J Comput Neurosci (2011) 30:471–487
DOI 10.1007/s10827-010-0272-1

A continuous mapping of sleep states through association of EEG with a mesoscale cortical model

Beth A. Lopour · Savas Tasoglu · Heidi E. Kirsch · James W. Sleigh · Andrew J. Szeri

Received: 10 November 2009 / Revised: 7 August 2010 / Accepted: 16 August 2010 / Published online: 1 September 2010

Abstract Here we show that a mathematical model of the human sleep cycle can be used to obtain a detailed description of electroencephalogram (EEG) sleep stages, and we discuss how this analysis may aid in the prediction and prevention of seizures during sleep. The association between EEG data and the cortical model is found via locally linear embedding (LLE), a method of dimensionality reduction. We first show that LLE can distinguish between traditional sleep stages when applied to EEG data. It reliably separates REM and non-REM sleep and maps the EEG data to a low-dimensional output space where the sleep state changes smoothly over time. We also incorporate the concept of strongly connected components and use this as a method of automatic outlier rejection for EEG data. Then, by using LLE on a hybrid data set containing both sleep EEG and signals generated from the mesoscale cortical model, we quantify the relationship between the data and the mathematical model. This enables us to take any sample of sleep EEG data and associate it with a position among the continuous range of sleep states provided by the model; we can thus infer a trajectory of states as the subject sleeps. Lastly, we show that this method gives consistent results for various subjects over a full night of sleep and can be done in real time.

Keywords Sleep · Sleep staging · Sleep scoring · Seizure · Locally linear embedding · Electroencephalogram · Cortical model · Mesoscale · Mean-field

Action Editor: Gaute T. Einevoll

B. A. Lopour · S. Tasoglu · A. J. Szeri (✉)
Department of Mechanical Engineering,
University of California, Berkeley,
CA 94720, USA
e-mail: Andrew.Szeri@berkeley.edu

B. A. Lopour
e-mail: bethlopour@berkeley.edu

H. E. Kirsch
Department of Neurology, University of California,
San Francisco, CA 94143, USA

J. W. Sleigh
Department of Anaesthetics, Waikato Hospital,
Hamilton, New Zealand

1 Introduction

The standard method of sleep scoring involves categorization of electroencephalogram (EEG) data into five separate stages (Niedermeyer and da Silva 2005). However, the *discrete* nature of these stages limits their utility as analytical and predictive tools. For example, in a study of human epilepsy, it may be observed that a seizure occurred during stage 2 sleep. This prompts further questions: Was the subject descending to deeper stages of sleep or arising from them? How quickly was the subject moving through each stage? Was a transition imminent when the seizure occurred?

The use of a mathematical model of the human sleep cycle may allow us to answer such questions by providing a *continuous* spectrum of sleep states, ranging from REM to the deepest slow-wave sleep. If the model can be directly associated with human sleep EEG data, it will be possible to track the subject's state to identify

the stage as well as changes in sleep depth and proximity to transitions. Ideally, this would be done in real-time, where the state is continuously determined as the subject sleeps. The process must be consistent over various subjects and robust to non-standard sleep cycles and periods of waking.

Here we utilize a technique called locally linear embedding (LLE) to make this connection between a model of the human sleep cycle and EEG data. First, we present a model of the human cortex with subcortical inputs represented by added driven noise, and we describe the associated mathematical representation of the sleep cycle (Section 2). We then introduce the technique of locally linear embedding (Section 3) and show that it provides the ability to distinguish between sleep stages when applied to EEG data (Section 4). These results demonstrate reliable separation between REM and NREM sleep data and provide a smooth temporal progression through the various stages of sleep. We also present the concept of strongly connected components as a method of outlier rejection for EEG data (Section 3.2) and introduce a method for automatic selection of LLE parameters (Section 4.3). Then, by performing LLE on a hybrid data set containing both sleep EEG and signals generated from the mathematical model, we are able to integrate the EEG and the model (Section 5). This allows us to take any sample of sleep EEG data and determine its position within the continuous range of sleep states provided by the model. We show that this method provides consistent results for various subjects over a full night of sleep, and it could be done online as the subject sleeps.

2 Mean-field cortical model

2.1 Background and mathematics

Mean-field models of the cortex are well-suited to the study of brain states described by EEG signals, including sleep. The variables in these models, representing quantities that are averaged over the millimeter scale, are comparable to the mesoscale measurements of EEG electrodes. More specifically, we choose a cortical model developed most recently in Liley et al. (2002) and Steyn-Ross et al. (1999, 2003). In addition to sleep, it has been used to model epileptic seizures (Kramer et al. 2005), anesthesia (Steyn-Ross et al. 2004; Bojak and Liley 2005), and the transition to seizure due to application of anesthetic agents (Liley and Bojak 2005).

Here, we use the dimensionless formulation of the model as described in Kramer et al. (2007), with two parameters Δh_e^{rest} and L added to represent neuromodulators that regulate the natural sleep cycle, as was done in Steyn-Ross et al. (2005):

$$\frac{\partial \tilde{h}_e}{\partial \tilde{t}} = 1 - \tilde{h}_e + \frac{\Delta h_e^{rest}}{h^{rest}} + L\,\Gamma_e\left(h_e^0 - \tilde{h}_e\right)\tilde{I}_{ee} + \Gamma_i\left(h_i^0 - \tilde{h}_e\right)\tilde{I}_{ie}\,, \tag{1}$$

$$\frac{\partial \tilde{h}_i}{\partial \tilde{t}} = 1 - \tilde{h}_i + L\,\Gamma_e\left(h_e^0 - \tilde{h}_i\right)\tilde{I}_{ei} + \Gamma_i\left(h_i^0 - \tilde{h}_i\right)\tilde{I}_{ii}\,, \tag{2}$$

$$\left(\frac{1}{T_e}\frac{\partial}{\partial \tilde{t}} + 1\right)^2 \tilde{I}_{ee} = N_e^\beta \tilde{S}_e\left[\tilde{h}_e\right] + \tilde{\phi}_e + P_{ee} + \tilde{\Gamma}_1\,, \tag{3}$$

$$\left(\frac{1}{T_e}\frac{\partial}{\partial \tilde{t}} + 1\right)^2 \tilde{I}_{ei} = N_e^\beta \tilde{S}_e\left[\tilde{h}_e\right] + \tilde{\phi}_i + P_{ei} + \tilde{\Gamma}_2\,, \tag{4}$$

$$\left(\frac{1}{T_i}\frac{\partial}{\partial \tilde{t}} + 1\right)^2 \tilde{I}_{ie} = N_i^\beta \tilde{S}_i\left[\tilde{h}_i\right] + P_{ie} + \tilde{\Gamma}_3\,, \tag{5}$$

$$\left(\frac{1}{T_i}\frac{\partial}{\partial \tilde{t}} + 1\right)^2 \tilde{I}_{ii} = N_i^\beta \tilde{S}_i\left[\tilde{h}_i\right] + P_{ii} + \tilde{\Gamma}_4\,, \tag{6}$$

$$\left(\frac{1}{\lambda_e}\frac{\partial}{\partial \tilde{t}} + 1\right)^2 \tilde{\phi}_e = \frac{1}{\lambda_e^2}\frac{\partial^2 \tilde{\phi}_e}{\partial \tilde{x}^2} + \left(\frac{1}{\lambda_e}\frac{\partial}{\partial \tilde{t}} + 1\right) N_e^\alpha \tilde{S}_e\left[\tilde{h}_e\right]\,, \tag{7}$$

$$\left(\frac{1}{\lambda_i}\frac{\partial}{\partial \tilde{t}} + 1\right)^2 \tilde{\phi}_i = \frac{1}{\lambda_i^2}\frac{\partial^2 \tilde{\phi}_i}{\partial \tilde{x}^2} + \left(\frac{1}{\lambda_i}\frac{\partial}{\partial \tilde{t}} + 1\right) N_i^\alpha \tilde{S}_e\left[\tilde{h}_e\right]\,. \tag{8}$$

The model contains two groups of equations: one that describes the evolution of the excitatory population (Eqs. (1), (3), (5), (7)) and one that governs the inhibitory population (Eqs. (2), (4), (6), (8)). Each variable is a function of dimensionless space ($\tilde{x}$) and time ($\tilde{t}$), and the subscript denotes its association with the excitatory or inhibitory population. For example, in the excitatory population, the mean soma potential is represented by $\tilde{h}_e$, while $\tilde{I}_{ie}$ is the input current from population i to population e. The synaptic currents are functions of local input, e.g. $N_e^\beta \tilde{S}_e$ where

$$\tilde{S}_e\left[\tilde{h}_e\right] = \frac{1}{1 + \exp\left[-\tilde{g}_e\left(\tilde{h}_e - \tilde{\theta}_e\right)\right]}\,; \tag{9}$$

this function converts the potential of the excitatory population into a mean firing rate. Synaptic currents are also affected by long-range corticocortical input $\tilde{\phi}_e$ and subcortical stochastic inputs such as $\tilde{\Gamma}_1$, which we define to be a function of zero-mean Gaussian white noise ξ_1:

$$\tilde{\Gamma}_1 = \alpha_{ee}\sqrt{P_{ee}}\,\xi_1\left[\tilde{x}, \tilde{t}\right]\,. \tag{10}$$

Here α_{ee} is a constant that determines the variance of the stochastic input. Please refer to Table 1 for further descriptions of all variables and parameters.

For completeness, we will include the full model in our simulations; however, it should be noted that a reduced version would suffice in this case. For example, we will utilize only the temporal evolution of variables, so it would be possible to convert the model to a system of ODEs by removing the spatial derivatives from Eqs. (7) and (8). In addition, the subdivision of local excitatory inputs, represented by $N_e^{\beta}\tilde{S}_e[\tilde{h}_e]$ in Eqs. (3) and (4), is unnecessary. Making these changes would perhaps reduce the computation time for numerical solutions to the model, but we would not expect them to affect the results.

For the purpose of modeling sleep, we will focus on the parameters L and Δh_e^{rest} and the variable $\tilde{h}_e$. The parameters represent the actions of neuromodulators adenosine and acetylcholine (ACh) that aid in the regulation of the human sleep cycle. Adenosine reflects the activity of the homeostatic drive to sleep, which is modulated by various somnogens. The ACh input into the cortex is a measure of the activity of the various brain stem controllers of sleep. Note that we have not specifically modeled the complex intrinsic interactions between the various brain stem nuclei. In this paper, we are primarily concerned with the interaction of their neuromodulator output with the cerebral cortex and thus model their effects only as extrinsic alterations in ACh.

In general, adenosine acts to reduce the resting potential of excitatory cells, thus making them less likely to fire; ACh does the opposite by raising the resting potential. These changes are represented in the model by Δh_e^{rest}, which adds directly to the resting potential of the excitatory population (disguised as a "1" in the dimensionless equations). In addition, ACh decreases the amplitude of the excitatory postsynaptic potential, effectively reducing the synaptic gain. In the model, this corresponds to a reduction in the effect of synaptic currents $\tilde{I}_{ee}$ and $\tilde{I}_{ei}$; therefore, the parameter L is multiplied by these quantities to simulate a change in synaptic gain. Lastly, as was done in Steyn-Ross et al. (2005), we take the mean excitatory soma potential $\tilde{h}_e$ to be representative of cortical activity; we will compare this variable to EEG measurements using locally linear embedding.

2.2 Model of the human sleep cycle

The mechanisms underlying human sleep and waking are complex; for recent, detailed reviews of the brain stem and hypothalamic control of sleep in thalamocortical systems see Fuller et al. (2006, 2007), McCarley (2007), Rosenwasser (2009), Saper et al. (2005a, b). In summary, the wakeful state may be characterized by high levels of activity in aminergic, cholinergic, orexinergic and glutamatergic neuronal populations in the brain stem and hypothalamus. The overall effect is to maintain the thalamo-cortical neurons in a depolarized,

Table 1 Dimensionless variables and parameters of the SPDE cortical model

Symbol	Description	Typical value
$\tilde{h}_{e,i}$	Spatially averaged soma potential for neuron populations	–
$\tilde{I}_{ee,ei}$	Postsynaptic activation due to excitatory inputs	–
$\tilde{I}_{ie,ii}$	Postsynaptic activation due to inhibitory inputs	–
$\tilde{\phi}_{e,i}$	Long-range (corticocortical) input to e and i populations	–
$\tilde{t}$	Time (dimensionless)	–
$\tilde{x}$	Space (dimensionless)	–
$\Gamma_{e,i}$	Influence of synaptic input on mean soma potential	4.6875×10^{-4}, 0.0105
$h^0_{e,i}$	Reversal potential	0, 1.0938
$T_{e,i}$	Neurotransmitter rate constant	12.0, 3.6
$\lambda_{e,i}$	Inverse length scale for corticocortical connections	11.2, 11.2
$P_{ee,ei}$	Subcortical input from excitatory population	25.0, 25.0
$P_{ie,ii}$	Subcortical input from inhibitory population	25.0, 25.0
$N^{\alpha}_{e,i}$	Number of distant (corticocortical) connections from excitatory populations to e and i populations	3710, 3710
$N^{\beta}_{e,i}$	Number of local synaptic connections from e and i populations	410, 800
$\tilde{g}_{e,i}$	Slope at inflection point of sigmoid function $\tilde{S}_e$	−29.021, −19.347
$\tilde{\theta}_{e,i}$	Inflection point for sigmoid function $\tilde{S}_e$	0.91406, 0.91406

Values for the dimensional parameters were taken from Wilson et al. (2006), with the exception of γ_i which was chosen to be $90s^{-1}$. The dimensionless parameters were then calculated as described in Kramer et al. (2007)

active, and continually firing state. These excitatory neurons also inhibit activity in various gamma-amino-butyric-acid(GABA)ergic cell populations, particularly in the ventro-lateral pre-optic area (VLPO), basal forebrain, and in the reticular nucleus of the thalamus. With the build up of homeostatic and circadian pressure to sleep (possibly mediated by various activity-dependent somnogens such as adenosine), the wake-active neurons are inhibited, which then allows the sleep-promoting neurons of the VLPO to start firing and trigger the transition from wakefulness to NREM sleep. This results in quiescence of the aminergic, orexinergic, and cholinergic brain-stem neuromodulator centers; which in turn allows hyperpolarization of the cortico-thalamic systems and hence the burst firing patterns characteristic of slow wave sleep. If these neurons are only moderately hyperpolarized, the EEG is dominated by the sleep spindles and K-complexes characteristic of stage 2 sleep. With more profound hyperpolarization the EEG is dominated by the delta waves of stages 3 and 4 (Steriade and Amzica 1998; Steriade and Timofeev 2001). This progressive slowing of the dominant frequency is captured by measures such as the permutation entropy index (Olofsen et al. 2008). The transition from NREM to REM sleep is associated with cortico-thalamic depolarization caused by activation of cholinergic and glutamatergic brain stem systems (mainly in and near the pedunculo-pontine tegmentum). The neuromodulatory environment of REM sleep differs from the wakeful state in that the amines and orexinergic systems are inactive in REM sleep, but active in the wakefulness; however, this distinction is not explicit in the present model.

Mathematically, we follow Steyn-Ross et al. (2005), where the representation of the sleep cycle is based on changes in neuromodulators L and Δh_e^{rest}. In order to visualize this, we look at steady-state solutions of h_e (without stochastic input) as L and Δh_e^{rest} are varied; these solutions create what we will refer to as the "sleep manifold" (Fig. 1). Notice that, for most parameter values, there is only one steady state solution. However, in certain cases, there are three solutions (two stable and one unstable), causing the manifold to fold over on itself. This fold is seen on the left side of Fig. 1. In this model, the top branch of solutions on the manifold is intended to be representative of REM sleep. Starting at this point, we can imagine that during sleep, there is a gradual descent to deep slow-wave sleep by following a trajectory down the right side of the manifold where there is only one steady state solution. This happens in a smooth continuous manner. Then the quick transition from slow-wave sleep to REM is simulated by a jump across the fold from the bottom branch

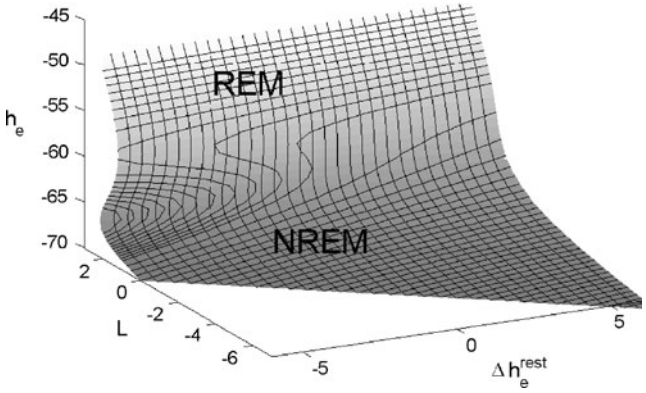

Fig. 1 The manifold of steady states in h_e from the mesoscale cortical model, hereafter referred to as the "sleep manifold." The parameters L and Δh_e^{rest} represent the actions of adenosine and acetylcholine, neuromodulators that vary over the course of the human sleep cycle. The manifold has two stable solutions on its left side; a jump from the bottom solution to the top solution represents the fast transition between NREM and REM sleep. However, the slow transition from REM to NREM occurs smoothly down the right side of the manifold, where there is only one set of solutions

of solutions to the top branch. This mimics the rapid transition from deep sleep to REM that is observed in human EEG recordings. This process of gradually moving from REM to NREM sleep and then quickly jumping back to REM represents one sleep cycle in the model.

This model has been previously studied. Steyn-Ross et al. (2005) calculated the EEG total power, fractions of high and low power, and correlation time exhibited by the model at the transition from slow-wave sleep to REM; it was found that they qualitatively matched both human clinical sleep recordings and cortical measurements from a cat. The model was also studied in two spatial dimensions to investigate stable oscillatory states similar to slow-wave sleep, and it was shown that a transition from one state to another can occur due to stochastic fluctuations (Wilson et al. 2005). Lastly, Wilson et al. (2006) interpreted the k-complex as a transient shift from a stable low-firing state to an unstable high-firing state and used this model to demonstrate the mechanism by which the transition may occur.

Because we are interested in comparing this model directly to human EEG recordings, we will use the sleep manifold as a way to generate model "EEG-like" signals. We will choose values of L and Δh_e^{rest}, find the numerical solution of the model for a given length of time, convert the dimensionless $\tilde{h}_e$ to mV, and downsample it to match the EEG recordings. By doing this for many different values of L and Δh_e^{rest} we can obtain representative signals of every sleep stage.

It has previously been argued that $\tilde{h}_e$ cannot be directly compared to measurements from cortical surface or scalp electrodes because those measurements are based on extracellular current flow, as opposed to the soma potential. This is important for the modeling of certain cortical phenomena; for example, in performing simulations of feedback control for the suppression of epileptic seizures, the value of the electrode measurement is fed directly back to the cortex to affect $\tilde{h}_e$, with little or no time delay (Lopour and Szeri 2010). In that case, the relationship between $\tilde{h}_e$ and the electrode measurement at any given time is very important. However, in the present analysis of EEG data using LLE, we are only interested in matching *scaled features* of the data that are calculated over 30-second intervals. We will not attempt to compare the temporal progression of $\tilde{h}_e$ directly to the EEG data. The previous work mentioned above has demonstrated a correspondence between $\tilde{h}_e$ and sleep EEG data with regard to these general features, so we feel confident in using it for our analysis without the addition of a scalp electrode model.

3 Locally linear embedding (LLE)

Locally linear embedding is a method of nonlinear dimensionality reduction that was originally introduced in Roweis and Saul (2000). It is useful for visualizing high-dimensional data sets as they would be embedded in a low-dimensional space, and it can often uncover relationships and patterns that are masked by the complexity of the original data set. It has been used to obtain maps of facial expressions and classify handwritten digits (Saul et al. 2003), as well as discriminate between normal and pre-seizure EEG measurements (Ataee et al. 2007). Here we will use LLE to characterize sleep EEG data and the numerical solutions of the cortical model. By embedding both in a two-dimensional space, we will be able to associate traditional EEG sleep stages with the continuous spectrum of states provided by the model.

3.1 The algorithm

Let us begin with a high-dimensional data set stored in a matrix $\mathbf{X}$ of size $D \times N$, where each column $\mathbf{X}_i$ represents one of the N D-dimensional data points. Then the LLE algorithm consists of three steps:

1. *Calculate the nearest neighbors of each data point* $\mathbf{X}_i$ *in the D-dimensional space.* This can be done in several ways; for example, we might choose the k closest points based on Euclidian distance, or we may choose only the points within a sphere of a given radius.
2. *Determine the best reconstruction of each point using only its nearest neighbors.* Mathematically, this takes the form of a least squares minimization problem:

$$\min_W \sum_{i=1}^{N} \left| \mathbf{X}_i - \sum_{j=1}^{k} W_{ij} \mathbf{X}_j \right|^2 , \tag{11}$$

   where k represents the number of nearest neighbors. Our goal is to choose the weights W that best reconstruct the original data points in the D-dimensional space, based on the criteria of least-squared error. Because we use only the nearest neighbors, we must have $W_{ij} = 0$ if $\mathbf{X}_j$ is not a neighbor of $\mathbf{X}_i$. In addition, we guarantee invariance to translations by enforcing $\sum_j W_{ij} = 1$. Note that the minimization can be calculated individually for every i.
3. *Compute the low-dimensional output vectors* $\mathbf{Y}_i$. These are chosen to provide the best global reconstruction using the weights W from the previous step. Again, this can be formulated as a least squares minimization:

$$\min_{\mathbf{Y}} \sum_{i=1}^{N} \left| \mathbf{Y}_i - \sum_{j=1}^{k} W_{ij} \mathbf{Y}_j \right|^2 . \tag{12}$$

   Here we are making the assumption that the weights that give the best reconstruction in D dimensions will also be the optimal weights in the lower-dimensional space. In this case, the N minimization problems are coupled by the elements of $\mathbf{Y}$, so they must be solved simultaneously.

A detailed description of the algorithm and several examples are provided in Saul et al. (2003). In addition, a Matlab implementation of LLE is available on the authors' website (Roweis and Saul 2009); it was used to generate all results presented here.

As a simple example, consider using LLE on a known 3D manifold (Fig. 2). In this toy example, the underlying manifold is known (although normally this would not be the case), and we recognize that it has only two dimensions, despite living in 3-dimensional space as shown in Fig. 2(a). The data set $\mathbf{X}$ consists of a random sampling of points from the manifold (Fig. 2(b)), and the LLE output for a reduction to two dimensions is displayed in Fig. 2(c). Here we see that LLE successfully unravels the manifold and uncovers its true 2D nature.

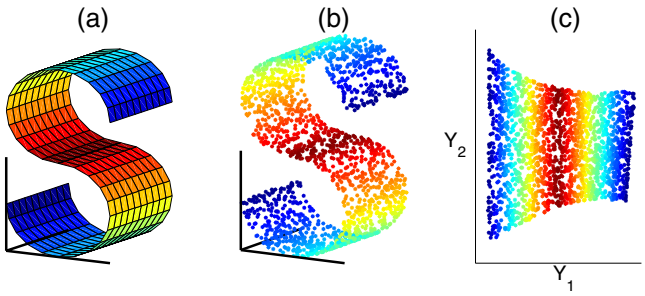

Fig. 2 A simple example of LLE, where three dimensions are reduced to two. (**a**) The underlying manifold, which lives in 3D space but has only two dimensions. In a typical LLE problem, the shape of this manifold is unknown and has too many dimensions to visualize easily. (**b**) A sampling of points from the manifold, which serves as the input to the LLE algorithm. (**c**) The result of applying LLE to the data set in (b). Note that in the $Y_1 - Y_2$ output space, the manifold has been flattened to reveal its two principal dimensions. This figure was generated using the "scurve.m" code from the LLE website (Roweis and Saul 2009)

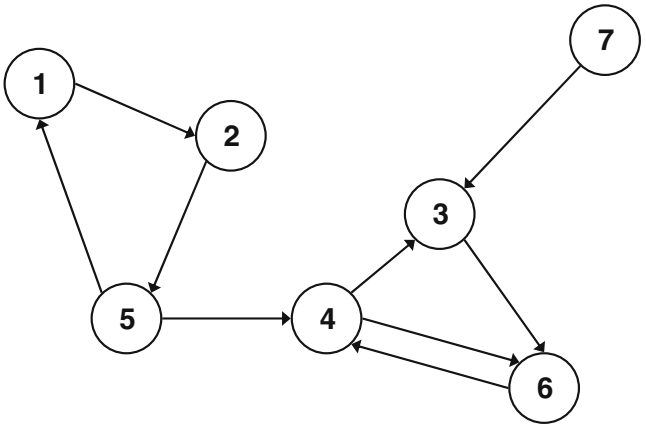

Fig. 3 An example of a directed graph generated by nearest neighbor associations. Here point 2 is a neighbor of point 1, point 3 is a neighbor of point 7, etc. In this case, the directed graph forms two strongly connected components: points 1, 2, 5, and points 3, 4, 6. In analyzing this data set, we would use LLE separately on each of these components and would remove point 7, which is not strongly connected to any other point

A possible source of confusion with locally linear embedding is the interpretation of output dimensions such as Y_1 and Y_2. Unlike linear methods such as principal component analysis, LLE does not provide a description of the output vectors in terms of the original D dimensions. The elements of $\mathbf{Y}$ are chosen to give the best *local* reconstructions based on a global minimization problem; this means that the interpretation of $\mathbf{Y}$ is different for every data point, and it cannot be described by a simple combination of the original dimensions.

3.2 Strongly connected components

The use of the LLE algorithm is based on the assumption that the entire data set lies on the same manifold in high-dimensional space. If more than one manifold is present, the locally linear reconstructions will no longer be accurate (imagine, for example, a point with nearest neighbors located on two separate manifolds). Therefore, before using LLE on a data set, we must verify this assumption.

The mathematics and terminology of directed graphs allows us to accomplish this task (Tarjan 1972). Note that when we calculate the nearest neighbors in the first step of the LLE algorithm, we create a directed graph based on the data points. For example, suppose there is a data set of seven points, and we have determined that point 2 is a neighbor of point 1, point 5 is a neighbor of point 2, etc. This can be depicted by arrows drawn from each point to its neighbors (Fig. 3). Then we can define a *strongly connected component* as a group of points where the arrows created by nearest neighbor associations allow for travel from every point in the group to every other point in the group (Tarjan 1972). *When a group of data points is strongly connected, this indicates that they lie on the same manifold* (Saul et al. 2003).

The example in Fig. 3 has two strongly connected components: points 1, 2, 5, and points 3, 4, 6. However, the two groups are not strongly connected together; one can move from the first group to the second through the connection between 4 and 5, but there is no way to get from the second group to the first. Point 7 is not strongly connected to any other point. Therefore, to use LLE on this sample data set, we would remove point 7 and use the algorithm separately on each strongly connected component.

There are several ways to identify the strongly connected components of a data set. The most traditional method involves an algorithm based on depth-first search of the directed graph (Tarjan 1972). An alternative method relies on analysis of the eigenspace that results from the LLE calculations (Polito and Perona 2001). It is also true that choosing the nearest neighbors in a different manner or increasing the value of k can change the structure of the strongly connected components. However, for the purposes of this study, we used the MATLAB function `dmperm` on a matrix containing the nearest neighbor associations for the data set. This function, based on the Dulmage-Mendelsohn decomposition, permutes the rows and columns of a matrix to put it into block diagonal form; by including the fact that every point is a neighbor with itself, we can guarantee that this permutation will be symmetric.

As output, `dmperm` provides the new order of rows and columns and identifies the blocks of the permuted matrix, where each block represents one strongly connected component within the data.

A remark about principal component analysis (PCA) is in order. This is perhaps the most common mode of dimensionality reduction, and it has also been used in the analysis of sleep EEG data (Gervasoni et al. 2004; Jobert et al. 1994; Corsi-Cabrera et al. 2000). However, PCA places the greatest importance on the directions of largest variance and relies on the assumption that the data is best reconstructed by a *linear* combination of the original measurements. While we tried PCA and achieved reasonable results, the nonlinear nature of the sleep manifold suggests that a more sophisticated solution is necessary. In addition, the concept of nearest neighbors on which the LLE algorithm is based enabled improvement in the separation of different sleep stages (see Section 4.4), and it played a crucial role in defining the quantitative relationship between the EEG data and mathematical model, as is discussed in Section 5.2.

4 LLE applied to sleep EEG data

Before examining the connection between EEG data and the mathematical model of the sleep cycle, we will first discuss the results of applying LLE to sleep EEG only. After introducing the data sets and our methods, we show that LLE can separate EEG data by sleep stage and provide a continuous representation of sleep depth.

4.1 Sleep EEG data

The EEG data used for this analysis was obtained from the Sleep-EDF database (Kemp 2009), which is part of the PhysioBank online resource of physiologic signals for biomedical research (Goldberger et al. 2000). We used four data sets (sc4002e0, sc4012e0, sc4102e0, and sc4112e0), each one consisting of a European data format (EDF) file and a file containing the hypnogram data. They were converted to ASCII format and then imported into Matlab.

The data were gathered in 1989 from healthy males and females between the ages of 21 and 35. Recordings were obtained over the course of one full day and include horizontal electrooculogram (EOG), two channels of EEG (Fpz-Cz and Pz-Oz sampled at 100 Hz), submental-electromyogram (EMG) envelope, oronasal airflow, and rectal body temperature. However, we used only the data from the Fpz-Cz EEG electrode pair in our analysis. The hypnogram data was generated via manual scoring according to Rechtschaffen & Kales using the two channels of EEG. For more details on the subjects, recording methods, and sleep staging, please see the full description in Mourtazaev et al. (1995).

4.2 LLE input based on EEG features

In order to use the EEG as an input to the LLE function, we need to define our high-dimensional data set. We do this by dividing the signal into non-overlapping windows and calculating both statistical and frequency-based *features* for each one. Therefore each window becomes one high-dimensional data point, where the dimension equals the number of features. Because the data was scored using 30-second epochs, this was a natural choice for the window length. Thus, if we have 100 minutes of EEG data and we calculate six features, we will input 200 six-dimensional points into LLE and seek the embedding in two dimensions.

We start with a pool of 17 features and use various subsets to perform the LLE analysis. An algorithm for the automated choice of feature combinations is discussed in Section 4.3. The 17 features are as follows:

Power in different frequency bands This group of five features consists of total power in the delta (up to 4 Hz), theta (4–7.5 Hz), alpha (7.5–12 Hz), beta (12–26 Hz), and gamma ranges (above 26 Hz).
Total power This is the total power in all five frequency bands.
Statistical measures These include variance, skewness, and kurtosis. Whereas the variance captures the spread of the data and is always positive, skewness is a measure of the asymmetry around the sample mean, i.e. negative skewness indicates that more data points lie below the mean than above. Kurtosis is a measure of how prone the distribution is to outliers; a signal with high kurtosis has infrequent large deviations from the mean.
Spindle score The spindle detector identifies segments of the EEG signal where the difference between consecutive points changes from positive to negative five times in a row, thereby creating two peaks and two troughs. The lag parameter τ_L defines the number of sample points spanned by each rise or fall within the sought-for spindle, so it can be adjusted to search for these motifs at lower frequencies. We set $\tau_L = 5$, which allows for detection of 8–12 Hz spindles in data sampled at 100 Hz (with the maximum response occurring for spindles at 10 Hz), and we used a minimum threshold of zero. The overall spindle score indicates the percentage of the

signal that was classified as spindle activity. Matlab code for this function is provided in McKay et al. (2010).

Permutation entropy Similar to the spindle score, the permutation entropy (PE) identifies motifs in the EEG data, such as peaks, troughs, and slopes. The PE has its maximum value when there is an equal distribution of all motifs and its minimum value when only a single motif is present. In this way, it is a measure of the "flatness" or "uncertainty" of the signal. Here we use the composite permutation entropy index (CPEI), which combines the PE with $\tau_L = 1$ and $\tau_L = 2$ with a minimum threshold level. In our study, we set the threshold at 1% of the interquartile range of the EEG data. Further descriptions of this measure and an associated MATLAB function can be found in Olofsen et al. (2008).

The CPEI has been found to be a good measure of anesthetic depth, and the motif-based methods used for permutation entropy and spindle detection are generally robust to noise. This is demonstrated in Olofsen et al. (2008), where the CPEI is calculated for both a time-varying signal and the same signal with added white noise of various magnitudes. As mentioned above, the noise threshold for PE is built into the calculation. These reasons (and the availability of published MATLAB code) led us to choose motif-based methods over more common parametric measures.

Properties of log power These four features are based on the log of the power spectral density (PSD), as obtained by Welch's method. First, we omit the delta and alpha peaks and calculate the slope and offset of a linear fit. We then determine the maximum value of the PSD above the linear estimate in the alpha range (8–17 Hz) and the maximum value of the PSD in the delta range (0.5–4 Hz). These values will generally be large when a prominent peak is present. The code for generating these features was based on a Matlab function found in Leslie et al. (2009).

Power fractions The low power fraction is obtained by summing the power in the delta and theta ranges and dividing by total power. Similarly, the high power fraction is calculated by summation of the power in the beta and gamma ranges and dividing by total power.

After the initial calculation, each feature was divided by its root mean square (RMS) value.

The selection of a subset of features from this list may seem like a difficult task. It is certainly an important one—the use of all 17 features or a "nonsensical" subset will give poor results. However, it is worth noting that there are *many* combinations that result in a satisfactory separation between sleep stages in the LLE embedding. While each one may be slightly different, there will be a large number of high quality with respect to discrimination.

4.3 Automated ranking of feature sets

When we apply LLE to the EEG data, there are essentially only three choices that we must make:

1. *How many nearest neighbors should we include?* In other words, what is the value of k? The LLE embedding will be stable over a range of values; we generally expect that k will be greater than the number of output dimensions and smaller than the original number of dimensions D (Saul et al. 2003).
2. *What should be the dimensionality of the LLE output space?* A nice property of the LLE algorithm is that each dimension is preserved as additional dimensions are added. Therefore, if we look at the results in two dimensions and do not achieve the desired mapping, we can add a third dimension without affecting the first two.
3. *Which combination of features should we use?* Employing all 17 features in our LLE analysis does not guarantee good separation between sleep stages because some of the features may not show consistent variation as the sleep depth changes. In addition, some features, such as the variance and the power in the delta band, show similar trends; we may achieve better results by eliminating these redundancies.

In this section, we focus on the last of these questions.

While we were able to identify many effective feature combinations through educated guesswork, we wanted to evaluate the utility of LLE as a method of sleep staging by identifying the best possible results. In this case, the "best" results are those that provide a large separation between sleep stages, especially between REM and deep slow-wave sleep. Because testing each combination of the features is an onerous task, e.g. choosing six features from a pool of 17 results in 12376 combinations, we developed an algorithm to evaluate the results automatically. It first identifies two groups of points: those marked as REM in the hypnogram and those determined to be stage 4. It then tracks two parameters based on the separation between those two groups of data points as they are embedded in the LLE output space.

We first measure the percent separation between REM and stage 4, calculated as

$$a_i = 100 \cdot erf\left(\sqrt{0.5}\,\frac{\mu_4 - \mu_{REM}}{\sigma_4 + \sigma_{REM}}\right), \quad (13)$$

where μ and σ are the mean and standard deviation, respectively. This is based on the assumption that the best separation occurs when the distance between the means is large and the total standard deviation is small. We perform this calculation in both the Y_1 and Y_2 directions and combine those measurements using the 2-norm to obtain the first parameter:

$$A = \sqrt{a_{Y_1}^2 + a_{Y_2}^2}\,. \quad (14)$$

The second parameter B uses the concept of nearest neighbors to evaluate separation; for example, if the stage 4 data points have only other stage 4 points as nearest neighbors, then we can infer that they are completely separated from the other sleep stages. More specifically, it measures the number of stage 4 points with REM points as nearest neighbors and divides that by the total number of stage 4 points. If the stage 4 group is isolated, we will have $B = 0$.

We determined the values of A and B for all possible combinations of six features. There were 267 feature sets where A exceeded a threshold of 90% separation in each direction: $A > \sqrt{90^2 + 90^2}$. We then identified the 267 feature sets with the lowest values of B. By finding the combinations that were common to both groups, we identified the 11 best feature sets. Visual inspection of the LLE results for these combinations confirmed the desired separation between REM and stage 4 sleep. Note that all 11 of these feature combinations provided results with $B = 0$.

4.4 Separation of sleep stages via LLE

Having described the EEG data set using frequency-based and statistical measures and having identified the most effective subsets of those features, we are now ready to apply the LLE algorithm. As a representative result, we choose one of the 11 feature sets from the previous section; the six features are power in the delta and theta bands, variance, spindle score, maximum height of the PSD above a linear estimate in the alpha band, and high power fraction. These are plotted in Fig. 4 for 178 epochs from the sc4002e0 data set. The corresponding hypnogram is included for reference. Note that the features were calculated in 30-second non-overlapping windows to match the sleep scoring of the hypnogram.

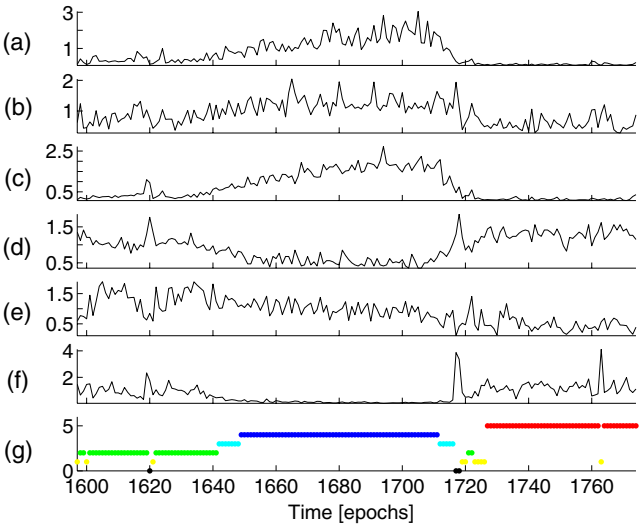

Fig. 4 Scaled features of EEG data set sc4002e0, as described in Section 4.2. The subfigures show power in the (**a**) delta and (**b**) theta bands, (**c**) variance, (**d**) spindle score, (**e**) maximum height of the power spectrum in the alpha band after subtraction of a linear estimate, and (**f**) high power fraction. Figure (**g**) shows the hypnogram of the EEG data, where the number and color indicate the sleep stage: awake (0, *black*), stage 1 (1, *yellow*), stage 2 (2, *green*), stage 3 (3, *cyan*), stage 4 (4, *blue*), and REM (5, *red*). The features were calculated for the data from epochs 1,597–1,774 in 30-second windows with no overlap

We then use these features as the high-dimensional input to the LLE algorithm. The 2D results for 13 nearest neighbors ($k = 13$) are displayed in Fig. 5(a). Every point in this figure represents a 30-second window of EEG data, and the color and symbol represent the sleep stage as determined by manual scoring. Here we see a very clear separation between the REM points (red circles) and those from stage 4 (blue stars), as required by our criteria for the automatic selection of the feature set. Stages 1 through 3 are located between those two groups and are arranged by sleep depth. In this example, we see a general trend of increasing sleep depth as we move to the upper right corner of the space. In addition, this low-dimensional embedding provides results with a smooth temporal progression. This is demonstrated by Fig. 5(b), where the LLE results from Fig. 5(a) are plotted versus time. In this example, the gradual transition to deep stage 4 sleep and the quick transition to REM are visible in the plot of Y_1.

We would like to emphasize the importance of identifying strongly connected components when using LLE. Figure 6(a) shows an example of the Y_1–Y_2 output space when LLE is performed on all 178 data points. Here, the feature set consisted of power in the delta, theta, and gamma bands, total power, maximum value of the PSD in the alpha band, and the

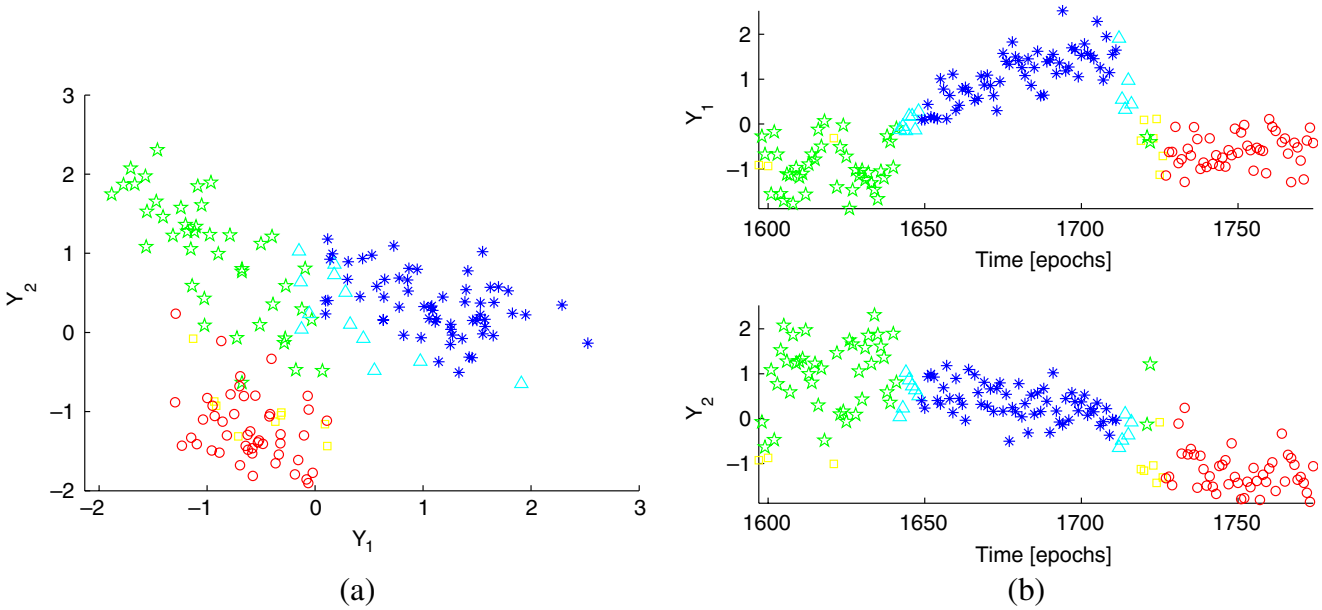

Fig. 5 (**a**) Results of applying LLE to EEG data using the six features in Fig. 4. The features were calculated for 30-second non-overlapping windows of data and the resulting 6-dimensional points were embedded in 2D space using LLE with $k = 13$; therefore, each point in this figure represents 30 s of EEG that has been characterized by the six features. The color and shape indicate sleep stage based on manual scoring: awake (*black* +), stage 1 (*yellow* □), stage 2 (*green* ⋆), stage 3 (*cyan* △), stage 4 (*blue* ∗), and REM (*red* ○). (**b**) LLE output dimensions Y_1 and Y_2 versus time, for the results shown in (a). This demonstrates that LLE provides a low-dimensional output where the sleep state changes smoothly over time

low power fraction. Again, each point represents 30-seconds of EEG data, and the symbol (and color) are assigned based on its designated sleep stage. While there is some visible separation between the stages, the overall trend is unclear. On the other hand, Fig. 6(b) shows the results when LLE is applied to the largest strongly connected component within the data. This component was identified as described in Section 3.2,

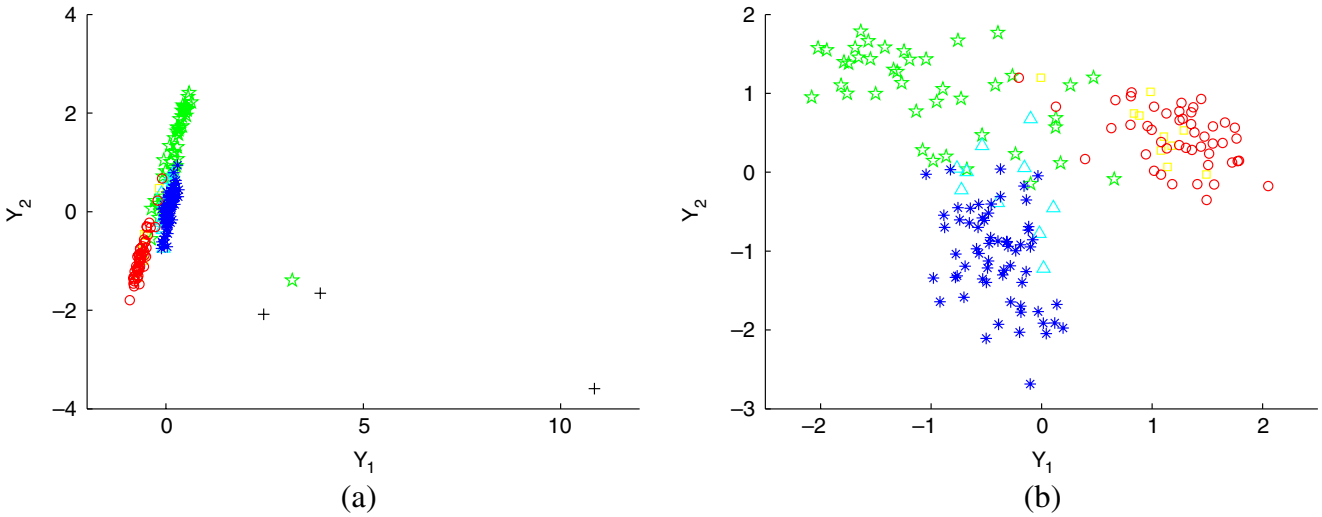

Fig. 6 LLE results on sleep EEG data before (**a**) and after (**b**) removal of eight weakly connected points. The six features were power in the delta, theta, and gamma bands, total power, maximum height of PSD above a linear estimate in the alpha band, and low power fraction, and we used $k = 13$. As before, the color and symbol indicate sleep stage: awake (*black* +), stage 1 (*yellow* □), stage 2 (*green* ⋆), stage 3 (*cyan* △), stage 4 (*blue* ∗), and REM (*red* ○). Note the dramatic improvement in separation between sleep stages when LLE is done on only one strongly connected component in (b)

and all other points were removed before using LLE. This greatly improves the results; the data points are spread further apart, and we see a grouping of sleep stages similar to Fig. 5. Sometimes the removal of weakly connected points has a very small impact on the results, but situations like this make it a necessity. The significant improvement for this feature set allowed it to be counted as one of the "best" 11 results discovered by the automatic algorithm.

In this case, analysis of the strongly connected components resulted in the removal of eight data points:

- 3 points from waking (epochs 1,620, 1,717, and 1,718)
- 2 points from stage 2 (epochs 1,619 and 1,624)
- 2 points from stage 3 (epochs 1,712 and 1,713)
- 1 point from stage 4 (epoch 1,665)

Based on Fig. 6(a), we can see why some of these were removed; there are four points that are clearly isolated from the rest of the data. However, the removal of points from stages 3 and 4 are much less obvious. It is important to realize that, by using the concept of strongly connected components, this decision is automatic—it allows us to avoid the subjective selection of outlier points.

5 Integration of EEG data and the model sleep cycle

Thus far, we have shown that LLE is capable of distinguishing between sleep stages using only one channel of EEG and that the embedding exhibits a smooth progression over time. However, remember that our original goal was to find the relationship between EEG data and the mathematical model of the sleep cycle. Here we accomplish this by applying LLE *simultaneously* to EEG data and simulated data from the model.

5.1 Model data set

To generate the model data set, we place a grid of points on the sleep manifold (Fig. 1) and obtain the numerical solution of the cortical model at each one. We vary L over the interval [0.5, 2] in increments of 0.1 and Δh_e^{rest} over $[-5, 5]$ in increments of 0.5. This gives us a total of 336 model signals for analysis; we then remove the initial transients and characterize each signal based on a subset of the features described in Section 4.2. In this way, the nonlinear sleep manifold is turned into "EEG-like" signals which are converted to high-dimensional data points for use with LLE.

For the model data set, the length of each signal is 10 s (as opposed to the 30-second windows used for the EEG data). We are able to use this shorter time because we can choose parameters in the model to simulate a stationary brain state, i.e. we can use constant values of L and Δh_e^{rest}. A test of the feature calculations for various window lengths indicated that, in many cases, the signal properties were stationary for windows greater than five seconds. Certain parts of the sleep manifold had transients lasting roughly 10 seconds.

In order to compare this model data set directly to EEG measurements, it is important that all of the basic properties match. For example, just as REM EEG signals have a much lower variance than those from stages 3 and 4, we expect that the signals from the topmost REM portion of the sleep manifold will have a smaller variance than those on the lower NREM section. However, we found that the use of a constant α, which defines the variance of the stochastic input to the model cortex in Eq. (10), does not reproduce this behavior. Therefore, we varied the value of α as we moved in the L-Δh_e^{rest} space. More specifically, we based it on the sleep manifold. Define $\boldsymbol{\mu}_e$ to be a matrix of the steady-state values of h_e after they have been shifted and scaled to have a range of [0, 1]. Then we define a matrix of α values:

$$\boldsymbol{\alpha} = \alpha \cdot \left(-7\boldsymbol{\mu}_e + 8\right) \ . \tag{15}$$

Therefore, the REM portion of the model sleep cycle (where $\mu_e \sim 1$)) will have stochastic inputs of α, while the lower NREM section (where $\mu_e \sim 0$) will have inputs of variance 8α. This stochastic input allowed us to successfully reproduce the desired range of variances in the model signals.

In addition to the variance, other features of the model data set mimic characteristics of sleep EEG. This can be verified by plotting the features as we traverse the sleep manifold. For example, power in the delta band, composite permutation entropy index (CPEI), "peak" height of the power spectral density in the alpha band, low power fraction, and high power fraction are shown in Fig. 7. The values of each feature are displayed for the grid of points in L and Δh_e^{rest} that covers the sleep manifold. For reference, the steady state values of h_e on the sleep manifold are shown in Fig. 7(a); note that this is similar to viewing Fig. 1 from the top and coloring the points based on their height. The lowest value is plotted in white and the highest value in black.

As desired, Fig. 7(b) indicates that the power in the delta band increases as the depth of sleep increases,

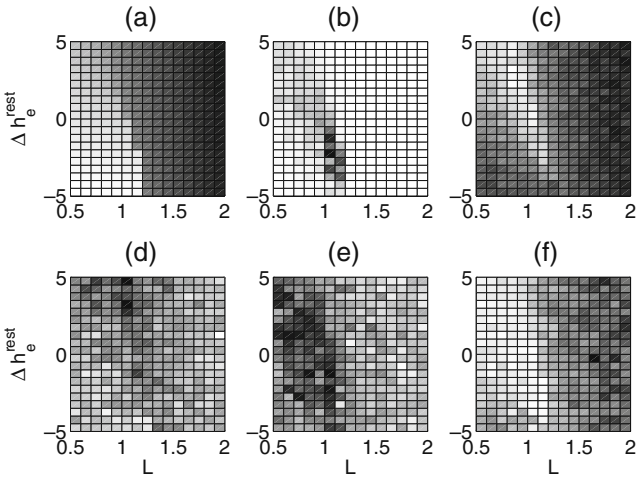

Fig. 7 Variation of five features as the surface of the sleep manifold is traversed in L-Δh_e^{rest} space. Each feature has been scaled by its RMS value and depicted in *grayscale*, with *white* indicating the lowest values and *black* representing the highest values. (**a**) The steady state values of h_e from the sleep manifold in Fig. 1. The *black points* represent the upper REM branch, the *white points* represent NREM, and the fold is located at roughly $L = 1.2$. The other subfigures show (**b**) power in the delta band, (**c**) permutation entropy, (**d**) maximum height of PSD above a linear estimate in the alpha band, (**e**) low power fraction, and (**f**) high power fraction. These five features use α as defined in Eq. (15). They show that the representation of REM and NREM in the model is consistent with the characteristics of sleep EEG

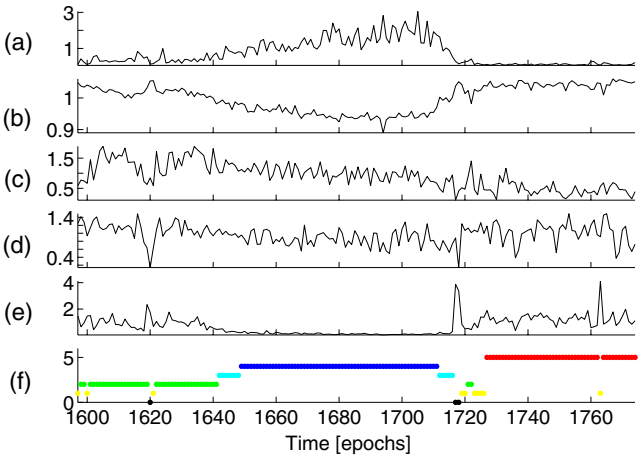

Fig. 8 Variation of the features from Fig. 7 when they are applied to sleep EEG data, rather than model data. The sample of EEG data was taken from sc4002e0, and each feature has been scaled by its RMS value. The subfigures show (**a**) power in the delta band, (**b**) permutation entropy, (**c**) maximum height of PSD above a linear estimate in the alpha band, (**d**) low power fraction, (**e**) high power fraction, and (**f**) hypnogram of the EEG data. The colors and numbering for the hypnogram are the same as those used for Fig. 4. Note that the values of these features (relative to sleep stage) are consistent with the model results in Fig. 7

with the largest values occurring near the quick transition to REM sleep. Similarly, Fig. 7(e) and (f) show that the fraction of power in the low frequencies is greater during NREM sleep, while the fraction of power at high frequencies is greater during REM sleep. Consistent with previous reports that the CPEI decreases with depth of anesthesia (Olofsen et al. 2008), we see in Fig. 7(c) that the CPEI decreases with sleep depth in the model. Figure 7(d) shows that the region of greatest alpha power is located in the upper left corner, for small values of L and large values of Δh_e^{rest}. As a means of comparison, the same five features were applied to a sample of EEG data and are displayed in Fig. 8.

5.2 Application of LLE to a hybrid data set

We now join the EEG measurements and the model data into one hybrid data set and use it as an input to the LLE algorithm. This simultaneously finds the low-dimensional embedding for both data types and allows us to infer a correspondence between them. For example, Fig. 9(a) shows the result of applying LLE to the grid of 336 model points and a full night's sleep from EEG data set sc4002e0 (epochs 800 to 2,000). The input data was composed of the five features from Fig. 7: power in the delta band, CPEI, maximum height of the PSD in the alpha band (relative to a linear estimate), and the low and high power fractions. We used $k = 14$, and only three points were removed by analysis of the strongly connected components.

In Fig. 9(a), the model data is represented by dots, where the color denotes the steady-state value of h_e associated with that point; in general, the red points represent the REM portion of the manifold, while the blue points represent NREM. On the other hand, the sleep EEG data points are rings, where the color is chosen based on sleep stage. Note that, for clarity, only the first 500 EEG data points were included in the figure.

The most important aspect of this result is that the EEG data points and model points overlap each other in the Y_1–Y_2 output space. This implies that model points have EEG data points as nearest neighbors (and vice versa) and verifies that LLE has associated the two data types with one another. Without fidelity of the model and careful choice of EEG features, we would have likely obtained a result with one cluster of EEG points and a completely separate cluster of model points. Further, LLE appears to have matched the sleep stages between the two data types—the deepest sleep (blue for both EEG and model) appears in the lower left corner, and REM (red) is embedded in a vertical

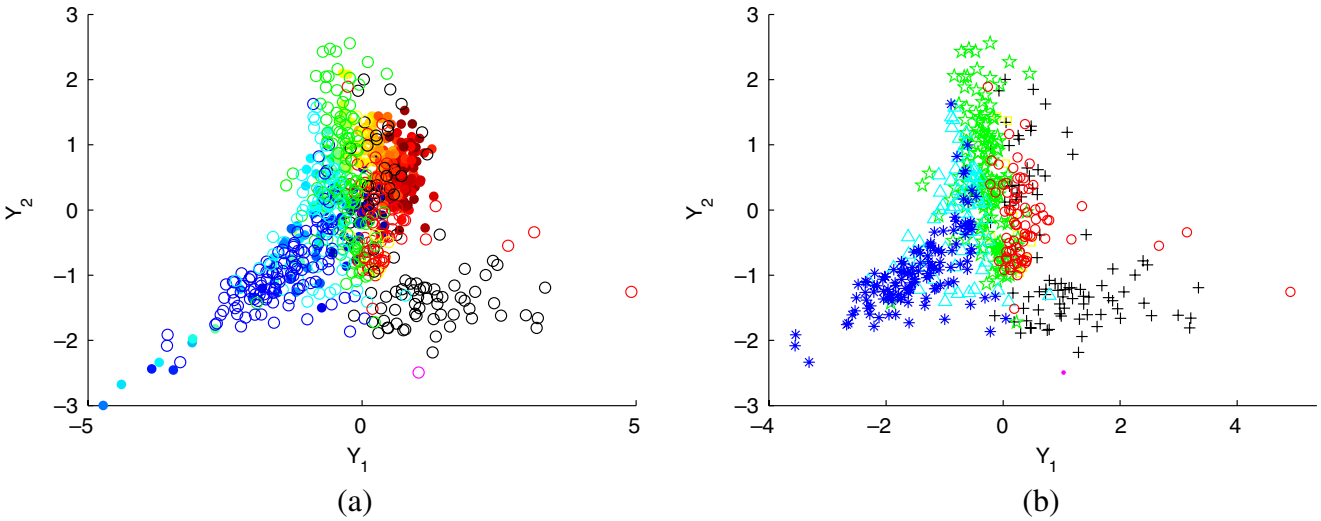

Fig. 9 (**a**) LLE results for a hybrid data set containing both sleep EEG data and numerical solutions of the cortical model. We used the five features from Fig. 7 and set $k = 14$. The *rings* represent EEG data and are colored by sleep stage. While the analysis included 1,200 windows of EEG data, only 500 are displayed here for clarity. The *solid dots* represent data from the model; they are colored based on the mean value of h_e at that point, where *red* represents the highest (REM) values, and *dark blue* marks the lowest (NREM) values. Note that the data and model points overlap in the output space and that the arrangement of sleep stages is very similar. (**b**) LLE results showing the EEG data *only*, using the same colors and symbols as Fig. 6. This allows us to see that the data has been roughly separated by sleep stage

band where Y_1 is in the range $[-1, 0]$. The separation between sleep stages can be seen more clearly in Fig. 9(b), which displays only the EEG data points from Fig. 9(a). Here we see that the stages are grouped; even the REM points and the awake points are separated, despite the fact that their EEG traces are characterized by very similar features. If we were to plot the Y_1 and Y_2 values of the EEG data points as they evolve in time, we would see a very similar result to the one in Fig. 5(b). Here, the Y_1 direction appears to be an approximate indicator of sleep depth.

5.3 Connection to the theoretical sleep manifold

So far, we have seen that LLE provides a qualitatively similar embedding for REM and NREM points in both EEG measurements and simulated model data. However, we would like to quantify this relationship. In other words, we would like to associate each EEG data point with a position on the sleep manifold in the L-Δh_e^{rest} space. This will allow us to infer the model trajectory of a subject's actual brain state as it moves along the manifold.

To do this, we use the results in Fig. 9(a) and again turn to the concept of nearest neighbors. Using $k = 14$, we calculate the nearest neighbors of every point in the Y_1–Y_2 space. We then identify *model* points that are nearest neighbors of *EEG* data points. Each one of those model points has an associated position on the sleep manifold; we assume that the L-Δh_e^{rest} positions of the model nearest neighbors will be the most closely associated positions for the EEG data point.

We can visualize this concept by creating histograms of the model nearest neighbors and separating them by sleep stage (Fig. 10(a)). Every time a model point is a nearest neighbor of an EEG point, we increment the count at the model point's associated location in L-Δh_e^{rest} for the sleep stage of the EEG point. We then create grayscale plots of the total counts, where white indicates that a location was never a nearest neighbor of that sleep stage and black indicates that it was a nearest neighbor many times.

For example, (i)–(vi) in Fig. 10(a) correspond to awake, REM, and stages 1–4, respectively. The thick vertical line at $L = 1.2$ marks the approximate location of the fold. As we move from REM to the deeper stages of sleep, we can see a continuous progression along the sleep manifold. In this example, REM and stage 1 sleep generally associate themselves with locations on the right half of the manifold (and a small piece of the lower left corner). Then in stage 2 sleep, we move to the left half of the manifold; here, we see two distinct groups of points, with a majority landing in the group that borders the area associated with REM and Stage 1. Stage 3 is associated with a cluster of points starting in the upper left-hand corner and approaching the fold. Stage 4 continues this progression and is located in a band of points leading up to the fold.

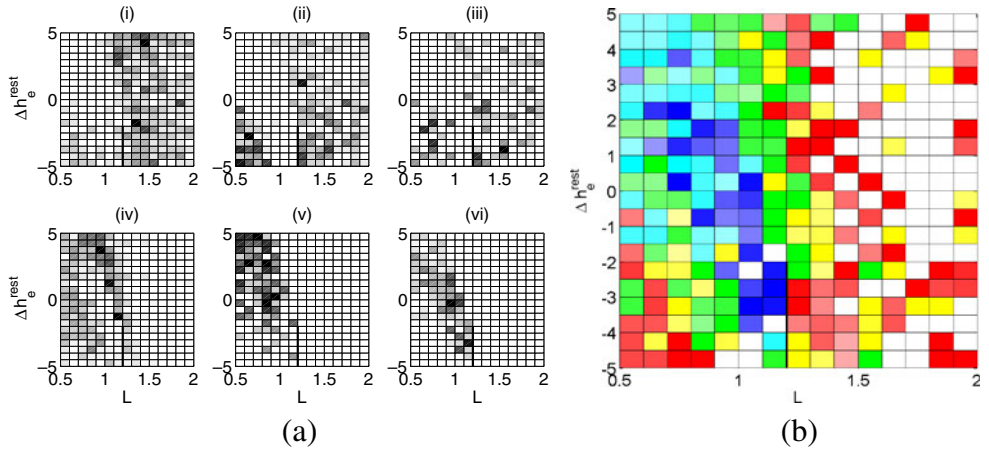

Fig. 10 Association between EEG data set sc4002e0 and the sleep manifold. Each picture shows the sleep manifold in L-Δh_e^{rest} space, with a *heavy black line* to indicate the location of the fold. (**a**) Histograms of nearest neighbors for (*i*) waking, (*ii*) REM, (*iii*) stage 1, (*iv*) stage 2, (*v*) stage 3, and (*vi*) stage 4 sleep. The shading of each square indicates the number of times that location on the sleep manifold was a nearest neighbor of EEG data in that stage. For example, (*vi*) shows that stage 4 sleep most often associates itself with the lower NREM branch of solutions leading up to the fold. (**b**) A composite picture of the results in (a), where each location is colored based on the sleep stage with the most neighbors at that point, relative to the total number of neighbors associated with that stage. Again, we use stage 1 (*yellow*), stage 2 (*green*), stage 3 (*cyan*), stage 4 (*blue*), and REM (*red*). The intensity of color is scaled based on the percentage of neighbors that come from that stage; the more saturated the color, the greater the percentage. Waking points were excluded

We can then create a composite plot that combines all five sleep stages. We neglect the waking points for this task because the current model does not effectively distinguish between the waking and REM states, although this is certainly an issue that may be addressed in the future. For every location on the manifold, we determine which sleep stage it was most closely associated with and color it accordingly. To do this, we scale the number of nearest neighbors for each stage by the total number for that stage; then, for every position on the manifold, we choose the stage with the highest value. This accounts for the fact that the subjects do not spend an equal amount of time in each sleep stage (otherwise, more time spent in a certain stage would lead to more nearest neighbors and a greater likelihood of dominating this composite plot). As in previous figures, we use red for REM, yellow for stage 1, green for stage 2, cyan for stage 3, and blue for stage 4. The intensity of the color is assigned based on the percentage of times it was associated with that sleep stage. Suppose a certain point on the manifold was a neighbor of stage 2 twelve times, a neighbor of stage 1 five times, and a neighbor of REM three times. We would color that point green to indicate stage 2 sleep, and its saturation value would be $12/(12+5+3) = 0.6$. In other words, the intensity of the color is a "confidence" measure; the more saturated the color, the more closely it is associated with that sleep stage. The composite figure for the data in Fig. 10(a) is shown in Fig. 10(b).

5.4 Inclusion of additional data sets

It is important that this method of analysis works consistently for different subjects with a variety of sleeping patterns. We tested this capability using the full night of sleep from each of the remaining three data sets: sc4012e0, sc4102e0, and sc4112e0. Rather than start from scratch and re-run the LLE algorithm, we projected the new data onto the existing embedding. For a new input $\mathbf{x}$, this is a three-step process (Saul et al. 2003):

1. Find the k nearest neighbors of each new data point among the points in the existing embedding.
2. Compute the best linear reconstruction w_j of each new point using only its nearest neighbors. Again, we enforce the constraint that the weights used in the reconstruction sum to one: $\sum_j w_j = 1$.
3. Calculate the output for the new data points: $\mathbf{y} = \sum_j w_j \mathbf{Y}_j$, where $\mathbf{Y}$ contains the original embedding coordinates and j cycles through the neighbors of $\mathbf{x}$.

This is more computationally efficient than running the entire algorithm again, and it guarantees that the output embedding will not change as we add new data. Most importantly, this makes it possible to do continuous real-time monitoring of EEG data; a new point could be projected onto the results every 30 s (or less) as the subject sleeps.

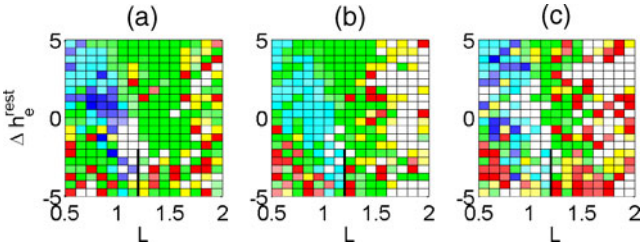

Fig. 11 Composite plots for EEG data sets (**a**) sc4012e0, (**b**) sc4102e0, and (**c**) sc4112e0, when they are projected onto the LLE results from Fig. 6(b), as described in Section 5.4. These pictures are analogous to Fig. 10(b) and use the same color scheme. Note that the results are consistent with those for sc4002e0 in Fig. 10(b); over various subjects, the sleep stages are generally associated with the same positions on the sleep manifold

When we project the sleep data from files sc4012e0, sc4102e0, and sc4112e0 onto the embedding derived from sc4002e0, we obtain the composite pictures in Fig. 11(a)–(c), respectively. All three results are consistent with one another, despite coming from different subjects and containing a minimal amount of stage 3 and 4 deep sleep. The only exception to this is stage 4 sleep in Fig. 11(c); however, it is important to note that only 21 points out of 1100 were denoted as stage 4 sleep for this subject, and those points were not all consecutive. Therefore, the subject had only transient movements into stage 4 from stage 3, and it is perhaps not surprising that the results show the stage 4 EEG points mixed in with those from stage 3. Also note that the placement of the sleep stages in Fig. 11 is consistent with the results in Fig. 10.

Lastly, we combine the results from all four data sets (the original embedding with sc4002e0 plus three projected data sets) to produce Fig. 12. The histograms in Fig. 12(a) were created by a simple summation of the nearest neighbor histograms for all four data sets. The composite plot in Fig. 12(b) was then generated according to the logic described in Section 5.3 using the combined histogram data. In all, these results are based on almost 40 h of EEG data from four different subjects. Again, they are consistent with the individual results and they show a clear picture of the sleep manifold regions associated with each sleep stage. It is also noteworthy that only a handful of points on the sleep manifold (colored white in the composite picture) were never nearest neighbors of an EEG data point.

This picture may be very useful in the analysis of seizures during sleep. Imagine taking another new sleep EEG data set, this time from an epileptic subject, and projecting it onto these results. By following the location in L-Δh_e^{rest} as the subject sleeps, we can get an idea of the sleep stage as it is traditionally defined, and *we can also identify that stage in more detail and detect nearness to transitions between stages*. The grid of points on the sleep manifold essentially gives us descriptions of 336 different brain states associated with sleep. We expect that future research will identify the locations on the sleep manifold where seizures are most likely to occur. With that knowledge, if the sleep state

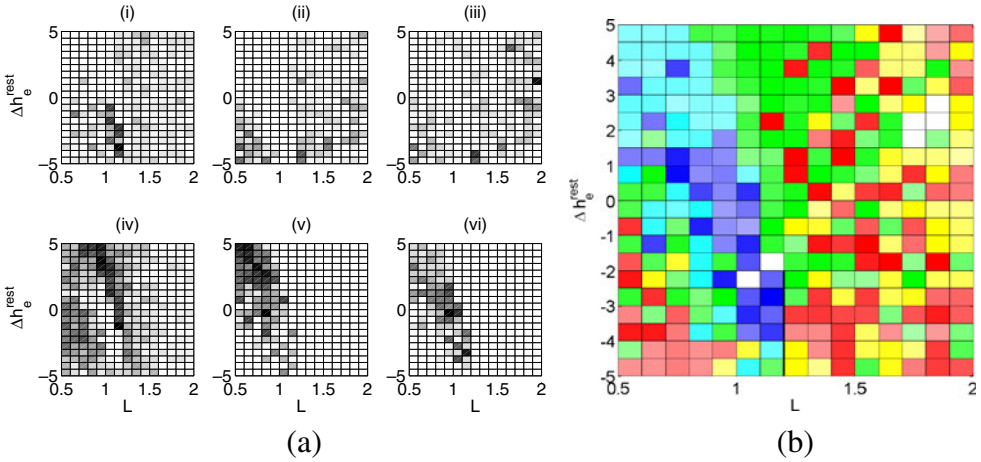

Fig. 12 Combined association between all four EEG data sets and the sleep manifold. The data set sc4002e0 was directly compared to the cortical model using LLE, and the remaining three data sets were projected onto those results as described in Section 5.4. (**a**) Total histograms of nearest neighbors, separated by sleep stage; these were calculated by summing the histograms from all four EEG data sets. The pictures show awake, REM, and stages 1–4 in (*i*) through (*vi*), respectively. (**b**) The total composite picture for all four data sets. This was generated from the histogram data in (a) and is analogous to Figs. 10(b) and 11. Again, this is consistent with previous results and shows the regions of the sleep manifold most closely associated with each sleep stage

characterization is done continuously while the subject is sleeping, this may allow for the prediction (and possibly prevention) of seizures.

We emphasize the fact that the coloring in the composite pictures (Figs. 10(b), 11, and 12(b)) is based on the subjective scoring of sleep data. The reliability of categorizing individual epochs of data has been reported at 73% for scorers from different labs (Norman et al. 2000) and as high as 90% for scorers from the same lab (Whitney et al. 1998). It has also been shown that reliability varies by sleep stage, with stage 2 having the highest level of agreement between scorers (78.3%) and stage 1 having the lowest (41.8%) (Norman et al. 2000). This certainly affects our results. For example, imagine if some of the points scored as REM that landed in the range $1.6 < L < 2$ on the sleep manifold were instead scored as stage 1. Then the right side of the composite picture would be completely yellow and the region associated with stage 1 would be more clear. Therefore, the composite pictures should be seen as "guides" to tie the analysis back to the traditional definitions of the sleep stages, not as the ultimate truth. As mentioned in the previous paragraph, we are most interested in the *position* on the sleep manifold, the trajectory that results as the subject sleeps, and the relationship of this trajectory to the regions where seizures may be most likely to occur.

6 Summary

Mathematical models represent an opportunity for exploration and prediction. In this case, a model of the human sleep cycle creates the possibility for a more detailed description of sleep states, with application to the prediction and analysis of seizures during sleep. The first step in such an endeavor is always to connect the model to the real world through experimental data.

Here we have used locally linear embedding to directly associate human sleep EEG data with the mathematical model. We first showed that LLE has the ability to distinguish between sleep stages when applied to EEG data alone. This analysis can reliably separate REM and NREM sleep data and provide a smooth temporal progression through the various stages of sleep. We also presented the concept of strongly connected components as a method of automatic outlier rejection for EEG data and discussed a method for the selection of EEG features used in the analysis. Then, by using LLE on a hybrid data set containing both sleep EEG and signals generated from the mathematical sleep cycle, we were able to quantify the relationship between the model and the data. This enabled us to take any sample of sleep EEG data and associate it with a position among the continuous range of sleep states provided by the model. In addition, this approach yields consistent results for various subjects over a full night of sleep and can be done online as the subject sleeps. This suggests a wide range of possibilities for future investigation.

Acknowledgements This work was supported through a National Science Foundation Graduate Research Fellowship to B. A. Lopour. It was also supported, in part, by a Mary Elisabeth Rennie Epilepsy and Epilepsy-related Research Grant. We extend a special thanks to Kelly Clancy and Albert Kao for work which served as the starting point for this project, done as part of a course in computational neuroscience at UC Berkeley taught by Professor B. A. Olshausen.

References

Ataee, P., Yazdani, A., Setarehdan, S. K., & Noubari, H. A. (2007). Manifold learning applied on EEG signal of the epileptic patients for detection of normal and pre-seizure states. In *Proceedings of the 29th Annual International Conference of the IEEE EMBS* (pp. 5489–5492).

Bojak, I., & Liley, D. T. (2005). Modeling the effects of anesthesia on the electroencephalogram. *Physical Review E*, 71(041902).

Corsi-Cabrera, M., Guevara, M. A., Río-Portilla, Y. D., Arce, C., & Villanueva-Hernández, Y. (2000). EEG bands during wakefulness, slow-wave and paradoxical sleep as a result of principal component analysis in man. *SLEEP, 23*(6), 1–7.

Fuller, P., Gooley, J., & Saper, C. (2006). Neurobiology of the sleep-wake cycle: Sleep architecture, circadian regulation, and regulatory feedback. *Journal of Biological Rhythms, 21*(6), 482–493.

Fuller, P., Saper, C., & Lu, J. (2007). The pontine rem switch: Past and present. *Journal of Physiology, 584*(3), 735–741.

Gervasoni, D., Lin, S.-C., Ribeiro, S., Soares, E. S., Pantoja, J., & Nicolelis, M. A. (2004). Global forebrain dynamics predict rat behavioral states and their transitions. *The Journal of Neuroscience, 24*(49), 11137–11147.

Goldberger, A. L., Amaral, L. A. N., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., et al. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation, 101*(23), e215–e220. Circulation Electronic Pages: http://circ.ahajournals.org/cgi/content/full/101/23/e215.

Jobert, M., Escola, H., Poiseau, E., & Gaillard, P. (1994). Automatic analysis of sleep using two parameters based on principal component analysis of electroencephalography spectral data. *Biological Cybernetics, 71*, 197–207.

Kemp, B. (2009). *The Sleep-EDF Database*. http://www.physionet.org/physiobank/database/sleep-edf/. Accessed August 2009.

Kramer, M. A., Kirsch, H. E., & Szeri, A. J. (2005). Pathological pattern formation and cortical propagation of epileptic seizures. *Journal of the Royal Society Interface, 2*, 113–127.

Kramer, M. A., Szeri, A. J., Sleigh, J. W., & Kirsch, H. E. (2007) Mechanisms of seizure propagation in a cortical model. *Journal of Computational Neuroscience, 22*, 63–80.

Leslie, K., Sleigh, J., Paech, M., Voss, L., Lim, C., & Sleigh, C. (2009). Dreaming and electroencephalographic changes during anesthesia maintained with propofol or desflurane. *Anesthesiology, 111*(3), 547–555.

Liley, D. T., & Bojak, I. (2005). Understanding the transition to seizure by modeling the epileptiform activity of general anesthetic agents. *Journal of Clinical Neurophysiology, 22*(5), 300–313.

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

Lopour, B. A., & Szeri, A. J. (2010). A model of feedback control for the charge-balanced suppression of epileptic seizures. *Journal of Computational Neuroscience, 28*(3), 375–387.

McCarley, R. (2007). Neurobiology of REM and NREM sleep. *Sleep Medicine, 8*(4), 302–330.

McKay, E., Sleigh, J., Voss, L., & Barnard, J. (2010). Episodic waveforms in the electroencephalogram during general anaesthesia: A study of patterns of response to noxious stimuli. *AIC, 38*(1), 102–112.

Mourtazaev, M., Kemp, B., Zwinderman, A., & Kamphuisen, H. (1995). Age and gender affect different characteristics of slow waves in the sleep EEG. *Sleep, 18*(7), 557–564.

Niedermeyer, E., & da Silva, F. L. (2005). *Electroencephalography: Basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins.

Norman, R. G., Pal, I., Stewart, C., Walsleben, J. A., & Rapoport, D. M. (2000). Interobserver agreement among sleep scorers from different centers in a large dataset. *SLEEP, 23*(7), 901–908.

Olofsen, E., Sleigh, J. W., & Dahan, A. (2008). Permutation entropy of the electroencephalogram: A measure of anaesthetic drug effect. *British Journal of Anaesthesia, 101*(6), 810–821.

Polito, M., & Perona, P. (2001). Grouping and dimensionality reduction by locally linear embedding. In *Advances in neural information processing systems 14* (pp. 1255–1262). MIT Press.

Rosenwasser, A. (2009). Functional neuroanatomy of sleep and circadian rhythms. *Brain Research Reviews, 61*, 281–306.

Roweis, S. T., & Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science, 290*(5500), 2323–2326.

Roweis, S. T., & Saul, L. K. (2009). *Locally linear embedding*. http://www.cs.toronto.edu/ roweis/lle/. Accessed June 2009.

Saper, C., Lu, J., Chou, T., & Gooley, J. (2005a). The hypothalamic integrator for circadian rhythms. *Trends in Neuroscience, 28*(3), 152–157.

Saper, C., Scammell, T., & Lu, J. (2005b). Hypothalamic regulation of sleep and circadian rhythms. *Nature, 437*(7063), 1257–1263.

Saul, L. K., Roweis, S. T., & Singer, Y. (2003). Think globally, fit locally: Unsupervised learning of low dimensional manifolds. *Journal of Machine Learning Research, 4*, 119–155.

Steriade, M., & Amzica, F. (1998). Coalescence of sleep rhythms and their chronology in corticothalamic networks. *Sleep Research Online, 1*(1), 1–10.

Steriade, M., & Timofeev, I. (2001). Natural waking and sleep states: A view from inside neocortical neurons. *Journal of Neurophysiology, 85*(5), 1969–1985.

Steyn-Ross, D. A., Steyn-Ross, M. L., Sleigh, J. W., Wilson, M. T., Gillies, I. P., & Wright, J. J. (2005). The sleep cycle modelled as a cortical phase transition. *Journal of Biological Physics, 31*, 547–569.

Steyn-Ross, M. L., Steyn-Ross, D. A., Sleigh, J. W., & Liley, D. T. J. (1999). Theoretical electroencephalogram stationary spectrum for a white-noise-driven cortex: Evidence for a general anesthetic-induced phase transition. *Physical Review E, 60*(6), 7299–7311.

Steyn-Ross, M. L., Steyn-Ross, D. A., Sleigh, J. W., & Whiting, D. R. (2003). Theoretical predictions for spatial covariance of the electroencephalographic signal during the anesthetic-induced phase transition: Increased correlation length and emergence of spatial self-organization. *Physical Review E, 68*, 021902.

Steyn-Ross, M. L., Steyn-Ross, D. A., & Sleigh, J. W. (2004). Modelling general anaesthesia as a first-order phase transition in the cortex. *Progress in Biophysics & Molecular Biology, 85*, 369–385.

Tarjan, R. (1972). Depth-first search and linear graph algorithms. *SIAM Journal on Computing, 1*(2), 146–160.

Whitney, C. W., Gottlieb, D. J., Redline, S., Norman, R. G., Dodge, R. R., Shahar, E., et al. (1998). Reliability of scoring respiratory disturbance indices and sleep staging. *SLEEP, 21*(7), 749–757.

Wilson, M. T., Steyn-Ross, M. L., Steyn-Ross, D. A., & Sleigh, J. W. (2005). Predictions and simulations of cortical dynamics during natural sleep using a continuum approach. *Physical Review E, 72*(051910).

Wilson, M. T., Steyn-Ross, A., Sleigh, J. W., Steyn-Ross, M. L., Wilcocks, L. C., & Gillies, I. P. (2006). The k-complex and slow oscillation in terms of a mean-field cortical model. *Journal of Computational Neuroscience, 21*, 243–257.