(2):71–83

Sedigh-Sarvestani M, Schiff SJ, Gluckman BJ (2012) Reconstructing mammalian sleep dynamics with data assimilation. PLoS Comput Biol 8(11):e1002–788

Sellers KK, Bennett DV, Hutt A, Williams JH, Fröhlich F (2015) Awake vs. anesthetized: layer-specific sensory processing in visual cortex and functional connectivity between cortical areas. J Neurophysiol 113(10):3798–3815

Sengupta B, Friston KJ, Penny WD (2015) Gradient-free mcmc methods for dynamic causal modelling. NeuroImage 112:375–381

Shayegh F, Fattahi RA, Sadri S, Ansari-Asl K (2011) A brief survey of computational models of normal and epileptic eeg signals: a guideline to model-based seizure prediction. J Med Signals Sens 1(1):62

Spellman T, Peterchev AV, Lisanby SH (2009) Focal electrically administered seizure therapy (feast): a novel form of ect illustrates the roles of current directionality, polarity, and electrode configuration in seizure induction. Neuropsychopharmacology 34(8):2002

Sporns O (2014) Contributions and challenges for network models in cognitive neuroscience. Nat Neurosci 17(5):652–660

Sporns O, Betzel RF (2016) Modular brain networks. Annu Rev Psychol 67:613–640

Staiger JF, Freund TF, Zilles K (1997) Interneurons immunoreactive for vasoactive intestinal polypeptide (vip) are extensively innervated by parvalbumin-containing boutons in rat primary somatosensory cortex. Eur J Neurosci 9(11):2259–2268

Stephan KE, Hilgetag CC, Burns GA, O'Neill MA, Young MP, Kotter R (2000) Computational analysis of functional connectivity between areas of primate cerebral cortex. Philos Trans R Soc London B 355(1393):111–126

Tamás G, Szabadics J, Lörincz A, Somogyi P (2004) Input and frequency-specific entrainment of postsynaptic firing by ipsps of perisomatic or dendritic origin. Eur J Neurosci 20(10):2681–2690

Thompson WH, Fransson P (2015) The frequency dimension of fmri dynamic connectivity: network connectivity, functional hubs and integration in the resting brain. NeuroImage 121:227–242

Trevelyan AJ, Schevon CA (2013) How inhibition influences seizure propagation. Neuropharmacology 69:45–54

Uhrig L, Dehaene S, Jarraya B (2014) Cerebral mechanisms of general anesthesia. Annales francaises d'anesthesie et de reanimation 33:72–82

Van Den Heuvel MP, Sporns O (2011) Rich-club organization of the human connectome. J Neurosci 31(44):15775–15786

Van den Hof PM, Dankers A, Heuberger PS, Bombois X (2013) Identification of dynamic models in complex networks with prediction error methods-basic methods for consistent module estimates. Automatica 49(10):2994–3006

van Rotterdam A, Da Silva FL, Van den Ende J, Viergever M, Hermans A (1982) A model of the spatial-temporal characteristics of the alpha rhythm. Bull Math Biol 44(2):283–305

Varela F, Lachaux JP, Rodriguez E, Martinerie J (2001) The brainweb: phase synchronization and large-scale integration. Nat Rev Neurosci 2(4):229

Von Stein A, Sarnthein J (2000) Different frequencies for different scales of cortical integration: from local gamma to long range alpha/theta synchronization. Int J Psychophysiol 38(3):301–313

Weaver KE, Wander JD, Ko AL, Casimo K, Grabowski TJ, Ojemann JG, Darvas F (2016) Directional patterns of cross frequency phase and amplitude coupling within the resting state mimic patterns of fmri functional connectivity. Neuroimage 128:238–251

Wendling F, Bartolomei F, Bellanger J, Chauvel P (2002) Epileptic fast activity can be explained by a model of impaired gabaergic dendritic inhibition. Eur J Neurosci 15(9):1499–1508

Wendling F, Hernandez A, Bellanger JJ, Chauvel P, Bartolomei F (2005) Interictal to ictal transition in human temporal lobe epilepsy: insights from a computational model of intracerebral EEG. J Clin Neurophysiol 22(5):343

Xiang W, Yang C, Karfoul A, Jeannès RLB (2016) Quantifying connectivity in a physiology based model using adaptive dynamic causal modelling. In: Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the, IEEE, pp 2818–2821

Yamamura D, Sano A, Tateno T (2017) An analysis of current source density profiles activated by local stimulation in the mouse auditory cortex in vitro. Brain Res 1659:96–112