Dynamic causal modelling of neurological pathology: Using neural mass models to understand dynamic dysfunction in NMDA-receptor antibody encephalitis

Richard E Rosch[1, 2], Gerald Cooray[1, 3], Karl J Friston[1]

[1]Wellcome Trust Centre for Neuroimaging, Institute of Neurology. University College London, UK. [2]Centre for Developmental Cognitive Neuroscience, Institute of Child Health, University College London, UK. [3]Department of Clinical Neurophysiology, Karolinska University Hospital, Stockholm, Sweden

Introduction

The invention of electroencephalography (EEG) in the early 20th century facilitated a revolutionary insight into the dynamic behaviour of the human brain. For the first time, clinicians and researchers were able to examine direct evidence of brain function, in both normal human participants and patients with neurological conditions (Jung & Berger 1979). Clinically this new way of assessing brain function has had significant impact on our understanding of a number of neurological and psychiatric conditions, but none more so than epileptic seizure disorders.

Epilepsy is the label given to a number of heterogeneous conditions characterised by an enduring risk of epileptic seizures – because of their heterogeneity, these conditions are now often referred to as 'the epilepsies' (Panayiotopoulos 2005). The term epileptic seizure describes the transient occurrence of signs and symptoms caused by abnormally excessive or synchronous activity in the brain (Fisher et al. 2014). Whilst the concept that abnormal electrical activity causes epileptic seizures predates the invention of EEG, much of our current physiological understanding and clinical decision making is based on EEG recordings from patients suffering from epilepsy (Eadie & Bladin 2001).

Descriptions and analyses of EEG recordings have remained virtually unchanged since its conception. Its clinical use largely rests on the description of visually recognisable features and their phenomenological categorisation, with the exception of some recently adopted advanced source localisation algorithms (Zschocke & Hansen 2012). But relying just on these visually apparent pathological patterns do not capture the entire breadth of information that is available in an EEG recording.

One of the main advantages of EEG (which it shares with magnetoencephalography, MEG) over other methods assessing brain function is its temporal resolution, which still remains unparalleled when it comes to investigating the human brain *in vivo*. This results in rich datasets, which capture interacting fluctuations of electric activity across frequencies that may be two or three orders of magnitude apart. Whilst surface EEG recorded non-invasively from the scalp has a limited spatial resolution, it does allow for the simultaneous recording of neuronal activity across almost the entirety of the cortical surface. Furthermore, the spatial resolution limitations have been addressed by developing invasive EEG recording devices that can be implanted neurosurgically where a better understanding of the spatial origins of an EEG signal are deemed clinically necessary.

Much of the information contained within these datasets is not accessible through visual inspection alone, but rather needs to be elicited utilising more quantitative analysis methods (Tong & Thakor 2009). Applying such quantitative analysis methods has led to the description of a wide variety of novel electrophysiological findings. For example, analysing the correlation between the time series of individual EEG channels will yield a matrix of channel-to-channel correlation measures. These can be read as indicators of *functional connectivity*, with the results interpreted in a graph theory framework for *functional network* analysis (Bullmore & Sporns 2009).

Another example can be found in the recent emergence of cross-frequency coupling as a potentially important mechanism for neuronal computation: Quantitative analysis of power at different EEG frequencies in humans has shown that amplitude fluctuations are measurably modulated and time locked to the phase of concurrent slower frequency oscillation, known as phase-amplitude cross-frequency coupling (Canolty & Knight 2010).

At the same time as advances in the detailed quantitative analysis of macroscopic EEG signals in health and disease, there has been an exponential increase in our understanding of the molecular, and to some extend cellular basis of many of the epilepsies (Helbig et al. 2008; Thomas & Berkovic 2014). Increasingly, knowledge of associated molecular abnormalities, such as the presence of relevant gene mutations or specific autoantibodies against synaptic targets, influences prognosis, clinical management and specific treatment decisions for patients with epilepsy.

Features of these disease-associated molecular abnormalities have also lead to a putative understanding of the pathophysiological mechanisms underlying different epileptic seizure disorders. For example, the frequency of mutations in ion channel genes has led to the concept of epilepsies and other paroxysmal neurological disorders being *channelopathies*, i.e. disorders in neuronal ion channel function (Spillane et al. 2015), which can be further investigated in appropriate animal model systems.

However, with the increase in knowledge, new challenges arise. The availability and increased clinical use of testing for specific mutation and autoantibodies has quickly led to the realisation that even apparently specific molecular abnormalities are associated with a wide variety of disease pictures in human patients. The same mutation in the voltage gated sodium channel gene *SCN1A* for example, can cause a diverse selection of phenotypes within the same family: ranging from comparatively mild phenotypes consisting of childhood febrile seizures, to a severe epileptic disorder characterised by difficult to control frequent daily seizures associated with global developmental delay (Miller & Sotero de Menezes 2014). Similarly, mutations in the *GRIN2A* gene, coding for a subunit of the N-methyl-D-aspartate receptor (NMDAR), can cause a range of electroclinical syndromes, even within the same family (Lesca et al. 2013).

This opens an explanatory gap: On one side there is an increased understanding of the putative molecular and cellular causes of dynamic disorders of the brain; on the other there are macroscale measures of abnormal brain function that, whilst loosely associated with some microscale abnormalities, do not allow a direct one-to-one mapping. Bridging this gap is likely to require an intermediate step – a conceptual bridge that can link information about molecular dysfunction with its expression in neuronal function at an intermediate level, the *mesoscale*, in order to understand the emergence of phenotypic variability observed in human patients.

In other fields within neuroscience, this approach is emerging as a necessary step for linking observations at the microscale (e.g. cellular neuronal circuits) with the observations at the macroscale (e.g. organism behaviour). Whilst descriptions of the cellular circuitry may include too many particular details to understand their

functional contributions to the overall behaviour, evaluating only a behaviour in the whole organism may not sufficiently represent the complexity of the underlying neuronal processes. Bridging this gap requires the identification recurrent themes at an intermediate functional level, such as neuronal computations, which may be implemented through differing neuronal circuits, but produce similar effects (Carandini 2012).

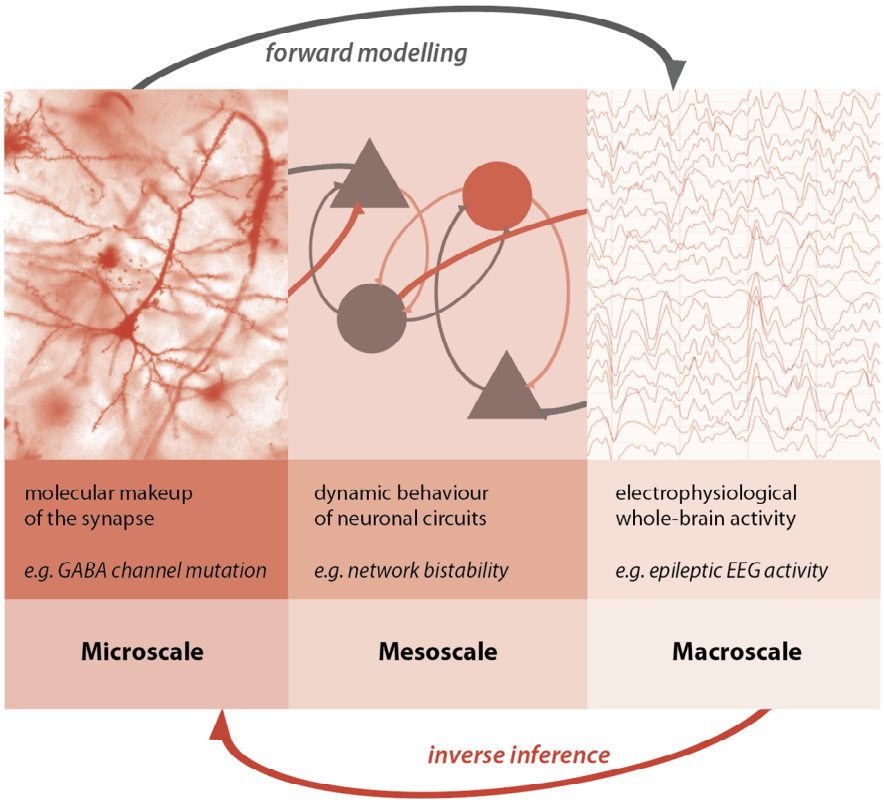

Fig. 1. Understanding epileptic dynamics at different scales: Different lines of evidence lead to descriptions of pathology on different scales. Clinical syndromes often rely on the description of recognisable phenotype at the macroscale. Recent advances in understanding associated molecular abnormalities have improved our pathophysiological understanding of many diverse epilepsies, but robustly linking clinical phenotypes with microscale abnormalities has proven difficult. Including an intermediate consideration of network dynamics may aid both prediction, and allow for addressing the inverse problem of inferring pathophysiology from whole-brain electrophysiological measurements.

In the context of clinical neurology, a similar approach would suggest that in order to link putative molecular causes (microscale) with diverse disease phenotypes observed in patients (macroscale), we need to consider the intermediate step: Dysfunction at the level of neuronal computations (mesoscale, Figure 1). Linking molecular abnormalities to *in vivo* neuronal dysfunction is now a standard component of identifying disease mechanisms in emerging genetics and autoimmune conditions (Pal & Helbig 2015). Attempts in relating macroscale findings to models of mesoscale neuronal dynamics have been less forthcoming, and will be the focus here.

This chapter will discuss the use of computational models of neuronal circuit function to link abnormalities in clinical neurophysiological measurements to pathophysiological mechanisms. An introduction of population models of neuronal function will be followed by an in-depth discussion of how such models can be used to make inference on the mechanism underlying specific electrophysiological changes. This approach will then be illustrated using an exemplar clinical case with a known underlying molecular diagnosis of NMDAR antibody encephalitis.

The Scientific Problem

The functional validation of possible molecular causes of neurological diseases is an essential step in any description of new pathophysiological mechanisms. In order to increase confidence that a genetic mutation that is epidemiologically linked with a specific phenotype could actually play a causative role in the disease, some evidence that this can have an effect on neuronal function is considered a standard requirement (Quintáns et al. 2014).

This approach usually relies on replicating the molecular abnormality in a model organism or system and evaluating the model for any resultant deficits, particularly in regards to neuronal function. For example, to provide evidence for the direct pathogenicity of NMDAR antibodies in the recently described NMDAR-antibody associated encephalitis (Dalmau et al. 2008), the antibody-rich patient cerebrospinal fluid was applied to murine hippocampal slices prepared for voltage-clamp recordings in order to measure the effects of antibody exposure on glutamate transmission. This provided evidence both for acute antagonism of the NMDAR by the antibody, as well as chronic reduction of NMDAR associated with antibody exposure (Hughes et al. 2010).

Whilst this approach is powerful, and necessary in order to evaluate candidate molecular causes of neurological disorders in the context of neuronal function, several problems remain unsolved when relying on this approach alone:

- **Animal models for human disease:** Model systems used to assess the pathological effects of molecular abnormalities are usually non-human organisms or

tissues, leaving uncertainty as to whether similar effects would be evident in the human brain.

- **Emergent properties at different scales:** There may be a gap between individual cell and small circuit abnormalities assessable in an experimental model system, and the inference drawn on larger networks and systems. These models are particularly prone to neglecting emergent properties at different scales (e.g. bistability of a network), and the effects of unknown modulators in the whole system that may enhance or suppress the observed microscale abnormality
- **Human phenotypic variability:** An unexpected result of the recent increase in molecular diagnoses in neurology is the discovery of large phenotypic variability even where a molecular cause has been identified and well characterized (Hildebrand et al. 2013). Functional investigations in homogeneous model systems do not address the mechanisms underlying phenotypic diversity.

Issues of disease pathology in humans, understanding whole-organism, and delineating relevant categories within phenotypically diverse groups are essential for translating basic neuroscientific findings into clinically relevant advances. In order to start addressing these issues, the inverse problem has to be addressed: How do macroscale abnormalities relate to underlying pathophysiology?

The EEG signal, despite containing a lot of rich information, is a poor measure of neuronal function at the cellular, or synaptic level: because of the spatial inaccuracies and the summation of many million individual neurons' activity into a composite signal, there are an infinite number of possible neuronal constellations that could cause the same measureable EEG signatures. Attempting to relate this composite, diffuse signal to underlying neuronal dysfunction is thus an ill-posed problem, where no unique solution exists.

Ill-posed problems are common in neuroscience, both in terms of problems researchers encounter when investigating nervous systems (e.g. the source localization problem for EEG signals, (Friston et al. 2008), and problems that nervous systems themselves have to address (e.g. the feature binding problem, (Di Lollo 2012)). These problems are not impossible to solve, as is evident in the successful application of source-reconstruction algorithms in identifying epileptogenic brain areas for surgery (Lantz et al. 2011), and the brain's successful and reliable decoding of visual information (Kawato et al. 1993).

With underdetermined ill-posed problems, providing constraints to the possible solutions is crucial (Friston 2005). Constraining the problem reduces the space of possible solutions and makes the problem more tractable. These constraints also help in keeping inverse solutions more interpretable and relevant to the scientific question at hand.

One way to constrain such inverse problems in neuroscience is the use of computational models of neuronal populations as a mesoscale representation of neuronal dynamics. This casts the inverse problem of attempting to infer microscale causes

of macroscopic phenomena into a more restricted problem: Assuming basic mechanisms of neuronal function and organization are met (i.e. the neuronal model applies), which setup of the known circuitry could produce an observed effect (i.e. which specific set of model parameters can produce the observed macroscale response)?

In the following, we will use advanced computational methods to address this problem, inferring underlying neuronal circuitry abnormalities from clinical EEG recordings of a patient with a known molecular abnormality: Specifically we will use mesoscale neuronal modelling to assess what synaptic abnormalities underlie paroxysmal EEG abnormalities in a paediatric patient with NMDAR antibody encephalitis.

NMDAR antibody encephalitis is a recently described autoimmune encephalitis (Dalmau et al. 2008), i.e. an inflammatory condition of the brain associated with, and most likely caused by, autoantibodies against molecular targets relevant for synaptic function. EEG abnormalities are commonly associated with the condition and have some diagnostic value (Gitiaux et al. 2013), but are very varied between patients (Florance et al. 2009) and evolve over time (Nosadini et al. 2015). In paediatric patients the common abnormalities described include non-specific sharp wave paroxysms, longer runs of rhythmic activity and more clearly epileptic spike and wave discharges (see Figure 2 for examples from our own clinical cohort).

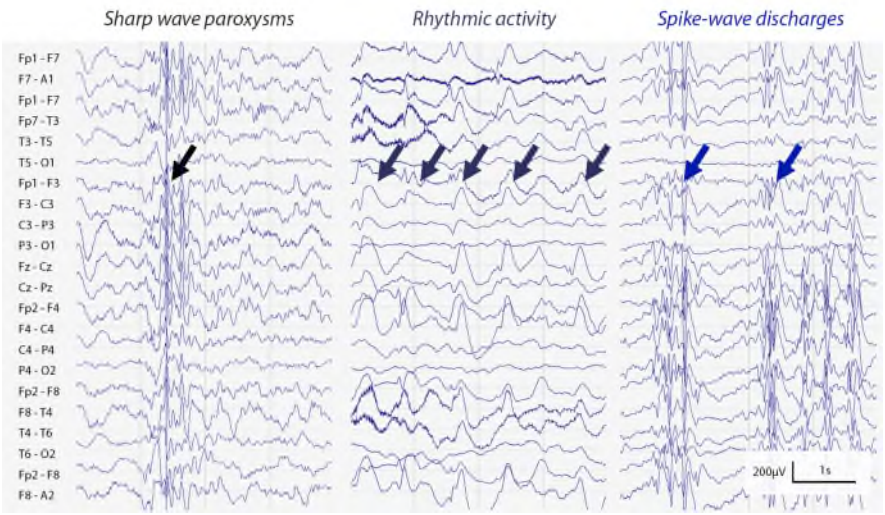

Fig. 2. EEG abnormalities observed in paediatric patients with NMDAR antibodies: The figure collates three different EEG findings from separate paediatric NMDAR antibody encephalitis patients. Abnormalities range from non-specific sharp wave paroxysms (left panel), to rhythmic activity with or without impairment of consciousness (middle panel) to clearly epileptic spike-wave activity (right panel).

In the following sections we will discuss mesoscale models of neuronal function and how they can be used to explain observed EEG phenomena by constraining the inverse problem. We will then highlight a specific computational approach – *dynamic causal modelling* (DCM) – and apply the method to EEG abnormalities observed in one NMDAR antibody encephalitis patient. We will then discuss our results in terms of their relation to other findings regarding the pathophysiology in NMDAR, as well as further implications for computational methods in the age of exponential discovery of candidate molecular mechanisms in neurology.

Computational Methods

In this section we will discuss how computational models of neuronal function can be used to make inference on causative mechanisms underlying phenomena observed in human EEG recordings. This will be done in three parts – the first will give a short overview of generative models of neuronal function and different approaches to linking them to empirical data. In the second part we will introduce the approach applied to our empirical case – namely *dynamic causal modelling* – in a little more detail. And in the concluding part of this section we will illustrate how dynamic causal modelling can be used to fit a generative neuronal model to empirical EEG data and make inference on underlying mechanisms using an illustrative case of NMDAR encephalitis.

Generative Models of Neuronal Population Activity

Neuronal systems are highly nonlinear coupled systems (Werner 2007). This means that predicting input-output relationships is challenging and often counterintuitive. One of the great strengths of computational models is that they can be used to explore input-output relationships systematically and help in identifying some of the unexpected effects produced by nonlinear interactions.

The pioneering work by Hodgkin and Huxley (Hodgkin & Huxley 1952) produced one of the first such computational models, and by some measures the most successful one developed so far. Using empirical voltage clamp measurements from the giant squid axon, they elegantly developed a model of neuronal membrane dynamics based entirely on voltage dependent ion channels, that could predict many patterns of neuronal behaviour observed empirically.

Since then, models are being developed on a multitude of different neuronal scales, ranging from subcellular compartment dynamic models, to models describing the output of whole neuronal populations. Because of the spatial scales of measurement, those models that represent whole neuronal populations are particularly informative when relating them to EEG measurements.

One of the earliest such models was the Wilson and Cowan neural mass model (Wilson & Cowan 1972) – they describe the behaviour of a whole set of interconnected neurons not individually but as whole aggregates, based on similar approaches in particle physics. They also provide a justification for this method based, interestingly, not just in its computational tractability, but rather the conceptually different inference this approach enables:

> "It is probably true that studies of primitive nervous systems should be focused on individual nerve cells and their precise, genetically determined interactions with other cells. […] [S]ince pattern recognition is in some sense a global process, it is unlikely that approaches which emphasize only local properties will provide much insight. Finally it is at least a reasonable hypothesis that local interactions between nerve cells are largely random, but that this randomness gives rise to quite precise long-range interactions." (Wilson & Cowan 1972)

Using this approach they arrive at a system of two ordinary differential equations describing a neuronal oscillator consisting of two different populations, one excitatory, one inhibitory:

$$\tau_e \frac{\mathrm{d}E}{\mathrm{d}t} = -E + (k_e - r_e E)\, S_e(c_1 E - c_2 I + P) \qquad (1.1)$$

$$\tau_i \frac{\mathrm{d}I}{\mathrm{d}t} = -I + (k_i - r_i I)\, S_i(c_3 E - c_4 I + Q) \qquad (1.2)$$

This system describes two neuronal populations, whose current states (*E*, *I*: proportion of cells in the population firing) influence each other through weighted connections (c_{1-4}, weights of population connections, see Figure 3A). This coupling is mediated through a sigmoid activation function ($S_{e/i}$), which acts like a switch integrating all incoming synaptic influences and translating them into a postsynaptic state change within a defined dynamic range (i.e. $0 - 1$). The sigmoid functions are population specific (and can therefore be parameterised independently) and are the source of non-linearity in the model.

Even despite the extreme simplification of these models of neuronal function, a whole range of dynamic behaviours can be reproduced with WC-type models at the scale of neuronal populations or cortical patches (Meijer et al. 2015; Wang et al. 2014; Heitmann et al. 2012). Because of the coupled nonlinearities, however, 'forward' predictions of model behaviour given specific parameters is non-intuitive and usually requires simulation of the model response. Recurrent simulations for varying parameter values can then be used to establish a link between model parameterisation and overall dynamic response (Figure 3B).

These parameter/response relationship can be exploited to make inference on model parameters underlying a given observations. Faced for example, with noisy measurements of a population oscillation (Figure 3C), one can use systematic variations of a model parameter to identify the specific parameter value that best fits the data (Figure 3D, illustrated for the stimulating current parameter *P*). However, even adding a single additional free parameter (e.g. the connection parameter c_1) creates

a complex model prediction error landscape that is much more difficult to optimise (Figure 3E) – an important problem that we will return to later.

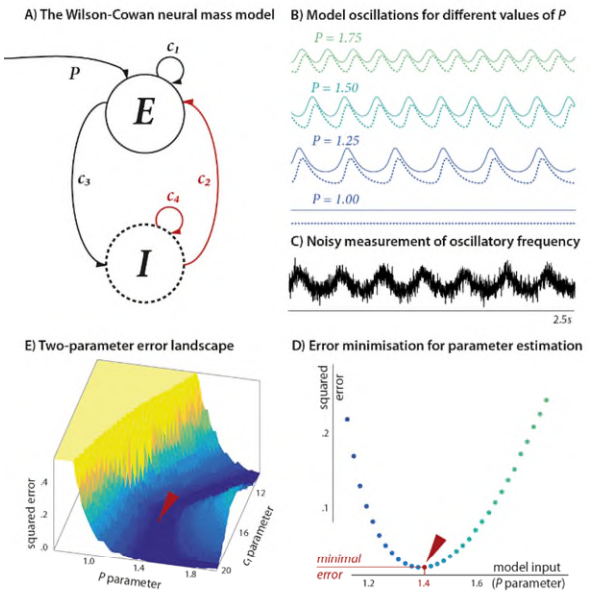

Fig. 3. The Wilson-Cowan neural mass model: A) The model consists of one inhibitory and one excitatory neuronal population, coupled through synaptic connections of a specific connection strength (c_i parameters). These can be excitatory (black) or inhibitory (red). The system receives external stimulating current input (*P* parameter) and acts as a neuronal oscillator. **B)** The model generates particular oscillation patterns for different parameter constellations – this figure illustrates steady state oscillatory responses with decreasing values for the input parameter *P*. Excitatory populations are represented by the solid lines, inhibitory populations by dashed lines. **C) S**ynthetic data illustrating a noisy measurement of neuronal population oscillation driven with *P=1.4*. We show how *oscillatory frequency* alone can be used to derive the *P*-parameter from noisy measurements such as this: **D)** Estimates of steady state oscillatory frequency can be derived from the model for a range of different values for *P*. Plotting the squared difference between estimated frequencies and that derived from the noisy synthetic signal, we can identify the *P* value that produces the minimal error. This approach identifies *P=1.4* as the value producing the minimal error (indicated by the red arrow). **E)** If more than one parameter is allowed to vary (e.g. input *P*, and self-excitation strength c_1) the error landscape becomes more complex and error minimisation alone does not produce unambiguous results – the red arrow indicates the same parameter constellation identified in 3D).
The model specifications were taken directly from the model's original description (Wilson & Cowan 1972). Parameters for the modelling were taken from one of the known oscillatory states and unless otherwise stated were: $c_1=16$, $c_2=12$, $c_3=15$, $c_4=3$, $P=1.25$.

WC models allow for generation of complex dynamics that remain computationally tractable enough to explore different parameter compositions and attempt inference on parameter combinations producing a certain dynamic response. However, in the original formulation consisting of a single excitatory and inhibitory population, they are limited in how well they can represent the range and complexity of cortical dynamics observed in the laminate cortex.

A major extension of the WC model was introduced by Jansen and Rit in 1995 (Jansen & Rit 1995) building on an extant literature of adaptations of the WC-type models (Lopes da Silva et al. 1974). The Jansen and Rit (JR) model explicitly models dynamics of a local cortical circuit by ascribing different neuronal populations to specific cortical lamina and describing their dynamics in terms of differential equations. In this model, an additional excitatory neuronal population is added allowing separate parameterisation for two excitatory neuronal populations.

$$\tau_e \frac{\mathrm{d}x_1}{\mathrm{d}t} = H_e\big(P + S_1(v_2)\big) - 2x_1 - \frac{v_1}{\tau_e} \tag{2.1}$$

$$\frac{\mathrm{d}v_1}{\mathrm{d}t} = x_1 \tag{2.2}$$

$$\tau_i \frac{\mathrm{d}x_2}{\mathrm{d}t} = H_i S_2(v_3) - 2x_2 - \frac{v_2}{\tau_i} \tag{2.3}$$

$$\frac{\mathrm{d}v_2}{\mathrm{d}t} = x_2 \tag{2.4}$$

$$\tau_e \frac{\mathrm{d}x_3}{\mathrm{d}t} = H_e S_3(v_1 - v_2) - 2x_3 - \frac{v_3}{\tau_e} \tag{2.5}$$

$$\frac{\mathrm{d}v_3}{\mathrm{d}t} = x_3 \tag{2.6}$$

This constellation of neuronal populations allows a diverse spectrum of frequency mixtures to be modelled, and is capable of producing a host of response dynamics also observed in empirical measurements of cortical potential fluctuations (Jansen & Rit 1995; Aburn et al. 2012; Goodfellow et al. 2012). An additional benefit that emerges from the laminar specificity of the JR model, is that it relates naturally to commonly available brain recordings in humans – specifically MEG and EEG. The electromagnetic activity measurable at the scalp is thought to mainly reflect postsynaptic currents in the apical dendrites of populations of pyramidal cells (Lopes da Silva 2010). These are explicitly modelled in the JR model, so that their selective contribution to EEG/MEG measurements can be distinguished from the activity of other cell populations.

The laminar specificity and the wide range of physiological frequencies that can be modelled mean that JR-type models are commonly employed in computational models of cortical function. They currently form one of model-based approaches for the analysis of large scale brain dynamics, such as the dynamic causal modelling (DCM) framework, which will be discussed in more detail later (David et al. 2006).

Because they aim to represent biophysical connectivity patterns found in actual cortical microcircuits, their architecture is also congruent with computational motifs thought to be the basis of cortical processing (e.g. predictive coding, (Bastos et al. 2012)).

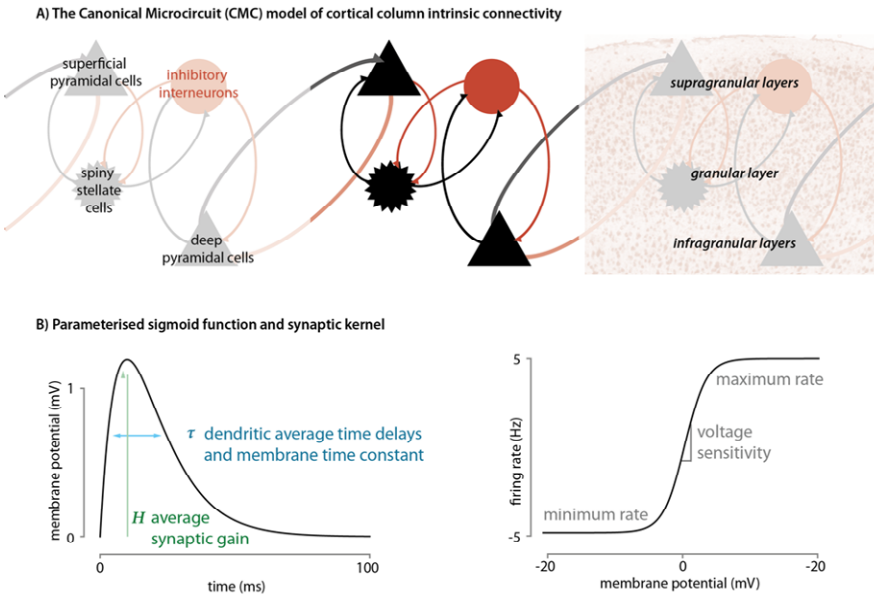

Fig. 4. The Canonical Microcircuit (CMC) Model : A) This extension of the Jansen-Rit model consists of four neuronal populations (left panel) mapping onto different cortical laminae (right panel). The middle panel shows the intrinsic excitatory (black) and inhibitory (red) connections contained in the model (for simplicity, recurrent self-inhibition present for each population is not shown). **B)** Two operators define the evolution of population dynamics: First a synaptic kernel performs a linear transformation of presynaptic input into an average postsynaptic potential, dispersed over time (left panel). This is parameterised by synaptic gain parameters and averaged time constants. Second there is a nonlinear transformation of average membrane potential into population firing rates, described as a parameterised, population-specific sigmoid function (right panel).

Within the DCM framework, several extensions of existing neuronal models have been developed to address specific hypotheses regarding neuronal function (Moran et al. 2013). One of the extensions to the classical JR model employed in DCM is the so-called ‘canonical microcircuit’ or CMC (Figure 4). Here the single JR pyramidal cell population is separated into distinct ‘superficial’ and ‘deep’ pyramidal cells. This allows not only for afferent and efferent projections to be separated into distinct cortical laminae, but also accommodates differences in spectral output among different layers of the same cortical column - that are seen empirically in invasive measurements (Buffalo et al. 2011).

The model consists of a simple extension of the differential equations given in equations 2.1 – 2.6. Using the mean-field approximation the model can also be re-conceptualised in terms of average membrane potentials and firing rates interacting through specific kernels that summarise the activity-dependent integration of input at the postsynaptic membrane, and the nonlinear transformation of all input into an output firing rate (Figure 4B) (David & Friston 2003).

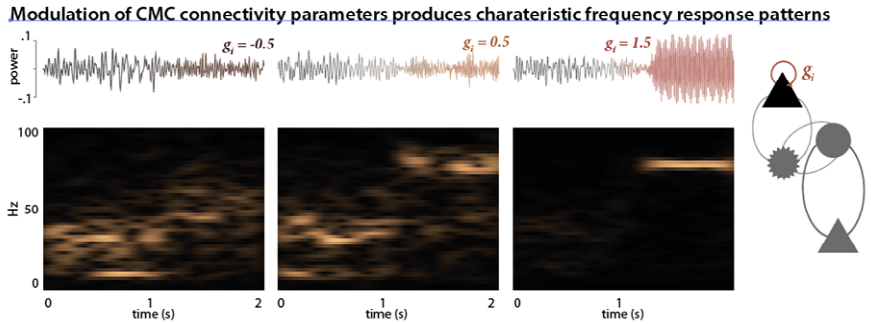

Fig. 5. Changes in intrinsic connectivity produce characteristic responses: Gradual changes to the recurrent self-inhibition gain parameter g_i are introduced to a CMC model of the cortical column at around 1s, changing the intrinsic modulation from 0 to -0.5, 0.5, and 1.5 respectively (left, middle, right panels). These changes produce characteristic signatures in the spectral output, apparent in both the time traces (top panels), and the spectral densities (bottom panels), with increases in self-inhibition leading to high power high frequency oscillations.

Because the architecture of the CMC model represents neuroanatomical features of the cortex, most of the modelling parameters are neurophysiologically meaningful and thus easily interpretable. The model parameters can be directly manipulated to reproduce many different dynamic behaviours – increasing the degree of self inhibition in superficial pyramidal cells for example will produce high frequency oscillations (Figure. 5, (Papadopoulou et al. 2015)).

Clearly the more intriguing question is whether inference on the model parameters can be made from empirical measurements, to identify which functional abnormality in the microcircuitry produced an abnormal measurement. This problem usually has more than one possible solution – meaning that many different possible constellations of parameters may cause identical appearing measurements, particularly where only some of the system's states are measureable, or observable (e.g. local field potentials), whilst many remain hidden (e.g. intracellular ion concentration fluctuations); the problem is ill-posed. This becomes particularly problematic where many different parameters can be used to explain a limited set of observed states. Even for the WC models with very few free parameters, simple optimisation routines as indicated in Figure 3D-E quickly become intractable, with complex multidimensional error landscapes that cannot be comprehensively mapped. The flexibility afforded by the increased number of free parameters in the CMC model comes at the cost of increased complexity of the space of possible solutions, making it difficult to evaluate which best explains observed behaviours.

There are several possible approaches to addressing this ill-posed problem, many of which have been employed in the computational modelling of epileptic seizures and EEG abnormalities. In the following, we will introduce a few of these approaches, with a focus on dynamic causal modelling. This will then be applied to address the question as to what abnormalities in the functional architecture can explain the paroxysms observed in patients with NMDAR antibody encephalitis.

Model Inversion for EEG Abnormalities

One of the most intuitive strategy to inversely link observed EEG features to changes in the parameters of an underlying generative model – i.e. to *invert* the model – is to systematically vary the parameters and evaluate how well the model simulations then fit the observed measurements. This can be done 'by hand', choosing individual parameter ranges and assessing the individual modelling outcomes (illustrated in Figure 3D-E for the simple case of estimating input currents producing a specific frequency output in the WC model). This approach can be informative, even in complex models of laminar cortical connectivity (Du et al. 2012), but is limited to a small numbers of varying parameters if comprehensive parameter mapping is attempted.

This limitation can be overcome by finessing (1) how the space of possible parameter values is explored to find a model that explains the data better (termed *optimization algorithms*), and (2) by utilising different measures to rank models against each other, i.e. to evaluate which model is the 'better' model (e.g. using a *cost function*). A large number of different approaches to both of these issues exist, of which a variety have been employed in inverting models of neuronal function to fit EEG data.

Optimisation algorithms

Optimisation algorithms describe computational strategies to identify parameter constellations within a range of possible values that produce a model output that best matches the observed results. There is a large literature regarding competing optimisation methods in a whole host of different areas of science and engineering, so in this section we will only discuss a few algorithms applied in fitting neuronal models to EEG data.

One of the most commonly applied algorithms is that of gradient descent (or ascent, depending on whether one is attempting to find minima or maxima). The basic idea is that from a random starting point, in order to find a minimum, one could iteratively take steps following the direction of the steepest downward gradient until no more changes are made at each step, i.e. the algorithm converges. Because the local gradient is defined by the first derivative of the cost function at any point, the cost function has to be locally differentiable for gradient ascent to be applicable. This approach is intuitive and easy to apply to a range of optimisation problems,

such as seizure classification (Thomas et al. 2008), or to refine aspects of EEG source reconstruction (Hansen & Hansen 2015).

There are a two major limitations to this approach, however, which apply to the problem at hand – namely inverting complex, multi-parameter neural mass model to fit EEG data:

(1) The gradient descent approach relies on the cost function to be smooth and continuous in order to be able to calculate the derivatives. Furthermore, in systems where there are unobserved variables in addition to unknown parameters that need to be inferred, calculating the derivatives directly is often not possible because of recursive dependencies between variables and parameters.
(2) It is designed to identify a local optimum, not the global optimum. Where the cost function is complex and has multiple local extrema, the local optimum identified in this approach may be far from the global optimum possible in the parameter ranges.

There are several alternative optimisation algorithms that address these problems. Genetic algorithms for example resemble the process of natural selection by producing random parameter variations and propagating the most 'successful' ones. After iteratively varying some of the parameters (introducing mutations) and then choosing the best variants (selection), the algorithm will converge to the best global solution, without requiring estimation of local gradients for its progression. This has been applied to fitting parameters of a detailed phenomenological model of individual EEG abnormalities in clinical EEG recordings, identifying patient-specific differences in the transition through parameter space (Nevado-Holgado et al. 2012). Similarly, algorithms such as particle swarm optimization, or simulated annealing use direct search strategies that do not rely on knowledge of the gradients. Thee algorithms converge to a global maximum without getting stuck in local optima. A variety of these have been used in model based analysis of EEG signals (Shirvany et al. 2012; Gollas & Tetzlaff 2005; van Dellen et al. 2012).

Therefore we have two broad classes of algorithms: (1) global direct search strategies, that yield robust convergence to global optima but come at a high computational cost, and (2) gradient descent algorithms that are more computationally efficient but may get stuck in local optima and not yield a global resolution. The balance of these competing limitations dictates which optimisation algorithm is most appropriate in a given situation.

When making inference on models with relatively few parameters, it is often possible to use one of the global algorithms for a model inversion, as the computational requirements for inverting a model of only a few parameters are usually manageable. However, in models, such as the CMC, where there are many free parameters that need to be fitted, the computational expense of these stochastic algorithms can be prohibitive and the more efficient gradient descent algorithms are called upon. In this setting, prior constraints are used to ensure model inversion is less

susceptible to arresting in local optima. Paradoxically, the local minima problem can also be finessed by have many free parameters (as 'escape routes' are more likely to be present where there are many different dimensions of parameter space).

The gradient descent approach can be further finessed to address some of the remaining problems: local linearization can be used to estimate gradient where the underlying cost function is expensive to calculate; expectation-maximisation (EM) algorithms can be employed to invert probabilistic models where not all variables are observed (Do & Batzoglou 2008), hierarchical model inversion can help to avoid local extrema (Friston et al. 2016). Each of these strategies is employed within the DCM framework, but crucially hinge on the cost-function employed – i.e. what is being optimised.

Cost Functions

In order to apply optimisation routines and improve how well a model represents data, we need to define which measure should be optimised. Often the most intuitive approach is to calculate the difference between the numerical predictions of the model states and the empirical measurements, and try and reduce the sum of squared errors between model prediction and empirical measurement. This approach has been successfully applied to EEG in a variety of ways (Sitnikova et al. 2008; Babajani-Feremi & Soltanian-Zadeh 2010).

If closeness of the model fit is the only criterion for the optimisation function, all free parameters within the models will be adjusted in order to produce the best model fit. Especially in models with many free parameters, this can lead to idiosyncratic results that resemble specific features of a given dataset, but show poor generalisability across different, similar datasets – a problem that has been termed *overfitting*. Several strategies can be employed to avoid overfitting and ensure generalisability of the modelling results.

One such approach has emerged naturally from reformulating the cost function not in terms of an absolute error that needs to be reduced, but rather in terms of the Bayesian model evidence (also known as the marginal likelihood) that needs to be maximised. The evidence is simply the probability of getting some data under a model of how those data were caused. This is generally evaluated by trying to estimate the underlying parameters of a model. In more detail: within the Bayesian framework, one estimates the probability of a given parameterisation ϑ, given a set of observations, or data *y*, by assuming that these were produced from a model *m* as follows:

$$p(\vartheta|y,m) = \frac{p(y|\vartheta,m)p(\vartheta,m)}{p(y|m)} \qquad (3.1)$$

This posterior probability is not easy to estimate directly, but various approaches can be used to approximate it. Variational Bayes is a generic approach to the analysis of posterior probability densities. In this approach, the *free energy* represents a

bound on the log of the model evidence and can therefore be used in optimisation routines to identify optima in the model evidence distribution (Friston et al. 2007). The (log-) evidence or marginal likelihood is defined as follows (where *D (||)* denotes the Kulback-Leibler, or *KL* divergence – a measure of the difference between two probability distributions; *y* denotes data; *m* denotes the model; ϑ denotes a set of model parameters; $q(\vartheta)$ denotes the variational density, i.e. the approximating posterior density which is optimised; thus $-\langle ln\ q(\vartheta)\rangle_q$ denotes the entropy and $\langle L(\vartheta)\rangle_q$ denotes the expected energy; *F* denotes the free energy):

$$\ln p(y|m) = F + D(q(\vartheta)||p(\vartheta|y,m)) \tag{4.1}$$

$$F = \langle \mathrm{L}(\vartheta)\rangle_q - \langle \ln\ \mathrm{q}(\vartheta)\rangle_q \tag{4.2}$$

The log evidence itself can be split into an accuracy and a complexity term, and thus automatically contains a penalty for overly complex models that are prone to overfitting. In the context of DCM the complexity of the model is established on the basis of how far parameters deviate from their prior values. Therefore, maximizing this Bayesian estimate of model evidence provides a compromise between goodness of model fit, and the generalizability of the model.

Specifically in regards to epilepsy there are further specific problems that need to be addressed: Often the changes of a parameter that varies with time are of interest (for example whilst trying to track network changes during the transition into a seizure). If no account were taken of the temporal contiguity between individual time steps, the already computationally expensive model inversion needs to be fully repeated at each time step, treating each window as independent sample.

For dynamic systems, where there is a temporal dissociation between fast varying states and more slowly changing underlying model parameters, this problem can be overcome through optimization approaches that take into account the temporal dependencies between parameter values at neighboring time points. One of the most successful of these approaches it the Kalman filter. This was originally developed for linear systems, but soon extended to nonlinear systems (Julier & Uhlmann 2004). The Kalman approach has been used very successfully to estimate parameters underlying transitions into seizure state, where it has proved to benefit from its ability to estimate unobserved (hidden) states (Freestone et al. 2014).

A similar (and mathematically equivalent) approach can be implemented within the DCM framework, where each time step receives the preceding model inversion posteriors as prior expectations, resulting in evidence accumulation (also known as Bayesian belief updating) across the whole modelling time (Cooray et al. 2016). More recently, a generic approach to estimating parameters at two modelling levels has allowed to accommodate arbitrary relationships between individual model inversion steps in a computationally efficient way (parametric empirical Bayesian approach, Friston, Zeidman, et al. 2015; Friston, Litvak, et al. 2015).

In summary, a Bayesian framework for the cost function allows incorporation of the required constraints to solve the inverse problem as *prior beliefs* regarding the parameters (consisting of expected value, and uncertainty measures). The use of priors furthermore allows the model evidence to be cast directly in terms of *accuracy – complexity*, and therefore preventing overfitting of excessively complex models. Furthermore, several computationally efficient techniques are available to accommodate modelling of time series data. We will now illustrate these procedures using a worked example.

Workflow for Analysis of NMDAR Antibody-Related Paroxysms

Patients with NMDAR antibody encephalitis show a whole variety of apparently different EEG paroxysms. The aim of the subsequent analysis is to identify possible causative mechanisms of how the molecular pathology is translated into an observable abnormal dynamic state. In order to address this aim, we call on the computational mechanisms introduced above.

Specifically, we utilise recent advances in parametric empirical Bayes within the DCM framework (Litvak et al. 2015; Friston et al. 2015; Friston et al. 2016), which allows for a two-stage modelling approach:

1. Fit parameters of canonical microcircuit neural mass model to both background and paroxysmal conditions separately, in order to find the parameter constellation that provides the best fit
2. Estimate the evidence for models of reduced complexity (i.e. fewer free parameters) to identify subset of parameters that explain most of the changes between background and paroxysms using Bayesian model comparison

The workflow for the analysis of an individual patient is illustrated in Figure 6. Note that in line with recent advances in dynamic causal modelling, first a 'full' model is inverted – i.e. all typically changing parameters are freed up and are allowed to change in order to explain observed data. Bayesian model comparison is conducted between models with reduced complexity, where the differences between conditions are explained by only a pre-defined subset of parameters. This second step allows for direct comparison of competing hypotheses (about which specific synaptic parameters mediate seizure onset) within the Bayesian framework.

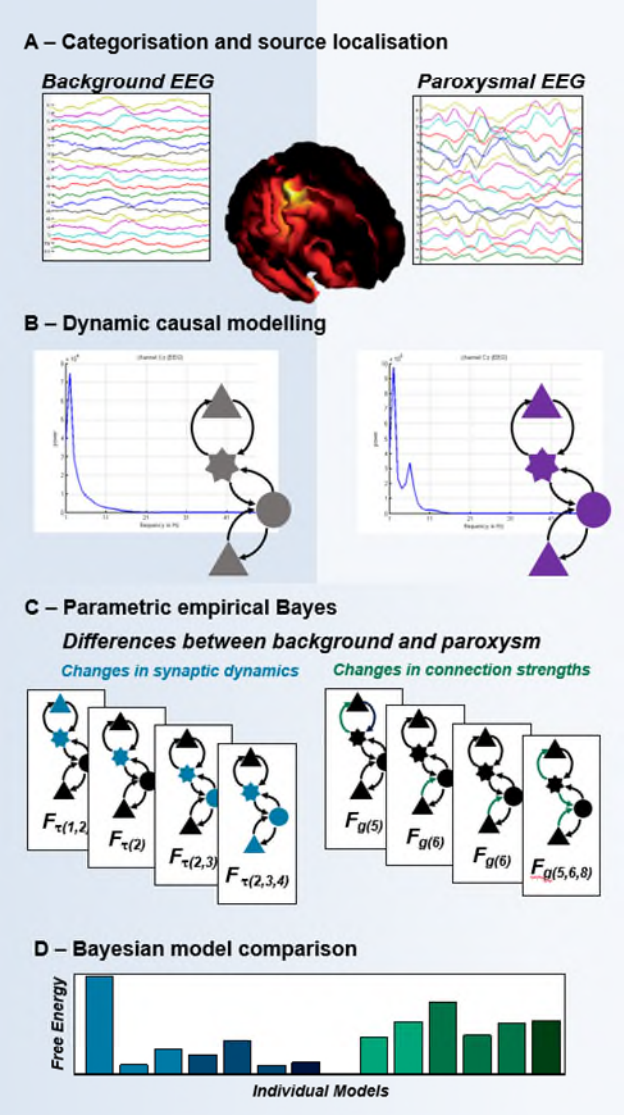

Fig. 6. Analysis workflow: A) Abnormal paroxysmal activity is visually identified on the EEG and source localised to a single source (Cooray et al. 2016) in order to extract a single 'virtual electrode' local field potential-like trace that contains the main spectral features of the activity. A matching number of time windows of background EEG activities is selected from artefact-free portions of the EEG. **B)** Dynamic causal modelling is used to respectively fit a single source canonical microcircuit to the paroxysmal, and the background activity. **C)** Parametric empirical Bayes is used estimate the free energy for models that explain both background, and paroxysmal activity with changes in only a subset of free parameters. **D)** Bayesian model comparison estimates the evidence for each of the reduced models from the previous step and is used to decide which model is most likely to have caused the observed data features.

The specific hypotheses tested in this analysis are founded in the existing knowledge of the molecular mechanism associated with abnormal EEG features: NMDA receptor antibodies affect glutamatergic, excitatory connections. Here we want to explore whether the resultant effects in the microcircuit mainly have an effect on the time constant of specific neuronal populations (parameterised as time constant, τ), or on the connection strength of excitatory connections (parameterised as g). The specific parameters of interest are summarised in table 1.

Table 1. Model parameters and combinations for reduced model comparison

Model parameters evaluated in reduced models			
τ_1	superficial pyramidal cells	g_1	connection from ss to sp
τ_2	spiny stellate cells	g_2	connection from dp to ii
τ_3	inhibitory interneurons	g_3	connection from ss to ii
τ_4	deep pyramidal cells		
Reduced models (i.e. combination of free parameters to explain both conditions)			
Model 1	τ_1	Model 16	g_1
Model 2	τ_2	Model 17	g_2
Model 3	τ_3	Model 18	g_3
Model 4	τ_4	Model 19	g_1, g_2
Model 5	τ_1, τ_2	Model 20	g_1, g_3
Model 6	τ_1, τ_3	Model 21	g_2, g_3
Model 7	τ_1, τ_4	Model 22	g_1, g_2, g_3
Model 8	τ_2, τ_3		
Model 9	τ_2, τ_4		
Model 10	τ_3, τ_4		
Model 11	τ_1, τ_2, τ_3		
Model 12	τ_1, τ_2, τ_4		
Model 13	τ_2, τ_3, τ_4		
Model 14	$\tau_1, \tau_2, \tau_3, \tau_4$		

Thus, the Bayesian model comparison in the second stage of the analysis will be used to decide whether changes in the *connection strength* of excitatory connections, or the *temporal integration dynamics* of individual neuronal subpopulations best explain the observed transitions from background EEG to paroxysmal abnormalities.

Results

Here we report a single case analysis of fitting a neuronal mass model of cortical dynamics to paroxysmal abnormalities in a patient with NMDA receptor encephalitis. Utilising the laminar and cell-type specificity of the canonical microcircuit and the computational efficiency of the variational Bayes approach to fitting the parameters, allows estimating the most likely constellations of biophysically relevant parameters that explain the observed EEG patterns.

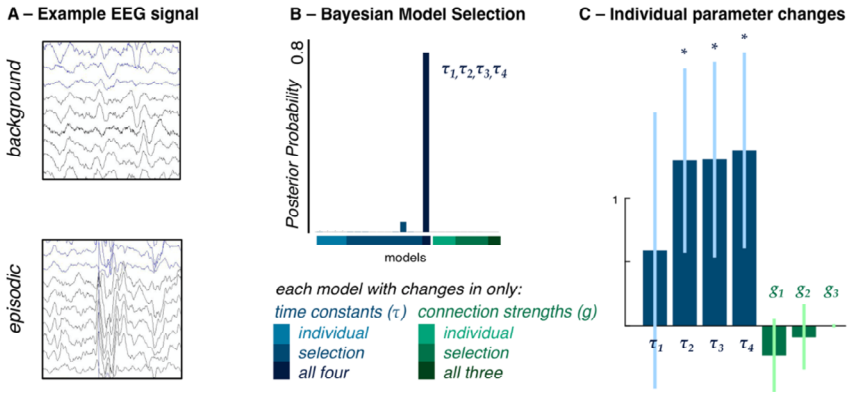

Fig. 7. Bayesian model selection and parameter averages: A) Two-second windows around all paroxysms (total: 12) and an equal number of background EEG windows were selected visually from the whole-scalp EEG - an exemplar single time window is shown here for background (top) and episodic (bottom) activity. **B)** Single source canonical microcircuits were fitted to average source localised background, and episodic power spectral densities respectively. Models that explained the difference between background and episodic activity with a limited set of free parameters (cf. Table 1) were then compared through Bayesian Model Selection - the winning model here was one where changes in all time constants were required to explain the difference between background EEG and episodic EEG. **C)** Bayesian model averaging was performed to estimate the parameter differences that explain the transition from background to episodic activity. A significant increase of time constants was estimated in spiny stellate, inhibitory interneuron and deep pyramidal cell populations.
Variables. *time constants*: τ_1 - superficial pyramidal cells, τ_2 - spiny stellate cells, τ_3 - inhibitory interneurons, τ_4 - deep pyramidal cells; *connection strengths*: g_1 - spiny stellate to superficial pyramidal cells, g_2 - deep pyramidal cells to inhibitory interneurons, g_3 - spiny stellate to inhibitory interneurons.

Time windows of 2 seconds around visually defined episodic EEG activity (Figure 7A shows an example) is source localised to a single point and their power spectral densities are averaged across time windows (total number = 12). The same number of time windows is randomly selected from artefact free background activity and source activity is estimated at the same cortical source for estimation of the power spectral densities.

In the first stage of the analysis, a single source canonical microcircuit is fitted to the empirical spectral densities for each of the two conditions separately. This results in two fully fitted DCMs, where all parameters are allowed to change between background and episodic activity. In order to assess which of these parameter changes is necessary for the observed differences conditions, Bayesian model selection was performed over a set of reduced model, where only a subset of parameters are allowed to change between episodic and background activity. This model space is laid out in detail in Table 1 - and broadly divides models into those with different combinations of changes in the synaptic dynamics (i.e. time constants, models 1-14), and the synaptic connection strengths (models 16-22). Comparing these models, the model with changes in all time constants provides the best fit (posterior probability ~ 0.76), followed by the model with changes in τ_2, τ_3, and τ_4 only (posterior probability ~ 0.13) as shown in Figure 7B.

Bayesian model averaging further provides estimates of the size and direction of changes of each parameter between background and episodic activity, shown in Figure 7C. Because the Bayesian model averaging takes into account uncertainty over specific parameter estimates, this allows for the calculation of a Bayesian confidence interval, and inference whether any given parameter change has a probability exceeding a certain significance threshold (here 99%). According to these estimates, we find that the episodic spectral densities are associated with a significant increase of time constants in the spiny stellate (τ_2), inhibitory interneuron (τ_3) and deep pyramidal cell (τ_4) populations.

The model fits can be seen in Figure 8. Figure 8A and B show the fits of the independently inverted full DCMs, whilst Figure C and D show the model fits for the reduced winning model where parameters across both conditions are identicial apart from the time constants τ_1, τ_2, τ_3, and τ_4. The paroxysms have a clear frequency peak in the low beta frequency range, which are present in both the full model fits (8B) and the reduced fits (8D) of that condition. Whilst the model fits for the full model are better for both he episodic and the background activity, most of the important differences between them are preserved well even in the reduced models where only a small subset of parameters contributes to explaining the differences seen. Most notably, the emergence of an additional frequency component in the beta range with an identical peak frequency is modelled well, whilst the relative power of high and low frequencies is not preserved as well in the reduced model prediction.

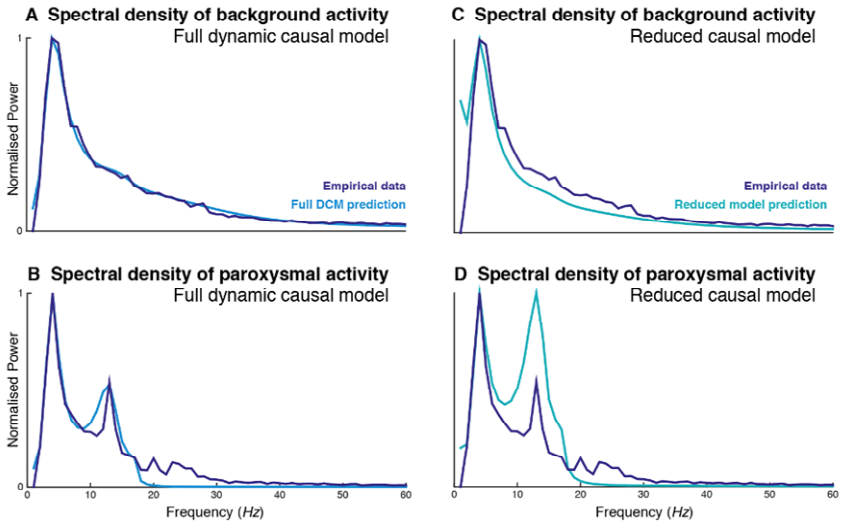

Fig. 8. Model fits for power spectral densities: Model fits are shown for the full DCM inversions (A, B) and the winning reduced models (C, D), where the differences between episodic and background activity are explained by changes in the time constants only. The top panels show the background activity, whilst the bottom panels show the paroxysmal, episodic abnormality. Model predictions and power estimation range from 4 - 60 Hz.

These findings are interesting in two ways. Firstly, identifying changes in parameters that carry biophysical relevance means that results from human EEG measurements can be used to evaluate hypotheses that emerge from molecular findings. Specifically for our case, a significant body of work has already established that antibodies against NMDAR have direct effects on glutamate transmission dynamics in a mouse model of NMDAR antibody encephalitis (Hughes et al. 2010).

These experiments investigating the blockade of NMDAR transmission show that (1) it affects mostly the temporal dynamics of the glutamate response, and not its size (as the latter is largely determined by the preserved AMPA-receptor response); and (2) the most significant effect on NMDAR availability in the mice treated with NMDAR antibody positive CSF was not within the dendritic spines (the sites of classical synaptic transmission), but on extrasynaptic NMDARs.

Our hypothesis space for the analysis presented was specifically designed to address the first aspect: In humans, can dynamic abnormalities on the EEG caused by NMDAR antibody exposure be best explained by changes in the time constants (as predicted from mouse models), or by a change in excitatory connection strengths. The findings from the Bayesian model selection in fact support changes in temporal dynamics underlying the observed EEG abnormalities, providing the first evidence from human studies that the mechanism observed in other model systems may explain the pathological features seen in patients.

However, our analysis pathway tried to explain transitions between background and paroxysmal activity within an individual patient (and during a single EEG recording) through changes in network parameters, even though the presumed underlying cause – the NMDAR antibody – are present and active throughout the recording. Our interpretations of the findings therefore do not suggest that NMDAR cause a permanent change in the time constants, but rather promote a volatility in time constants that facilitates the transient appearance of paroxysmal abnormalities.

Time constants themselves are known to be a composite measures that depend on particular physiological states and are therefore not actually constant, but are themselves dynamic in their expression (Koch et al. 1996). One identified mechanism of activity-dependent changes in the temporal profile of postsynaptic integration is the recruitment of extrasynaptic NMDAR during excessive stimulation at AMPAR-only synapses (Clark & Cull-Candy 2002). The observations made in animal models suggest that NMDAR antibodies change the balance between extrasynaptic and synaptic NMDAR, and are therefore likely to change the dynamics of time-constant changes governing physiological synaptic transmission, making the transient changes in temporal dynamics described in our model a plausible pathophysiological mechanism.

The main effect as estimated from the canonical microcircuit is an increase of the time constants in a variety of neurons. The increased time constants may facilitate temporal integration of neuronal signal and therefore result in an increase in the coupling between superficial and deep pyramidal cells, potentially explaining the high amplitude paroxysmal activity observed on the EEG.

Limitations

In this study we applied DCM to episodic abnormalities in patient EEGs. In the process of this study, many simplifying assumptions are required to render such an empirical inversion tractable.

We have chosen to investigate intrinsic changes in neuronal population coupling and dynamics, based on a single trace extracted from a ‘virtual electrode’ at a source location estimated from the paroxysmal abnormalities. Thus we have not specifically addressed any larger scale topographic heterogeneity or network level interactions in this study.

Furthermore, the current analysis describes only the state switching between short paroxysmal abnormalities and the patient-specific background, and not the transition into pathological EEG patterns at the onset of illness (as this data is rarely available). This means that the inference we draw is one regarding fluctuations in an already pathological state, that may explain the variations in the EEG phenotypes observed.

The study also is specifically designed to investigate different mechanisms intrinsic to a single cortical source - we therefore do not model differences in input or

extrinsic connectivity. However, these are likely to contribute to state switching between the different dynamical states illustrated here, and will be the focus of further modelling research.

Take home message for Clinicians

This chapter offers an introduction to using empirical electrophysiological data to inform the parameterisation of advanced mesoscale neuronal models. This approach is particularly suited to link conditions where long-lasting, or even permanent pathologies (such as a lesion, or a molecular abnormality) find their pathophysiological expression only transiently in abnormal neuronal dynamics. The prime example for this is epilepsy, where abnormalities in the neuronal network produce intermittent and unpredictable abnormal states – epileptic seizures; but the same approach is also relevant to neuropsychiatric conditions, or the encephalitides, as discussed in this chapter.

Linking macroscopic, and often transient observations – such as clinical features, or EEG measurements – to underlying causes, even where they are understood in some detail is far from intuitive. Whilst the discovery of NMDAR binding antibody in the context of clinical autoimmune encephalitis clearly suggest a direct pathophysiological role for the antibody, understanding how it affects synaptic transmission in order to produce the abnormalities observed in neuronal states still remains difficult. This currently limits the prognostic and diagnostic value of EEG recordings, as well as hindering the development of targeted therapies.

The chapter introduces mesoscale computational modelling as a possible link between molecular or microstructural pathology, and macroscale phenotypes. Exploiting recent advances in both neuronal models of cortical function, and the fitting of parameters to empirical data within the well-established framework of Dynamic Causal Modelling allows for the testing of specific mechanistic hypotheses. The approach presented here allows researchers to directly address specific questions emerging from other disease models and evaluate whether evidence for similar mechanisms can be identified in human patients.

These computational models can facilitate a thorough understanding of the dynamic effects of apparently static abnormalities within an organism. Whilst they are not set up to reproduce the complexity of whole organisms, they allow the mapping of changes in the model parameters and dynamic outputs of the model. They are therefore an ideal tool to further explore hypotheses derived from newly identified genetic mutations, other molecular causes, or animal models of specific conditions.

Regarding the example of NMDAR encephalitis, the computational approach presented here provides empirical evidence for electrophysiological abnormalities being caused by changes in the temporal dynamics of synaptic transmission, rather than changes in connection strength. This replicates findings from animal models, providing converging lines of evidence that the observations made in the animal models is in fact related to the dynamic abnormalities we see in human patients.

Take home message for Computationalists

Advanced computational modelling in the analysis of electrophysiological signals is currently limited to a few source localisation algorithms in routine clinical practice. Yet recent advances both in machine learning algorithms, and the increased availability of computational resources provide an opportunity to integrate advanced computational analysis of neuronal signals into clinical practice.

The approach presented in this chapter is deliberately using empirical data to parameterise an existing, full generative model of neuronal population function (as opposed to more data driven machine learning approaches): The generative model both constrains the inverse problem we face by attempting to make inference on mesoscale mechanisms from macroscale recordings. But crucially it also forces the results of the computational analysis to be cast within biophysically plausible terms.

The work presented here should not be seen in isolation, but instead provides a novel, and necessary perspective on an existing scientific question. When attempting to identify causative mechanisms in NMDAR encephalitis, neither computational nor animal-based approaches will give us the full answer. Rather the strength of the evidence lies in the use of existing evidence from other model systems to constrain the computational analysis to only address specific, competing hypotheses. This 'evidence accumulation' is easiest where all lines of evidence refer to similar neurophysiological concepts (e.g. connection strengths, time constants, gain parameters). Indeed on can foresee the application of dynamic causal modelling to data from animal models to provide a formal integration of animal and human measurements.

The Dynamic Causal Modelling approach presented here furthermore has the benefit that it will provide estimates of model evidence (to decide between competing hypotheses) as well as individual parameter estimates (to evaluate specific effects), derived from fitting the model to empirical data. This combines the benefits of data driven analysis: the DCM can provide direct empirical measures for, or against specific hypotheses, as well as being utilised as a generative model of neuronal dynamics whose parameter space can be explored in detail.

The example presented shows that the approach is uniquely flexible and can be applied to a wide variety of contexts. All software used here, including model inversion techniques, canonical microcircuitry models and classical EEG analysis modules, is freely available as part of the Statistical Parametric Mapping (SPM) academic freeware (www.fil.ion.ucl.ac.uk/spm).

Acknowledgments

RER is funded by a Wellcome Trust Clinical Research Fellowship (106556/Z/14/Z). KJF is funded by a Wellcome Trust Principal Research Fellowship (088130/Z/09/Z).

References

Aburn, M.J. et al., 2012. Critical Fluctuations in Cortical Models Near Instability. *Frontiers in Physiology*, 3, p.331.

Babajani-Feremi, A. & Soltanian-Zadeh, H., 2010. Multi-area neural mass modeling of EEG and MEG signals. *NeuroImage*, 52(3), pp.793–811.

Bastos, A.M. et al., 2012. Canonical Microcircuits for Predictive Coding. *Neuron*, 76(4), pp.695–711.

Buffalo, E.A. et al., 2011. Laminar differences in gamma and alpha coherence in the ventral stream. *Proceedings of the National Academy of Sciences*, 108(27), pp.11262–11267.

Bullmore, E. & Sporns, O., 2009. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3), pp.186–198.

Canolty, R.T. & Knight, R.T., 2010. The functional role of cross-frequency coupling. *Trends in Cognitive Sciences*, 14(11), pp.506–515.

Carandini, M., 2012. From circuits to behavior: a bridge too far? *Nature Neuroscience*, 15(4), pp.507–509.

Clark, B.A. & Cull-Candy, S.G., 2002. Activity-dependent recruitment of extrasynaptic NMDA receptor activation at an AMPA receptor-only synapse. *Journal of Neuroscience*, 22(11), pp.4428–4436.

Cooray, G.K. et al., 2016. Dynamic causal modelling of electrographic seizure activity using Bayesian belief updating. *NeuroImage*, 125, pp.1142–1154.

Dalmau, J. et al., 2008. Anti-NMDA-receptor encephalitis: case series and analysis of the effects of antibodies. *The Lancet Neurology*, 7(12), pp.1091–1098.

David, O. et al., 2006. Dynamic causal modeling of evoked responses in EEG and MEG. *NeuroImage*, 30(4), pp.1255–1272.

David, O. & Friston, K.J., 2003. A neural mass model for MEG/EEG: *NeuroImage*, 20(3), pp.1743–1755.

Van Dellen, E. et al., 2012. MEG Network Differences between Low- and High-Grade Glioma Related to Epilepsy and Cognition D. R. Chialvo, ed. *PLoS ONE*, 7(11), p.e50122.

Do, C.B. & Batzoglou, S., 2008. What is the expectation maximization algorithm? *Nature Biotechnology*, 26(8), pp.897–899.

Du, J., Vegh, V. & Reutens, D.C., 2012. The Laminar Cortex Model: A New Continuum Cortex Model Incorporating Laminar Architecture L. J. Graham, ed. *PLoS Computational Biology*, 8(10), p.e1002733.

Eadie, M.J. & Bladin, P.F., 2001. *A Disease Once Sacred. A History of the Medical Understanding of Epilepsy* 1st ed., New Barnet: John Libbey & Co Ltd.

Fisher, R.S. et al., 2014. ILAE Official Report: A practical clinical definition of epilepsy. *Epilepsia*, 55(4), pp.475–482.

Florance, N.R. et al., 2009. Anti-N-methyl-D-aspartate receptor (NMDAR) encephalitis in children and adolescents. *Annals of Neurology*, 66(1), pp.11–18.

Freestone, D.R. et al., 2014. Estimation of effective connectivity via data-driven neural modeling. *Frontiers in Neuroscience*, 8, p.383.

Friston, K., 2005. A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1456), pp.815–836.

Friston, K. et al., 2008. Multiple sparse priors for the M/EEG inverse problem. *NeuroImage*, 39(3), pp.1104–1120.

Friston, K. et al., 2007. Variational free energy and the Laplace approximation. *NeuroImage*, 34(1), pp.220–234.

Friston, K., Zeidman, P. & Litvak, V., 2015. Empirical Bayes for DCM: A Group Inversion Scheme. *Frontiers in Systems Neuroscience*, 9(November), pp.1–10.

Friston, K.J. et al., 2016. Bayesian model reduction and empirical Bayes for group (DCM) studies. *NeuroImage*, 128, pp.413–431.

Gitiaux, C. et al., 2013. Early electro-clinical features may contribute to diagnosis of the anti-NMDA receptor encephalitis in children. *Clinical Neurophysiology*, 124(12), pp.2354–2361.

Gollas, F. & Tetzlaff, R., 2005. Modeling brain electrical activity in epilepsy by reaction-diffusion cellular neural networks. In R. A. Carmona & G. Linan-Cembrano, eds. p. 219.

Goodfellow, M., Schindler, K. & Baier, G., 2012. Self-organised transients in a neural mass model of epileptogenic tissue dynamics. *NeuroImage*, 59(3), pp.2644–2660.

Hansen, S.T. & Hansen, L.K., 2015. EEG source reconstruction performance as a function of skull conductance contrast. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, pp. 827–831.

Heitmann, S., Gong, P. & Breakspear, M., 2012. A computational role for bistability and traveling waves in motor cortex. *Frontiers in Computational Neuroscience*, 6, p.67.

Helbig, I. et al., 2008. Navigating the channels and beyond: unravelling the genetics of the epilepsies. *The Lancet Neurology*, 7(3), pp.231–245.

Hildebrand, M.S. et al., 2013. Recent advances in the molecular genetics of epilepsy. *Journal of Medical Genetics*, 50(5), pp.271–279.

Hodgkin, A.L. & Huxley, A.F., 1952. A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology*, 117(4), pp.500–44.

Hughes, E.G. et al., 2010. Cellular and Synaptic Mechanisms of Anti-NMDA Receptor Encephalitis. *Journal of Neuroscience*, 30(17), pp.5866–5875.

Jansen, B.H. & Rit, V.G., 1995. Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biological Cybernetics*, 73(4), pp.357–66.

Julier, S.J. & Uhlmann, J.K., 2004. Unscented Filtering and Nonlinear Estimation. *Proceedings of the IEEE*, 92(3), pp.401–422.

Jung, R. & Berger, W., 1979. Hans Bergers Entdeckung des Elektrenkephalogramms und seine ersten Befunde 1924-1931. *Archiv fuer Psychiatrie und Nervenkrankheiten*, 227(4), pp.279–300.

Kawato, M., Hayakawa, H. & Inui, T., 1993. A forward-inverse optics model of reciprocal connections between visual cortical areas. *Network*, 4(4), p.4150422.

Koch, C., Rapp, M. & Segev, I., 1996. A brief history of time (constants). *Cerebr Cortex*, 6, pp.93–101.

Lantz, G., Grouiller, F. & Spinelli, L., 2011. Localisation of Focal Epileptic Activity in Children Using High Density EEG Source Imaging. *Epileptologie*, 28, pp.84–90.

Lesca, G. et al., 2013. GRIN2A mutations in acquired epileptic aphasia and related childhood focal epilepsies and encephalopathies with speech and language dysfunction. *Nature Genetics*, 45(9), pp.1061–1066.

Litvak, V. et al., 2015. Empirical Bayes for Group (DCM) Studies: A Reproducibility Study. *Frontiers in Human Neuroscience*, 9(Dcm), pp.1–12.

Di Lollo, V., 2012. The feature-binding problem is an ill-posed problem. *Trends in Cognitive Sciences*, 16(6), pp.317–321.

Lopes da Silva, F.H., 2010. *MEG: An Introduction to Methods*, Oxford University Press.

Lopes da Silva, F.H. et al., 1974. Model of brain rhythmic activity. The alpha-rhythm of the thalamus. *Kybernetik*, 15(1), pp.27–37.

Meijer, H.G.E. et al., 2015. Modeling Focal Epileptic Activity in the Wilson–Cowan Model with Depolarization Block. *The Journal of Mathematical Neuroscience*, 5(1), p.7.

Miller, I.O. & Sotero de Menezes, M.A., 2014. SCN1A-Related Seizure Disorders. *GeneReviews*.

Moran, R., Pinotsis, D. a & Friston, K., 2013. Neural masses and fields in dynamic causal modeling. *Frontiers in Computational Neuroscience*, 7(May), p.57.

Nevado-Holgado, A.J. et al., 2012. Characterising the dynamics of EEG waveforms as the path through parameter space of a neural mass model: Application to epilepsy seizure evolution. *NeuroImage*, 59(3), pp.2374–2392.

Nosadini, M. et al., 2015. Longitudinal Electroencephalographic (EEG) Findings in Pediatric Anti-N-Methyl-D-Aspartate (Anti-NMDA) Receptor Encephalitis: The Padua Experience. *Journal of Child Neurology*, 30(2), pp.238–245.

Pal, D. & Helbig, I., 2015. Commentary: Pathogenic EFHC1 mutations are tolerated in healthy individuals dependent on reported ancestry. *Epilepsia*, 56(2), pp.195–196.

Panayiotopoulos, C., 2005. *The Epilepsies* 1st ed., Oxford: Bladon Medical Publishing.

Papadopoulou, M. et al., 2015. Tracking slow modulations in synaptic gain using dynamic causal modelling: Validation in epilepsy. *NeuroImage*, 107, pp.117–126.

Quintáns, B. et al., 2014. Medical genomics: The intricate path from genetic variant identification to clinical interpretation. *Applied & Translational Genomics*, 3(3), pp.60–67.

Shirvany, Y. et al., 2012. Non-invasive EEG source localization using particle swarm optimization: A clinical experiment. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, pp. 6232–6235.

Sitnikova, E. et al., 2008. Granger causality: Cortico-thalamic interdependencies during absence seizures in WAG/Rij rats. *Journal of Neuroscience Methods*, 170(2), pp.245–254.

Spillane, J., Kullmann, D.M. & Hanna, M.G., 2015. Genetic neurological channelopathies: molecular genetics and clinical phenotypes. *Journal of Neurology, Neurosurgery & Psychiatry*, p.jnnp–2015–311233.

Thomas, E.M. et al., 2008. Seizure detection in neonates: Improved classification through supervised adaptation. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, pp. 903–906.

Thomas, R.H. & Berkovic, S.F., 2014. The hidden genetics of epilepsy—a clinically important new paradigm. *Nature Reviews Neurology*, 10(5), pp.283–292.

Tong, S. & Thakor, N.V., 2009. *Quantitative EEG Analysis Methods and Clinical Applications*, Artech House.

Wang, Y. et al., 2014. Dynamic Mechanisms of Neocortical Focal Seizure Onset B. Ermentrout, ed. *PLoS Computational Biology*, 10(8), p.e1003787.

Werner, G., 2007. Metastability, criticality and phase transitions in brain and its models. *Biosystems*, 90(2), pp.496–508.

Wilson, H.R. & Cowan, J.D., 1972. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical journal*, 12(1), pp.1–24.

Zschocke, S. & Hansen, H.-C., 2012. *Klinische Elektroenzepalographie* 3rd ed., Berlin, Heidelberg: Springer-Verlag.