Evidence-based modeling of network discharge dynamics during periodic pacing to control epileptiform activity

Keith Bush[a], Gabriella Panuccio[b], Massimo Avoli[b], Joelle Pineau[c]

[a]Department of Computer Science, University of Arkansas at Little Rock (UALR), Little Rock, Arkansas, USA 72204
[b]Montreal Neurological Institute and Departments of Neurology, Neurosurgery, and Physiology, McGill University, Montreal, Quebec, Canada H3A 2B4
[c]School of Computer Science, McGill University, Montreal, Quebec, Canada H3A 2A7

Abstract

Deep brain stimulation (DBS) is a promising therapeutic approach for epilepsy treatment. Recently, research has focused on the implementation of stimulation protocols that would adapt to the patients need (adaptive stimulation) and deliver electrical stimuli only when it is most useful. A formal mathematical description of the effects of electrical stimulation on neuronal networks is a prerequisite for the development of adaptive DBS algorithms. Using tools from non-linear dynamic analysis, we describe an evidence-based, mathematical modeling approach that 1) accurately simulates epileptiform activity at time-scales of single and multiple ictal discharges, 2) simulates modulation of neural dynamics during epileptiform activity in response to fixed, low-frequency electrical stimulation, 3) defines a mapping from real-world observations to model state, and 4) defines a mapping from model state to real-world observations. We validate the real-world utility of the model's properties by statistical comparison between the number, duration, and interval of ictal-like discharges observed *in vitro* and those simulated *in silica* under conditions of repeated stimuli at fixed-frequency. These validation results confirm that the evidence-based modeling approach captures robust, informative features of neural network dynamics of *in vitro* epileptiform activity under periodic pacing and support its use for further implementation of adaptive DBS protocols for epilepsy treatment.

Keywords: Deep brain stimulation, Epilepsy, Manifold embedding, Modeling, Neural network dynamics

1. Introduction

It is well established that electrical stimulation can manipulate the dynamics of the neuronal activity generated by brain slices maintained *in vitro* to reduce the frequency, duration, or amplitude of ictal-like discharges (Bawin et al., 1986; Nakagawa and Durand, 1991; Durand, 1993; Durand and Warman, 1994; Schiff et al., 1994; Jerger and Schiff, 1995; Gluckman et al., 1996; Barbarosie and Avoli, 1997; Warren and Durand, 1998; D'Arcangelo et al., 2005). In particular, fixed-frequency stimulation, also known as periodic pacing, has achieved robust suppression of epileptiform activity in a diversity of *in vitro* chemical models of ictogenesis, e.g., high-K^+ (Jerger and Schiff, 1995), low-Mg^{2+} (Schiff et al., 1994), and 4-aminopyridine (4AP) (Barbarosie and Avoli, 1997; D'Arcangelo et al., 2005). The next logical step beyond periodic pacing is the deployment of an adaptive control system that intelligently selects the timing of stimulations to maximize the suppression of ictal-like discharges while minimizing current pulses. Such intelligent control systems rely on accurate knowledge of causal relationships in the dynamics of the system of interest. Therefore, they are feasible only to the degree with which the modulating effects of electrical stimulation can be accurately modeled and predicted. However, a complete understanding of the dynamics underlying the response of neuronal networks to electrical stimulation is still lacking.

Predictive models must accurately capture both the system's state (a numerical representation of the system's current behavior) and its transition function (a mapping from the current state and control action, e.g. stimulation, onto the future state). The state and transition function can be modeled by one of two distinct approaches. The first principles approach builds up the model from neurophysiology theory. The evidence-based approach builds up the model from abstract rules that best describe observations of the real-world system. With respect to neural network dynamics of *in vitro* epileptiform activity under periodic pacing, only first principles models have been studied (see for example Biswal and Dasgupta (2002); Franaszczuk et al. (2003)). However, two weaknesses make them poorly suited for use in the construction and/or validation of neural control systems. First, they are unable to quantitatively reproduce the efficacy of dynamic modulation of neuronal activity as a function of stimulation frequency with the accuracy necessary to make effective control decisions. Second, they provide no mapping between the model's state space and real-world observations.

Here, we describe an evidence-based modeling approach that combines

non-linear dynamic analysis (time-delay embeddings) to identify the neural system's state with machine learning (nearest neighbor methods) to approximate the neural system's transition function. Directly from the data, this approach 1) simulates the occurrence of epileptiform activity *in vitro* at time-scales of single and multiple ictal discharges, 2) simulates modulation of epileptiform activity in response to periodic pacing protocols, 3) defines a mapping from real-world observations to model state (allowing the model to act as a stand-alone predictor or as part of a control system), and 4) defines a mapping from model state to real-world observations (allowing the model to act as a simulator). We quantitatively validate the model's predictive accuracy under multiple stimulation protocols by first-order and higher-order statistical comparison between dynamics of the model and of a previously unseen dataset. We also compare the model's simulated field potentials directly against previously unseen real-world field potential recordings as qualitative, visual evidence of the model's efficacy in describing neural network dynamics of *in vitro* epileptiform activity under periodic pacing.

2. Materials and Methods

2.1. Theory

The state of a dynamic system is a real-valued vector that uniquely determines the future state of the system according to the system's rules. These rules are termed the transition function. Models of real-world dynamic systems are founded on accurate identification of the system's underlying state and transition function. With regard to neuronal networks studied *in vitro*, obstacles to system identification are numerous: dynamic variance, nonstationarity, limited exploration, noise, transients, and limited knowledge of first principles; however, the most challenging obstacle is partial observability, which implies that all of the components required to uniquely quantify the neuronal networks state cannot be directly observed. Indeed, network activity generated by brain slices maintained *in vitro* contains millions of hypothetically measurable parameters, yet it can only be observed through a small number of field potentials.

Nonlinear dynamic systems theory provides means of reconstructing complete state observability from partial observability via the method of delayed embeddings, formalized by Takens' Theorem (Takens, 1980). Here we present the key points of Takens' Theorem utilizing the notation of Huke (Huke, March, 2006) in a deterministically forced system.

3

Let the system state, $\mathbf{s}$, be a M-dimensional, real-valued, bounded vector space indexed by time, t, and a is a real-valued control action. Consider the state update function, f, is a deterministic function:

$$\mathbf{s}(t+1) \;=\; f(\mathbf{s}(t), a(t)). \tag{1}$$

Let the action choice $a(t)$ be set by a deterministic control function, π:

$$a(t) \;=\; \pi(\mathbf{s}(t)), \tag{2}$$

then Equation 2 may be composed with Equation 1:

$$\begin{aligned}
\mathbf{s}(t+1) \;&=\; f(\mathbf{s}(t), \pi(\mathbf{s}(t))) \\
&=\; \phi(\mathbf{s}(t)), \tag{3}
\end{aligned}$$

where ϕ specifies the discrete time evolution of the forced system. If ϕ is a smooth map, $\phi : \mathbb{R}^M \to \mathbb{R}^M$, and the system is observed through partial state, $\tilde{s}$, via function, y, such that:

$$\tilde{s}(t) \;=\; y(\mathbf{s}(t)),$$

and $y : \mathbb{R}^M \to \mathbb{R}$, then if ϕ^{-1} exists, and ϕ, ϕ^{-1}, and y are continuously differentiable, we may apply Takens' Theorem (Takens, 1980) to reconstruct the complete state space of the observed system. Thus, for each $\tilde{s}(t)$, we can construct a vector $\mathbf{s}_E(t)$,

$$\mathbf{s}_E(t) \;=\; [\tilde{s}(t), \tilde{s}(t-1), ..., \tilde{s}(t-(E-1))], \; E > 2M,$$

such that $\mathbf{s}_E$ lies on a subset of $\mathbb{R}^E$ which is an embedding of $\mathbf{s}$. Because embeddings preserve the connectivity of the original vector-space, in the context of system dynamics, a mapping ψ,

$$\mathbf{s}_E(t+1) \;=\; \psi(\mathbf{s}_E(t)), \tag{4}$$

may be substituted for f (Eqn. 3), and vectors $\mathbf{s}_E(\mathrm{t})$ may be substituted for vectors $\mathbf{s}(t)$ without loss of generality.

The properties of a correctly constructed embedding vector, $\mathbf{s}_E$, guarantee that we can reconstruct the complete state of a large class of partially observable dynamical systems directly from a sequence of their observations and that this reconstruction may occur independently of identification of

the transition function, ψ. We exploit both of these facts in our methodology to use simple but powerful semi-parametric modeling methods, which differentiates our approach from important earlier approaches to modeling forced physiological systems using parametric functions of Volterra-Wiener kernels (Marmarelis and Orme, 1993; Marmarelis, 1997; Marmarelis et al., 1999).

Takens' Theorem does not define how to compute the embedding dimension, E, of arbitrary sequences of observations nor does it provide for a test to determine if the theorem is applicable to a specific dataset. In practice the intrinsic dimension, M, of a system is often unknown. Finding high-quality embedding parameters of challenging domains, such as chaotic or noise-corrupted nonlinear signals, occupy much of the fields of subspace identification and nonlinear dynamic analysis. Numerous methods of note exist (Galka, 2000; Kennel and Abarbanel, 2002; Katayama, 2005), drawn from both disciplines. We build-upon a spectral approach (Galka, 2000) premised on the singular value decomposition (SVD). It has the advantage of being non-parametric, computationally efficient, and robust to additive noise—all of which are useful in practical application.

We summarize the spectral parameter selection algorithm as follows. Given a sequence of state observations $\tilde{\mathbf{s}}$ of length $\tilde{S}$, we choose a *sufficiently large* embedding dimension, $\hat{E}$. Sufficiently large refers to a cardinality of dimension which is *certain* to be greater than twice the dimension in which the actual state space resides. For each embedding window size, $t_{\min} \in \{\hat{E}, ..., \tilde{S}\}$, we define a matrix $\mathbf{S}_{\hat{E}}(t_{\min})$ having row vectors, $\mathbf{s}_{\hat{E}}(t), t \in \{t_{\min}, ..., \tilde{S}\}$, constructed according to

$$\mathbf{s}_{\hat{E}}(t) = [\tilde{s}(t), \tilde{s}(t-\tau), ..., \tilde{s}(t-(\hat{E}-1)\tau)], \qquad (5)$$

where $\tau = t_{\min}/(\hat{E}-1)$. We compute the SVD of the matrix $\mathbf{S}_{\hat{E}}(t_{\min})$, and we record the vector of singular values, $\boldsymbol{\sigma}(t_{\min})$. Embedding parameters of $\tilde{\mathbf{s}}$ are found by analysis of the sequence of second singular values, $\sigma_2(t_{\min})$, $t_{\min} \in \{\hat{E}, ..., \tilde{S}\}$. The value of $t_{\min}$ at the first local maxima of this sequence is the approximate embedding window size, $T_{\min}$. The approximate embedding dimension, E, is the number of *non-trivial* singular values of $\boldsymbol{\sigma}(T_{\min})$, where we define non-trivial as a value greater than the longterm trend of $\boldsymbol{\sigma}(t_{\min})$ with respect to $\boldsymbol{\sigma}(T_{\min})$, $t_{\min} \gg T_{\min}$.

The spectral method provides a mechanism for efficiently projecting into, and out of, the E-dimensional state space. These projections exist because

the SVD returns a decomposition of a matrix, $\mathbf{X}$, as $\mathbf{X} = \mathbf{U \Sigma V}^T$. It is well-known that matrix $\mathbf{V}$ defines an orthonormal basis of covariance matrix, $\mathbf{X}^T \mathbf{X}$, with vectors ordered by the corresponding singular values, $diag(\mathbf{\Sigma})$. Thus, $\mathbf{V}_{\hat{E}}[1, ..., E](T_{\min})$, i.e., the first E columns of matrix $\mathbf{V}$ returned by the SVD of the embedding $\mathbf{S}_{\hat{E}}(T_{\min})$, form a basis, renamed $\mathbf{V}_E$, of a low-dimensional manifold, $\mathbf{S}_E$, which maximally captures the variance of the embedded observations. Therefore, we make a change of coordinates such that $\mathbf{V}_E \mathbf{S}_E = \mathbf{S}_{\hat{E}}(T_{\min})$. We achieve this by computing the Moore-Penrose pseudoinverse, $\mathbf{V}_E^\dagger$, of $\mathbf{V}_E$ and left-multiply such that $\mathbf{S}_E = \mathbf{V}_E^\dagger \mathbf{S}_{\hat{E}}(T_{\min})$. $\mathbf{S}_E$ is model's state space.

The matrix $\mathbf{V}_E^\dagger$ defines the projection from an $\hat{E}$-dimensional vector of embedded observations, given by Equation 5, into the reconstructed, reduced E-dimensional state space, $\mathbf{S}_E$. $\mathbf{V}_E$ defines the opposite projection (from the E-dimensional state space to the space of embedded observations). Thus, any numerical system state can be mapped, without loss of generality, to an approximate sequence of observations. Likewise, sequences of observations can be mapped to a numerical state.

The preservation of locality and dynamics afforded by such an embedding allows approximation of the underlying dynamic system. To model these dynamics we assume that the derivative of the region surrounding each state is well-approximated by the derivative of the state itself, a *nearest-neighbors* derivative (Parlitz and Merkwirth, 2000). We simulate trajectories as iterative numerical integration of the local state and gradient. We define the model and integration process formally.

Consider a dataset $\mathcal{D}$ of observations, $\tilde{s}(t)$, $t \in \{1, ..., \tilde{S}\}$. Applying the spectral embedding method to $\mathcal{D}$ yields parameters E and $T_{\min}$. Embedding $\mathcal{D}$ according to Equation 5 and projecting into state space via $\mathbf{V}_E^\dagger$ yields a sequence of vectors $\mathbf{s}_E(t)$ in $\mathbb{R}^E$ indexed by $t \in \{T_{\min}, ..., \tilde{S}\}$. A model, $\mathcal{M}$, of $\mathcal{D}$ is the set of vectors, $\mathbf{m}(t) = \mathbf{s}_E(t)$, $t \in \{T_{\min}, ..., \tilde{S}\}$.

Consider a state vector $\mathbf{x}(i)$ in $\mathbb{R}^E$ indexed by simulation time, i. To numerically integrate this state we define the gradient according to our definition of locality: nearest neighbor. The model's nearest neighbor of $\mathbf{x}(i)$,

denoted $\mathbf{m}(t_{\mathbf{x}(i)})$, is defined as[1]

$$\mathbf{m}(t_{\mathbf{x}(i)}) = \operatorname*{argmin}_{\mathbf{m}(t) \in \mathbf{M}} \|\mathbf{m}(t) - \mathbf{x}(i)\|_2. \tag{6}$$

The model gradient and numerical integration are defined, respectively, as,

$$\begin{aligned} \nabla_{\mathbf{x}(i)} &= \mathbf{m}(t_{\mathbf{x}(i)} + 1) - \mathbf{m}(t_{\mathbf{x}(i)}) \quad \text{and} \tag{7} \\ \mathbf{x}(i+1) &= \mathbf{x}(i) + \nabla_{\mathbf{x}(i)} + \eta, \tag{8} \end{aligned}$$

where η is a vector of noise. Applying Equations 6–8 iteratively simulates a trajectory of the underlying system, termed a *surrogate dataset*, $\tilde{\mathcal{D}}$, or simply *surrogate*. Surrogates are initialized from some state, $\mathbf{x}(0)$. Equation 8 assumes that dataset $\mathcal{D}$ contains noise. This noise biases the derivative estimate in $\mathbb{R}^E$, via Equation 5. In practice, a small amount of additive noise facilitates generalization. Note, our technique does not preclude the use of non-local function approximation, but here we assume a sufficient density of data exists to reconstruct the embedded state space with minimal bias.

As defined, the model generates surrogates that mimic real-world neural network dynamics without stimulation (experimental control condition). We extend this model to account for electrical stimulation effects as follows. Let dataset $\mathcal{D}$ also include sequence $\mathbf{a}$, comprised of actions, $a(t)$, $t \in \{1, ..., \tilde{S}\}$. We now define the model, $\mathcal{M}$, of $\mathcal{D}$ to be the set of 2-tuples, $\mathbf{m}(t) = \{\mathbf{s}_E(t), a(t)\}$, $t \in \{T_{\min}, ..., \tilde{S}\}$, and we add the operation $\mathcal{Z}(\mathbf{m}(t)) \equiv [\mathbf{s}(t), \omega a(t)]$ where ω is a parameter which scales the action dimension to the model's state space. In this formulation, we redefine the nearest neighbor to be conditioned on the action input to the model, $a(i)$:

$$\mathbf{m}(t_{\mathbf{x}(i)}) = \operatorname*{argmin}_{\mathbf{m}(t) \in \mathbf{M}} \|\mathcal{Z}(\mathbf{m}(t)) - [\mathbf{x}(i), \omega a(i)]\|_2, \ a \in \mathbb{R}. \tag{9}$$

Using Equation 9, numerical integration of the model can generate surrogates for arbitrary stimulation protocols.

We extend the model definition further by introducing the notion of a labeled state space. Let the dataset $\mathcal{D}$ also include the sequence $\mathbf{l}$, comprised of labels $l(t)$, $t \in \{1, ..., \tilde{S}\}$. We define the model, $\mathcal{M}$, of $\mathcal{D}$ to be the set of 3-tuples, $\mathbf{m}(t) = \{\mathbf{s}_E(t), a(t), l(t)\}$, $t \in \{T_{\min}, ..., \tilde{S}\}$, and we add the operation

---

[1]The operation $\operatorname*{argmin}_{x} f(x)$ returns the argument, x, at which the function, $f(x)$, is at its global minimum.

$\mathcal{L}(\mathbf{m}(t)) \equiv l(t)$, which extracts the label assigned to model state, $\mathbf{m}(t)$. This extension provides the necessary mathematical tools for analysis of the model's simulation performance described in Equation 11.

2.2. Calculation

To construct and validate the evidence-based modeling approach using real-world observations, we collect a dataset, $\mathcal{D}$, containing field potential recordings under each stimulation protocol, ρ, from the set of all protocols, P. For each observation, $\tilde{s}(t)$, assign the label, $l(t)$, and the control action, $c(t)$, as either *idle*=0 or *stimulation*=1. We then perform spectral embedding (c.f. Section 2.1) to identify the parameters E and $T_{\min}$. For each observation, we define the time elapsed since the last $c = 1$, termed time-since-stimulation or t_{ss} and the maximum t_{ss}, denoted $t_{\max}$, as the largest interval between stimulations in the dataset. We filter the data such that if $t_{ss}(t) > t_{\max}$ then $t_{ss}(t) = t_{\max}$ (i.e., scale the t_{ss} of the control protocol to $t_{\max}$) and then normalize t_{ss} by $t_{\max}$. Finally, we define the model action, $a(t) = t_{ss}(t)$. We then partition the dataset into disjoint training dataset, $\mathcal{D}^{trn}$, and test dataset, $\mathcal{D}^{tst}$, ensuring that both datasets contain approximately the same distribution of stimulation protocols. Assuming that N unique partitions are possible, we term this N-fold cross validation and we define the set of cross-validation tests, $\mathcal{N} \equiv \{1, ..., N\}$. For all datasets $\mathcal{D}_n^{trn}$ and $\mathcal{D}_n^{tst}$, $n \in \mathcal{N}$, we embed the datasets via parameters E and $T_{\min}$ and compute the projections V_E and $V_E^\dagger$. We then compute the reduced-dimensional states via change of coordinates and combine the states with the corresponding actions and labels to form the training models, $\mathcal{M}_n^{trn}$, and testing models, $\mathcal{M}_n^{tst}$, $n \in \mathcal{N}$.

We define measurements, Θ, and we measure each training dataset, $\mathcal{D}_n^{trn}$, $n \in \mathcal{N}$, yielding results vector, $\theta_n^{trn} = \Theta(\mathcal{D}_n^{trn})$, $n \in \mathcal{N}$. These measurements summarize important characteristics of the system that should be reproduced by the model. As an example, measurements of stimulated epileptiform activity could include the efficacy of various fixed-frequency policies in suppressing ictal-like discharges. For a wide range of parameters η and ω, we numerically integrate $\mathcal{M}_n^{trn}$, forming surrogates, $\tilde{\mathcal{D}}_n^{trn}[\eta, \omega]$ and then measure the surrogates, $\tilde{\theta}_n^{trn}[\eta, \omega] = \Theta(\tilde{\mathcal{D}}_n^{trn}[\eta, \omega])$. The best training model parameters, for each training dataset, are defined as:

$$\{\eta_n, \omega_n\} \quad = \quad \operatorname*{argmin}_{\{\eta, \omega\}}(\| \tilde{\theta}_n^{trn}[\eta, \omega] - \theta_n^{trn} \|_2). \tag{10}$$

This minimization over parameters η_n and ω_n maximizes, according to the

measurement Θ, the likelihood that dataset $\mathcal{D}_n^{trn}$ would be observed by simulation of model $\mathcal{M}_n^{trn}$.

Using parameters η_n and ω_n, we numerically integrate test models $\mathcal{M}_n^{tst}, n \in \mathcal{N}$, according to Equations 6–8, yielding test surrogates, $\tilde{\mathcal{D}}_n^{tst}, n \in \mathcal{N}$: these are the model's predictions. We report predictions as distributions of N means of K randomly initialized trials executed on each test model. Each prediction, therefore, requires the generation of $K \times N$ surrogates.

2.3. Experimental

Male, adult Sprague-Dawley rats (250–300 g) were decapitated under deep isoflurane anesthesia. The brain was quickly removed and placed in cold (0–2° C) artificial cerebro-spinal fluid (ACSF) of the following composition (mM): 124 NaCl, 2 KCl, 2 $MgSO_4$, 2 $CaCl_2$, 1.25 KH_2PO_4, 26 $NaHCO_3$ and 10 D-glucose, continuously bubbled with gas mixture (CO_2 5% and O_2 95%) to equilibrate at pH $\sim$ 7.35. Partially disconnected combined hippocampus-EC slices (450 μm thick) were cut as previously described (Panuccio et al., 2010) using a VT1000S vibratome (Leica, Germany). In these brain slices, which included the most ventral part of the hippocampal formation, fast CA3-driven interictal-like activity disclosed by 4AP bath-application was restrained to the hippocampus proper and did not propagate to the EC (cf., Avoli et al. (1996), but see also Avoli et al. (2002)). Slices were then transferred to an interface recording chamber, lying between warm ($\sim$ 32° C) ACSF and humidified gas (CO_2 5% and O_2 95%), where they were allowed to recover for $\geq$ 1 hour before beginning continuous bath-application of 4AP ($\sim$ 1 ml/min). Chemicals were acquired from Sigma-Aldrich Canada, Ltd. (Oakville, Ontario). All the procedures were carried on in accordance to the Canadian Council on Animal Care and McGill University guidelines. Field potential recordings were performed with ACSF-filled pipettes (tip diameter < 10 μm; resistance < 10 MΩ) pulled from borosilicate capillary tubing (World Precision Instruments Inc., Sarasota, FL, USA) using a P-97 puller (Sutter Instrument, Novato, CA, USA). Extracellular signals were fed to a Cyberamp 380 amplifier (Molecular Devices, Palo Alto, CA) connected to a digital interface device (Digidata 1320A, Molecular Devices). Data were acquired at a sampling rate of 5 KHz and low-passed at 2 KHz, using the software Clampex 8.2 (Molecular Devices), stored on the hard drive and analyzed off-line. Recording electrodes were placed in the EC deep layers and in the subiculum. Extracellular current pulses (100–250 μA, pulse width 100 μs) were delivered in the subiculum through a bipolar concentric Pt-Ir

electrode (FHC, Bowdoin, ME, USA) plugged onto a high voltage stimulus isolator unit (A360, WPI Inc., Sarasota, Florida, USA) connected to the pulse generator Pulsemaster A300 (WPI Inc., Sarasota, Florida, USA). Stimulus intensity was established prior to beginning the experimental protocols in order to reliably induce an interictal-like event in the EC. The following periodic pacing protocols were implemented: 0.2 Hz, 0.5 Hz, 1 Hz, and 2 Hz. Each stimulation phase was preceded by a control period and followed by a post-stimulation recovery period, which served as the control recording for the following stimulation protocol. Recordings were pursued until at least 4 ictal-like discharges were generated (control and ineffective stimulation protocols) or for a period of ≥ 3 times the previously observed interval between ictal-like discharges (effective stimulation).

We assigned a label to each data point of the recordings: ictal=1 and non-ictal=0. The latter includes both baseline and interictal-like discharges. We also assigned a time-since-stimulation value, t_{ss}, where $t_{\max}$, the maximum interval between pulses, is 5.0 s, corresponding to the 0.2 Hz protocol. We low-pass filtered (50 Hz) the recordings and desampled them (500 Hz). We then centered the data by a filter such that baseline data points had a mean potential of 0.0 mV. We combined recordings from 5 brain slices to form a dataset of 30,421 s, including 46 ictal-like discharges. We partitioned the dataset into cross-validation training datasets, $\mathcal{D}_n^{trn}$ and test datasets, $\mathcal{D}_n^{tst}$, $n \in \mathcal{N}$, $\mathcal{N} \equiv \{1, ..., 6\}$, comprised of data from 3 and 2 *in vitro* experiments, respectively (6 partitions constructed from the 5 slice recordings ensured that each training and testing pair contained an approximately balanced composition of control and stimulation protocols). We formed models from each of these datasets, $\mathcal{M}_n^{tst}$ and $\mathcal{M}_n^{tst}$, $n \in \mathcal{N}$. A visual example of the embedding spectrum of one training dataset is shown in Fig. 1a. Here, by the methods described in Sections 2.1 and 2.2, we computed embedding parameters $E = 3$ and $T_{\min} = 0.5$. A projection of the data, given by Fig. 1b and c, illustrates the geometry of the embedding as well as the relative positions of non-ictal and ictal data points.

A model of neural network dynamics of *in vitro* epileptiform activity under periodic pacing should reproduce modulation of ictal-like discharges in response to stimulation. To enforce these characteristics in the model, we defined the fraction of ictal-like discharges to be μ,

$$\mu = \frac{\sum_{i=1}^{|\mathcal{D}|} \mathcal{L}\left(\mathcal{D}(i)\right)}{|\mathcal{D}|}, \tag{11}$$

as the measurement, Θ, to be optimized during selection of the numerical integration parameters, η and ω.

Thus, for each $\mathcal{M}_n^{trn}$ and for each pair of parameters η and ω drawn from the set of pairs defined by the cross product of ranges $\eta = [0.0, 0.001]$ sampled at intervals 0.0001 and $\omega = [0.0, 1.0]$ sampled at intervals of 0.05, we simulated $K = 30$ surrogates of $1,800$ seconds length for each protocol, ρ, $\rho \in P = \{\text{control}, 0.2 \text{ Hz}, 0.5 \text{ Hz}, 1.0 \text{ Hz}, 2.0 \text{ Hz}\}$. We computed $\tilde{\theta}_n^{trn}[\eta, \omega]$ by averaging all K values of μ for each stimulation protocol. We computed similar measurements of the training datasets, and we solved for the parameter pair η_n and ω_n that yielded minimum error according to Eq. 10. Once found, parameters η_n and ω_n remained unchanged throughout the experiments.

3. Results

3.1. State Validation

Embedding theory implies that a single E-dimensional state of the model captures the dynamic information contained in a sequence of historical observations. In the context of modeling neural network dynamics of *in vitro* epileptiform activity under periodic pacing, the dynamic information is the state's label. If the distribution of labels in the model's state space (see Fig. 1b and c) is both accurate and generalizable, then test data projected into the model's state space will have similar structure and distribution. Adherence of the model to embedding theory, therefore, can be measured experimentally by using the model to predict the labels of previously unseen observations.

For each M_n^{trn} and $\mathcal{D}_n^{tst}$, $n \in \mathcal{N}$, we embedded dataset $\mathcal{D}_n^{tst}$ using model's E and $T_{\min}$ parameters and mapped these data into a state space using the model's projection, $V_E^{\dagger}$. For each element in this projection, we applied the $\mathcal{L}$ operation to the result of Equation 6; the resulting label is the model's prediction, which we compared against the actual labels of the elements of $\mathcal{D}_n^{tst}$.

To assess the structural quality of our embedding we computed the performance for each of the training models in classifying the corresponding test dataset. We measured performance in terms of likelihood ratios (Hogg et al., 2004) (LR_n^+) of ictal classification. Likelihood ratios measure how much the odds of a positive event change in response to a classifier's positive prediction of that event. A random classifier yields LR+=1.0, indicating that the prediction provides zero information. For an ideal classifier, LR^+ approaches

∞. Our classifier (i.e., model) achieved $LR^+ = 9.33$ (2.37–16.28 for the 95% confidence limits), indicating that a point classified as ictal is 933% more likely to actually be ictal.

3.2. Transition Function Validation

Embedding theory also implies that transitions between model states are equivalent to transitions in the underlying dynamic system. Thus, state sequences, and hence label sequences, of both real and surrogate datasets should be similar. We compared similarities between sequences of labels in the test datasets and the test surrogates at two time-scales: macroscopic (multiple ictal discharges) and mesoscopic (single ictal discharge).

To validate macroscopic temporal predictions of the model we compared the distributions of fractions of ictal labels (see Eq. 11) between the test datasets and test surrogates for each stimulation protocol. As illustrated in Fig. 2a, the model accurately captures the non-linear functional relationship between the stimulation policy and the fraction of ictal labels observed (i.e., the model accurately reproduces the degree of control of ictal-like discharges in response to stimulation). To validate mesoscopic temporal predictions, we compared the distributions of ictal-like discharge duration (Fig. 2b) and interval (Fig. 2c) between the test datasets and test surrogates for each stimulation protocol. Our results confirm that the model is faithfully capturing higher-order, non-linear functional relationships between the stimulation policy and the dynamic response of stimulated neural networks.

These results do, however, illustrate a difficulty of fairly evaluating the model and real data using cross-validation. When ictal samples are sparsely distributed in the dataset, as is the case for periodic pacing results in both the test dataset and the test surrogate, then the impact of outliers can be augmented. Only a single ictal discharge interval exists in the test dataset under the 0.2 Hz protocol. However, due to 6-fold partitioning of the dataset as part of cross validation of the *in vitro* experiments, this single ictal discharge interval appears three times attributing to a large mean-value having zero variance. A similar artifact occurs for the test surrogate at 1.0 Hz periodic pacing. Only one of the six cross validation models produced a mean ictal discharge duration. This single value, likely caused by noise in the model, dominates the mean value.

12

3.3. Generation of Surrogate Data

The spectral embedding method yields a projection, $\mathbf{V}_E$, from the model's state space to a vector of observations (given by Eq. 5). The first element of the resultant vector is the simulated observation of the field potential corresponding to the model's current state. During numerical integration of the model, each state may be projected onto such a field potential; the resulting sequence is a surrogate field potential recording. Fig. 3 illustrates examples of test surrogate field potentials alongside examples of field potential recordings for each of the five stimulation protocols.

The time correspondence between the test dataset and surrogate (which is, in general, very good) arises from structural consistencies in the dataset that are exploited by the evidence-based modeling approach. However, excellent temporal reproduction does not imply the ability to accurately predict specific trajectories, particularly, future epileptiform discharges. The presence of noise and the nonlinear (potentially chaotic) nature of this dynamical system make prediction of specific closed-loop trajectories difficult; previous work in dynamic modeling of epileptiform systems suggests that such predictions may be infeasible (Lopes da Silva et al., 2003).

Another visual aspect of the plots in Fig. 3 is the degradation of temporal detail reproduced in the surrogate as the frequency of stimulation increases. This observation stands in contrast to the agreement between the test surrogate and test dataset results depitected in Fig. 2a across stimulation frequencies. Because the spectral embedding method extracts embeddings that maximally capture dataset variance, we speculate that model fidelity is preserved in high-variance regimes of the dataset (i.e., lower-frequency stimulation and control) at the expense of low-variance regimes (i.e, higher-frequency stimulation), which is supported by the visual evidence.

4. Discussion

We proposed and validated an evidence-based model that faithfully reproduces epileptiform discharges as well as the effects of low-frequency periodic pacing on ictal-like activity generated by a brain slice preparation *in vitro*. Our use of cross-validation for reporting the model performance is a well-known machine learning technique for identifying how model predictions will generalize to previously unseen data. The technique succeeds because it approximates an unbiased estimate of the actual distribution to which the model is being fit (Kohavi, 1995).

The key physiological insight to be drawn from successful application of the evidence-based approach to this dataset is its ability to validate the existence of a low-dimensional manifold that captures canonical dynamics of a complex neural circuit governed by millions of variables. Cross-validated simulations using this manifold confirm, topologically, the robustness of neural dynamic modulation during epileptiform activity in response to fixed, low-frequency electrical stimulation that has been reported in the literature. Further, we know of no other computational modeling approach that is capable of generating surrogate data featuring: 1) spontaneous ictal-like discharges having frequency, duration, amplitude, and higher-order dynamics similar to that of real-world observations, 2) spontaneous interictal-like discharges that do not lead to ictal-like events, and 3) post-ictal depression.

The versatility of this approach, however, comes at a cost: evidence-based models are limited in the testable hypotheses that they can inform. We consider this approach most appropriate for hypotheses in which accurate prediction of causal relationships is paramount to understanding the physiological roots of these relationships.

One can easily imagine a large set of testable hypotheses that require designing and implementing neural control systems, be it automated exploration of neural dynamics or treatment of neurological diseases (Sun et al., 2008; Jahangiri et al., 1997; Schiff et al., 1994; Pineau et al., 2009; Schiff and Sauer, 2008). Currently, first-principles approaches, while unlimited in the testable hypotheses that they can inform, are unsuitable for control applications because they do not model the causal relationships in real-world neural networks with the accuracy necessary to make effective control decisions; they also provide no mapping between real-world observations and the models state, making it difficult to query the model in a specific real-world scenario.

We anticipate specific need for the evidence-based modeling approach in the fields of intelligent and adaptive control. Training algorithms for these classes of control systems rely on accurate representations of the neural system's state and its transition function to succeed. We leave the application of the model to control systems for future work; however, the evidence-based approach does solve a key challenge posited by prior neural control research using Kalman filtering (Schiff and Sauer, 2008) in that it does not assume the form of the system's dynamics *a priori*. Compared to past methods, this difference makes the evidence-based approach scalable to real-world neural control systems.

Acknowledgements

References

Avoli, M., Barbarosie, M., Lcke, A., Nagao, T., Lopantsev, V., Khling, R., 1996. Synchronous GABA-mediated potentials and epileptiform discharges in the rat limbic system *in vitro*. Journal of Neuroscience 16, 3912–3924.

Avoli, M., D'Antuono, M., Louvel, J., Khling, R., Biagini, G., Pumain, R., D'Arcangelo, G., Tancredi, V., 2002. Network and pharmacological mechanisms leading to epileptiform synchronization in the limbic system *in vitro*. Prog. Neurobiol. 68, 167–207.

Barbarosie, M., Avoli, M., 1997. CA3-driven hippocampal-entorhinal loop controls rather than sustains *in vitro* limbic seizures. The Journal of Neuroscience 17(23), 9308–9314.

Bawin, S., Abu-Assal, M., Sheppard, A., Mahoney, M., Adey, W., 1986. Long-term effects of sinusoidal electric fields in penicillin-treated rat hippocampal slices. Brain Research 399, 194–199.

Biswal, B., Dasgupta, C., 2002. Neural network model for apparent deterministic chaos in spontaneously bursting hippocampal slices. Physical Review Letters 88, 088102–1.

D'Arcangelo, G., Panuccio, G., Tancredi, V., Avoli, M., 2005. Repetitive low-frequency stimulation reduces epileptiform synchronization in limbic neuronal networks. Neurobiology of Disease 19, 119–128.

Durand, D., 1993. Ictal patterns in animal models of epilepsy. Journal of Clinical Neurophysiology 10, 181–297.

Durand, D., Warman, E., 1994. Desynchronization of epileptiform activity by extracellular current pulses in rat hippocampal slices. Journal of Physiology 71, 2033–2045.

Franaszczuk, P., Kudela, P., Bergey, G., 2003. External excitatory stimuli can terminate bursting in neural network models. Epilepsy Research 53, 65–80.

Galka, A., 2000. Topics in nonlinear time series analysis: with implications for EEG analysis. World Scientific.

Gluckman, B., Neel, E., Netoff, T., Ditto, W., Spano, M., Schiff, S., 1996. Electric field suppression of epileptiform activity in hippocampal slices. Journal of Neurophysiology 76, 4202–4205.

Hogg, R. V., Craig, A., McKean, J. W., 2004. Introduction to mathematical statistics. Prentice Hall.

Huke, J., March, 2006. Embedding nonlinear dynamical systems: a guide to Takens' Theorem. Tech. rep., Manchester Institute for Mathematical Sciences, University of Manchester.

Jahangiri, A., Durand, D., Lin, J., 1997. Singular stimulus parameters to annihilate spontaneous activity in Hodgkin-Huxley model with elevated potassium. In: Proceedings of the 19th International Conference of the IEEE Engineering in Medicine and Biology Society. pp. 165–178.

Jerger, K., Schiff, S., 1995. Periodic pacing and *in vitro* epileptic focus. Journal of Neurophysiology 73(2), 876–879.

Katayama, T., 2005. Subspace methods for system identification. Springer.

Kennel, M., Abarbanel, H., 2002. False neighbors and false strands: a reliable minimum embedding dimension algorithm. Physical Review E 66, 026209.

Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: International Joint Conference on Artificial Intelligence. pp. 1137–1145.

Lopes da Silva, F., Blanes, W., Kalitzin, S., Parra, J., Suffczynski, P., Velis, D., 2003. Dynamical diseases of brain systems: Different routes to epileptic seizures. IEEE Transactions on Biomedical Engineering 50(5), 540–548.

Marmarelis, V., 1997. Modeling methodology for nonlinear physiological systems. Annals of Biomedical Engineering 25, 239–251.

Marmarelis, V., Juusola, M., French, A., 1999. Principal dynamic mode analysis of nonlinear transduction in a spider mechanoreceptor. Annals of Biomedical Engineering 27, 391–402.

Marmarelis, V., Orme, M., 1993. Modeling of neural systems by use of neuronal modes. IEEE Transactions on Biomedical Engineering 40(11), 1149–1158.

Nakagawa, M., Durand, D., 1991. Suppression of spontaneous epileptiform activity with applied currents. Brain Research 567, 241–247.

Panuccio, G., D'Antuono, M., de Guzman, P., Lannoy, L. D., Biagini, G., Avoli., M., 2010. *In vitro* ictogenesis and parahippocampal networks in a rodent model of temporal lobe epilepsy. Neurobiol Dis.

Parlitz, U., Merkwirth, C., 2000. Prediction of spatiotemporal time series based on reconstructed local states. Physical Review Letters 84(9), 1890–1893.

Pineau, J., Guez, A., Vincent, R., Panuccio, G., Avoli, M., 2009. Treating epilepsy via adaptive neurostimulation: a reinforcement learning approach. International Journal of Neural Systems 19(4), 227–240.

Schiff, S., Jerger, K., Duong, D., Chang, T., Spano, M., Ditto, W., 1994. Controlling chaos in the brain. Nature 370, 615–620.

Schiff, S., Sauer, T., 2008. Kalman filter control of a model of spatiotemporal cortical dynamics. Journal of Neural Engineering 5, 1–8.

Sun, F., Morrell, M., Wharen, R., 2008. Responsive cortical stimulation for the treatment of epilepsy. Neurotherapeutics: The Journal of the American Society for Experimental NeuroTherapeutics 5, 68–74.

Takens, F., 1980. Detecting strange attractors in turbulence. In: Young, D. A. R. . L. S. (Ed.), Dynamical Systems and Turbulence. Vol. 898. Warwick, pp. 366–381.

Warren, R., Durand, D., 1998. Effects of applied currents on spontaneous epileptiform activity induced by low calcium in the rat hippocampus. Brain Research 806, 1078–1085.
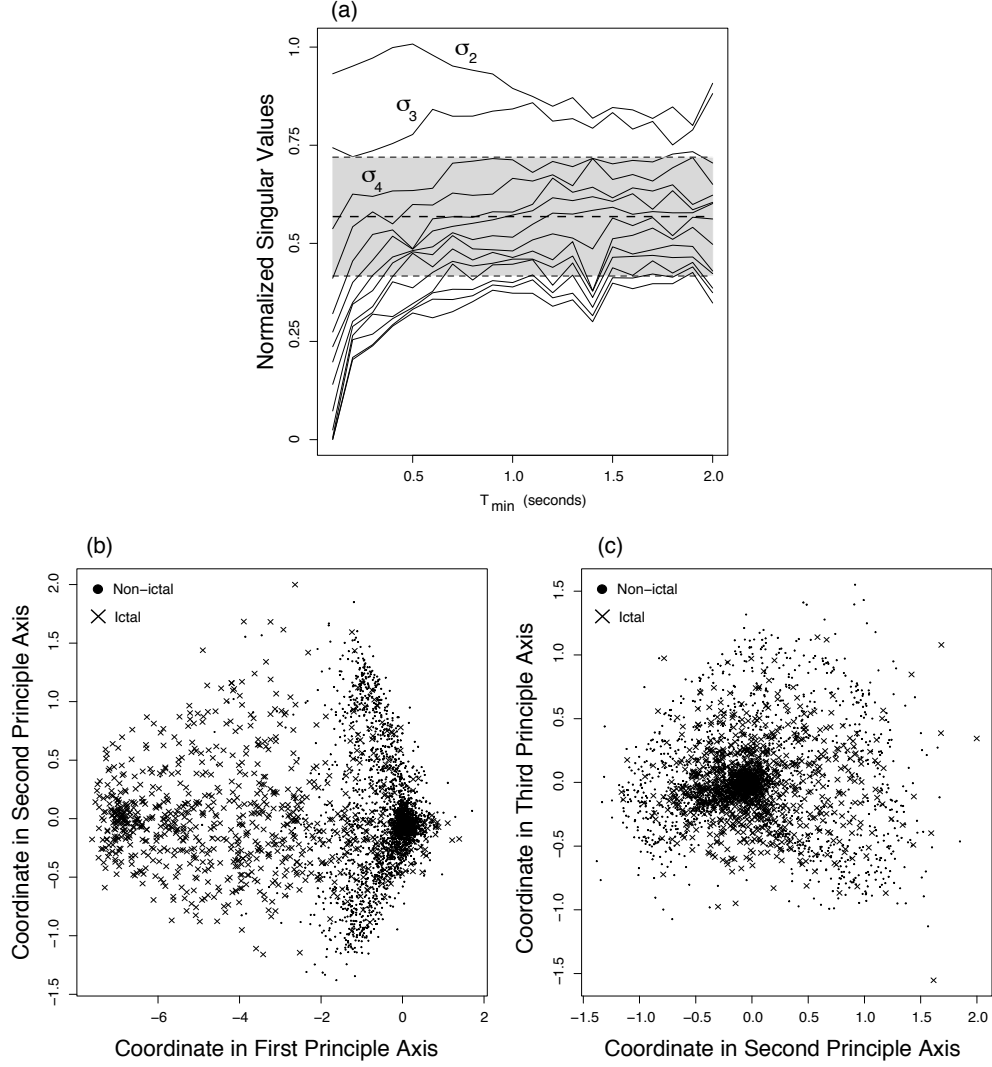
Figure 1: **(a)** Embedding spectrum assuming $E = 15$; E was determined empirically to be sufficiently large to contain all the observable dynamics (magnitudes of σ_E at greater E are indistinguishable within a small constant). Maxima σ_2 occurred at approximately $T_{\min} = 0.5$ s. The mean and standard deviation of the spectra of singular values (σ_2–σ_{15} over 0.0–2.0 s) are presented as bold dashed (mean) and dashed ($\pm$ std. dev.) horizontal lines. Singular values 1–3 fall above the first standard deviation of this trend at $T_{\min} = 0.50$, which indicates that the intrinsic dimension of the embedded system is $E = 3$. **(b)** and **(c)** are examples of a real training dataset after embedding and projection into the model's state space; the dataset is viewed along the 1st and 2nd as well as 2nd and 3rd principal axes of V_E, respectively.
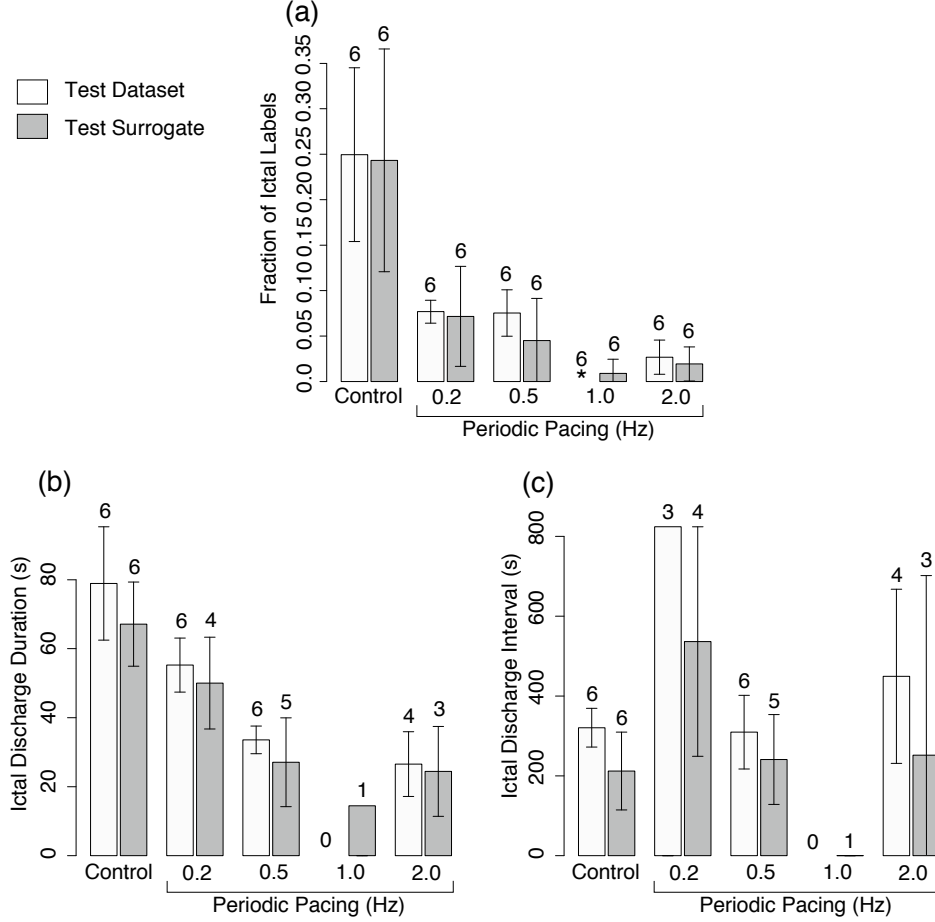
18

Figure 2: **(a)** Comparison of the fraction of ictal-like discharges observed as a function of stimulation protocol between test datasets and test surrogates. Bars summarize the means of the distributions of the cross-validation experiments and error bars indicate the 95% confidence limits. Numbers above the error bar indicate the number of cross-validation experiments comprising the distribution. Asterisk (*) indicates that the distribution has zero mean and variance and cannot be plotted. **(b)** Comparison of ictal-like discharge durations as a function of stimulation protocol between test datasets and test surrogates. **(c)** Comparison of ictal-like discharge intervals as a function of stimulation protocol between test datasets and test surrogates. To construct these comparisons we generated $K = 30$ randomly initialized surrogates of $1,800$ seconds length for each of $N = 6$ test models, $\mathcal{M}_n^{tst}$, $n \in \mathcal{N}$, for each stimulation protocol. For each surrogate we computed the fraction of ictal labels, according to Eq. 11, as well as the duration and interval of ictal-like discharges. We separated the data by stimulation protocol and averaged together the K quantities (i.e., fractions of ictal labels or ictal-like discharge durations and intervals) computed for each protocol on each test model, yielding $N = 6$ quantities for each protocol. We computed the analogous N quantities of the test datasets for each stimulation protocol.
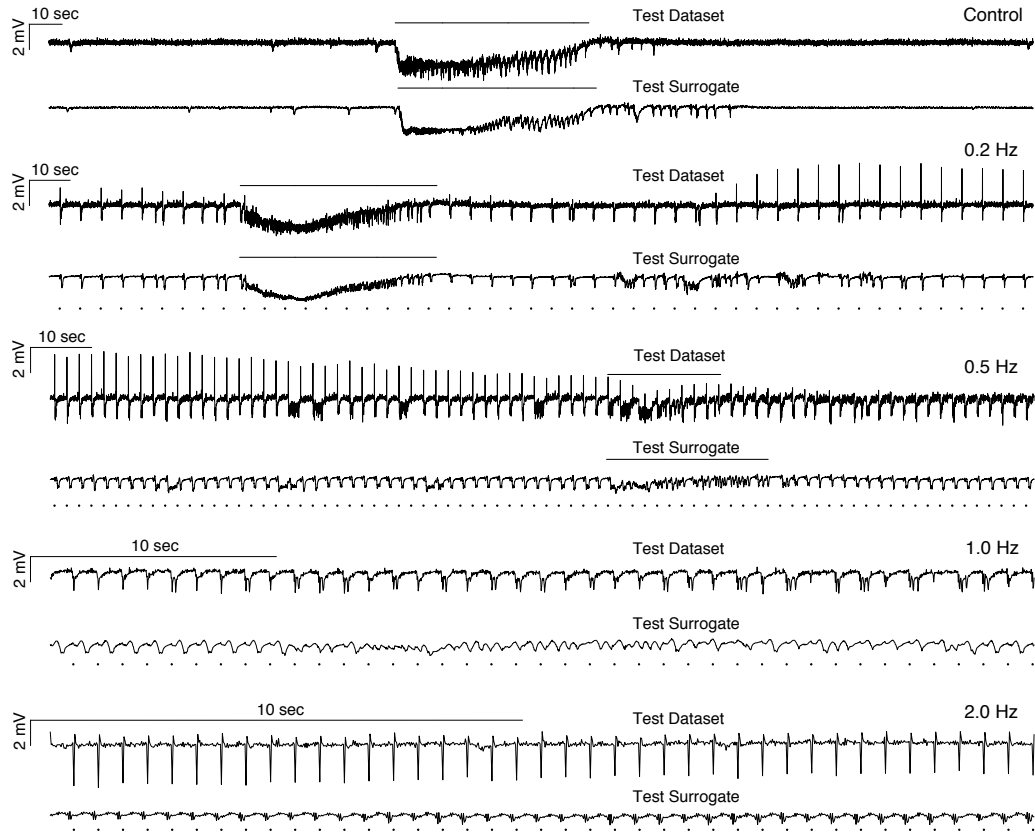
19

Figure 3: Comparison of selected segments of surrogate field potential traces versus selected segments of the test dataset that share similar dynamics. Dots indicate the timing of stimulations. Horizontal black lines indicate ictal labels. The test dataset labels were assigned by humans whereas the test surrogate labels were assigned automatically by the model. No effort was made to identify traces that are visually similar. Rather, visual similarity between the dynamic structure of the surrogates and datasets was observed to be a general attribute of the evidence-based modeling approach.