NETWORKS AND AN SIRS MODEL OF EPILEPSY EEG DATA

NETWORK STRUCTURES AND THEIR EFFECT ON A STOCHASTIC SIRS MODEL OF EPILEPSY EEG DATA

By EVAN MITCHELL, B. ARTS SC.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree Master of Science

McMaster University MASTER OF SCIENCE (2017) Hamilton, Ontario (Mathematics)

TITLE: Network Structures and their Effect on a Stochastic SIRS Model of Epilepsy EEG Data AUTHOR: Evan Mitchell, B. Arts Sc. (McMaster University) SUPERVISOR: Professor Benjamin Bolker NUMBER OF PAGES: vii, 17

Abstract

In this thesis, we consider a stochastic SIRS model of EEG data. The model is built over three different network structures: a random network, a scale-free network, and a small-world network. These models are then fit to an EEG signal from a control individual and an EEG signal from an individual experiencing an epileptic seizure. We are interested in determining whether these models can distinguish between the two data sets, and whether any of the network structures offer a significantly better fit to the data than others; there is also a broader interest in the effects of different network structures on the time series characteristics of an SIRS system.

Acknowledgements

I want to extend a special thanks to my supervisor, Dr. Ben Bolker, for his patience and advice. Completion of this thesis would not have been possible without his expertise, guidance, and coding knowledge. I also want to thank the staff and professors at McMaster University for continuing to foster my love of mathematics, and my family for their support over the course of this work.

Contents

List of Figures

List of Tables

List of Abbreviations

EEG → electroencephalogram
SIRS → susceptible-infected-recovered-susceptible epidemiological model

1 Introduction

Epilepsy is a neurological condition in which an individual experiences a period of abnormal neuronal activity. This activity manifests as repeated seizures, although the period of repetition can vary greatly from hours to years [10]. Often accompanying these seizures is a loss of consciousness, causing the individual to retain no memory of the event [10]. Epilepsy treatments consist of drugs aimed at suppressing abnormal firing of neurons to reduce the frequency of seizures [10].

A common way of diagnosing epilepsy is through the use of an electroencephalogram (EEG). In an EEG, electrodes placed on the skull capture the voltage changes that occur as neurons conduct electrical potentials [10]. These so-called "brain waves" are sent to a computer and depicted as a plot of voltage versus time [10]. During a seizure, an individual's EEG shows distinct differences from a non-seizure control condition. Differing frequencies and increased amplitudes can be treated as indicators of both the presence and duration of an epileptic event [10].

Since the EEG measures changes in a quantity over time, tools from the fields of differential equations and time series analysis can be brought to bear to describe these signals. Mathematical models of EEGs range from the purely phenomenological to those based firmly in biological principles. As an example of the former, Ghorbanian et al. investigated a model of EEGs of individuals with and without Alzheimer's disease [8, 9]. Their model is based on coupled Duffing-van der Pol oscillators with white noise and consists of a system of stochastic differential equations [8, 9]. They fit the model to the data based on both the power spectrum and information content (as measured by various types of entropy), and concluded that such a model can accurately match EEG data [8, 9].

An example of modelling with a more biological foundation can be found in the classic integrate-and-fire model. Here, the change in a neuron's membrane potential is related to the sum of all incoming synaptic signals, some of which are excitatory while others are inhibitory [6]. The neuron in this model is said to be "leaky" as the membrane potential decays at a fixed rate [6]. When the membrane potential reaches a given threshold, the neuron fires and a voltage spike is generated; the potential is then reset to the resting potential and the dynamics repeat [6]. While numerically straightforward to analyze, the integrate-and-fire model fails to take into account any of the specific physiological features of the neuron.

Models such as Hodgkin-Huxley and FitzHugh-Nagumo (essentially a simplified version of Hodgkin-Huxley) seek to incorporate these physiological features. The potential is governed by the flow of ions–specifically sodium, potassium, and chloride ions–across the neuronal membrane [3]. These models in-

clude terms that account for the movement of these ions and how they relate to the change in membrane potential [3]. While these additions make the model more biologically sound, they also make it more numerically complex to analyze.

As a sort of middle ground between the integrate-and-fire model and the Hodgkin-Huxley and FitzHugh-Nagumo models, we can look at the work of Acedo and Moraño [1]. There are three stages in which a neuron can be at any given time: resting (the "natural" or baseline state), firing (when a voltage change is actively taking place), or refractory (when a voltage change has recently occurred and the neuron is unable to go through another voltage change)[1, 10]. Acedo and Moraño draw parallels between these stages and the stages of the typical SIRS epidemiological model: as an action potential (voltage change) moves from brain cell to brain cell, a neuron goes from susceptible to an incoming voltage change (S) to undergoing a voltage change (I) to a state where another voltage change cannot occur (R), back to a susceptible state where the cycle can repeat [1]. Instead of assuming homogeneous mixing, they assign an underlying structure to the population of neurons in the form of a random network. In such a network, the average degree and the total number of nodes (neurons) are chosen, the desired number of edges/connections is calculated based on those choices, and those edges are randomly distributed among all pairs of nodes [1]. In their research, Acedo and Moraño found that a certain range of model parameter values resulted in sustained oscillations in the model output and they suggest that this model could be used to describe the oscillations found in EEG signals [1].

Inspired by the work of Acedo and Moraño, this thesis seeks to address a couple of questions. First: Can the stochastic SIRS model built over certain network structures distinguish between epilepsy and non-epilepsy (control) EEG data? Second: Does the specific network structure affect how well the model fits the data? We are also interested in the more general question of how different network structures affect the time series characteristics of a SIRS system.

2 Methods

2.1 Collecting and preparing the data

The EEG recordings used in this thesis come from a study by Andrzejak et al. [2]; the authors have kindly made their data freely accessible for any research purposes. Two sets of EEGs were downloaded: a recording from a healthy control subject and a recording from an individual experiencing an epileptic seizure. Each recording was taken at a sampling frequency of 173.61 Hz for a time of 23.6 seconds, and the values were standardized to be in units of

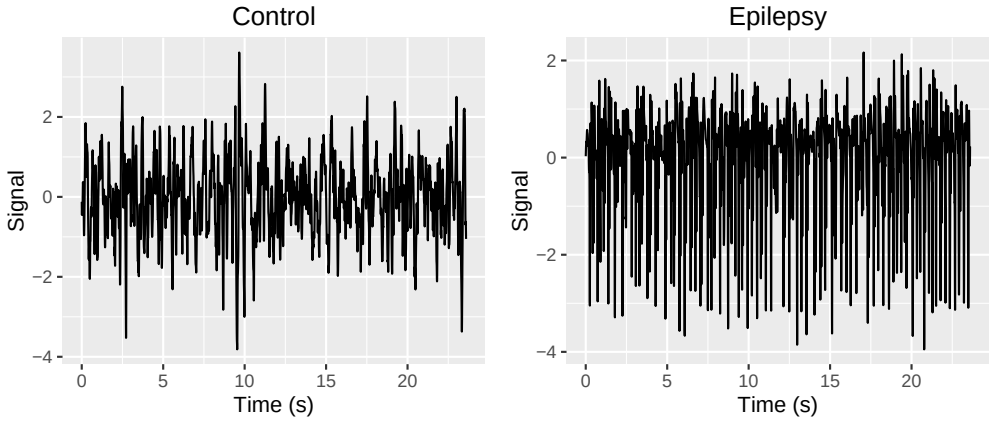

Figure 1: Filtered and standardized control and epilepsy EEG signals.

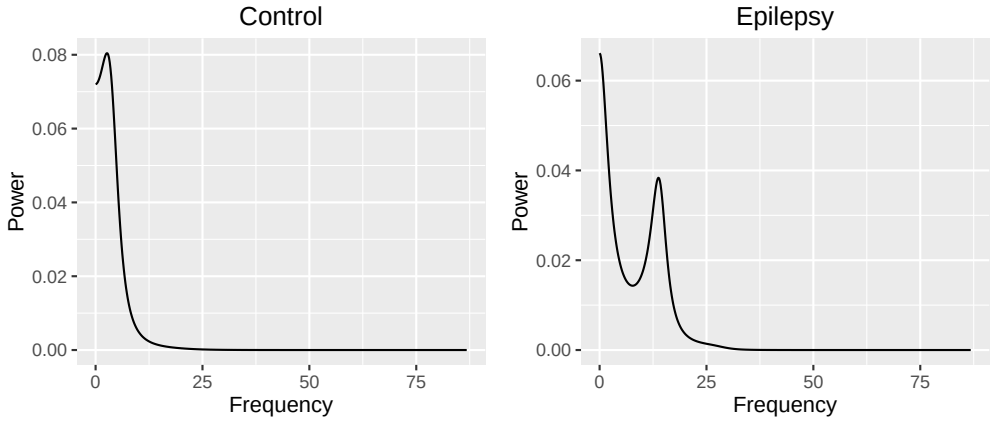

Figure 2: Power spectra for control and epilepsy data sets.

standard deviations from the mean. As outlined in the paper by Andrzejak et al. [2], a band-pass filter tuned to the range of 0.53-40 Hz (with a response roll-off of 12 dB/octave) was applied to the data prior to analysis. Figure 1 shows the filtered control and epilepsy EEG signals, and Figure 2 shows the power spectra for the control and epilepsy data sets.

2.2 Constructing the networks

We consider three network structures in this study: a random network (as described by Acedo and Moraño, [1]), a scale-free network, and a small-world network. In all three cases, a directed network was constructed. Also note

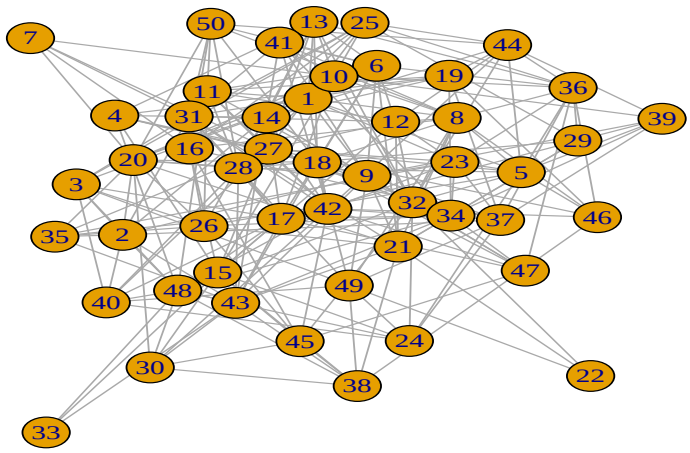

Figure 3: An example of a random network with 50 nodes.

that the term "degree" refers here to the number of incoming connections to a node (which was set to be equal to the number of outgoing connections), not the sum of both incoming and outgoing connections.

2.2.1 Random network

Random networks are ones in which the desired number of edges are distributed uniformly randomly among the total number of nodes, leading to a Poisson degree distribution [1]. An average degree k and total number of nodes N are chosen, then the kN edges are distributed among the N nodes. In our construction, we set $k = 5$ and $N = 10,000$. Figure 3 shows a smaller ($N = 50$) example of a random network.

2.2.2 Scale-free network

A scale-free network is one in which the degree distribution follows a power law, i.e., if $\tilde{k}$ represents a given degree, then the degree distribution $p(\tilde{k})$ is given by

$$p(\tilde{k}) = C\tilde{k}^{-\alpha}, \tag{1}$$

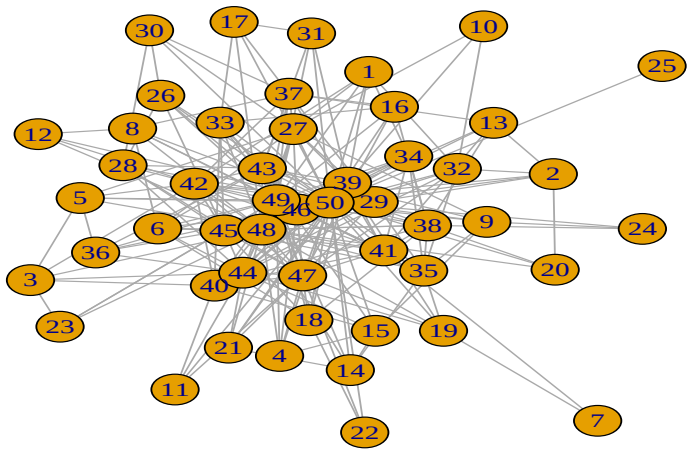

Figure 4: An example of a scale-free network with 50 nodes.

where $\alpha > 2$ is the power law exponent and $C = \alpha - 1$ is a normalization constant [4, 7]. This leads to the formation of "hubs": a small number of nodes with a large degree. As with the random network, we choose a total number of nodes $N = 10,000$ and an average degree $k = 5$. To achieve this average, we use Equation (1) to compute the mean of the power law distribution as

$$\sum_{\tilde{k}=1}^{N-1} \tilde{k} p(\tilde{k}) = (\alpha - 1) \sum_{\tilde{k}=1}^{N-1} \tilde{k}^{-\alpha+1}, \tag{2}$$

where the range on $\tilde{k}$ is chosen so that we do not end up with any unconnected (orphan) nodes or any loops (a node connected to itself). We then set Equation (2) equal to our desired average degree $k = 5$ and numerically solve (using `R`'s `uniroot` function) for the power law exponent α. With this exponent in hand, the network is constructed by distributing the desired number kN of edges among the total number N of nodes with a probability given by Equation (1). Figure 4 shows a smaller ($N = 50$) example of a scale-free network.

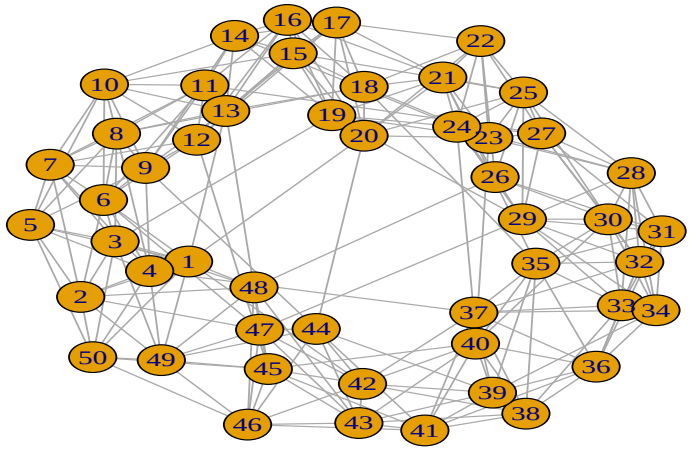

Figure 5: An example of a small-world network with 50 nodes.

2.2.3 Small-world network

Small-world networks are characterized by a short average path length between any two nodes and a high degree of clustering [12]. Again, we choose an average degree $k = 5$ and a total number of nodes $N = 10,000$. Following the algorithm of Watts and Strogatz [12], we start with a one-dimensional ring lattice of size N and randomly rewire each edge with probability p; here, we set $p = 0.5$. The construction is carried out using the algorithms built into the `igraph` package for `R`. Figure 5 shows a smaller ($N = 50$) example of a small-world network.

2.3 The stochastic SIRS model and conversion to an EEG signal

The deterministic continuous-time SIRS model is defined by the following equations:

$$\frac{dS}{dt} = -\beta SI + \varphi R \tag{3a}$$

$$\frac{dI}{dt} = \beta SI - \gamma I \tag{3b}$$

$$\frac{dR}{dt} = \gamma I - \varphi R. \tag{3c}$$

If we assume a fixed finite population size N, then we can recast this as a system of two equations:

$$\frac{dS}{dt} = -\beta SI + \varphi(N - S - I) \tag{4a}$$

$$\frac{dI}{dt} = \beta SI - \gamma I. \tag{4b}$$

However, these equations assume each individual can come into contact with any other individual in the population. If we wish to introduce a network structure, then a neuron will only interact with the other neurons to which it is connected. Since the degree of each node in these networks is a random variable drawn from a probability distribution—Poisson in the cases of random and small-world networks [1, 12], and a power law distribution in the case of scale-free networks [4]—incorporating the network structures adds stochasticity to the SIRS model.

To implement this model, we use a stochastic, discrete-time, discrete-state process with three parameters: the rate β of infection per susceptible neuron per infected neuron; the average period t_{inf} of infection; and the average period t_{imm} of immunity (refractory period). The first decision to be made is what time step should be used in this discretization. Since the sampling frequency of the data is 173.61 Hz, our time step should be no larger than 1/173.61 seconds so that the output contains at least as much information as the data. Shrinking the time step will reduce the noise, but will also increase the computation time of the model. It was found that using time steps smaller than $1/(5 \cdot 173.61)$ produced no noticeable change in the model output, so we use this value as our time step.

The initial values of susceptible, infected, and recovered neurons are set to their respective equilibrium values in the deterministic differential equations, and we keep track of which neurons are in which class at each stage of the process.

Each time step begins by cataloguing the movement of neurons from S to I. To accomplish this, we need to convert the rate β of infection per susceptible per infective into a probability of infection in a single time step. Suppose that $p_S(t)$ is the probability that an initially susceptible neuron is still in S at time t. Then the probability that a single member of I infects a susceptible neuron in time $(t, t+dt)$ is given by

$$p_S(t)\beta dt. \tag{5}$$

It follows that

$$\frac{dp_S}{dt} = -p_S\beta, \tag{6}$$

so $p_S(t) = e^{-\beta t}$. If β_τ represents the probability that a single infected neuron infects a single susceptible neuron in one time step τ, then

$$\beta_\tau = 1 - p_S(\tau) = 1 - e^{-\beta\tau}. \tag{7}$$

To determine whether a susceptible neuron becomes infected in a given time step, we need to know the probability $\tilde{\beta}_\tau$ that at least one of the infected neurons to which it is connected causes an infection. If $\tilde{k}$ is the degree of a given neuron and $\tilde{I}$ is the total number of infected neurons at the start of a given time step, then

$$\tilde{\beta}_\tau = 1 - (1 - \beta_\tau)^{\frac{\tilde{k}\tilde{I}}{N}}. \tag{8}$$

We compute this probability for each susceptible neuron, construct a separate vector of probabilities chosen from a uniform distribution, and those neurons for which $\tilde{\beta}_\tau$ is greater than the corresponding randomly chosen probability will be moved from S to I. Finally, an infection time must be assigned to each of the new infected neurons, i.e., we must choose the number of time steps for which the neuron will remain in the I class. For each new infected neuron, we choose a number of time steps from a geometric distribution (although a Poisson distribution could be and was also considered) with a mean of t_{inf}/τ to use as the neuron's infection time.

After infection has occurred, recovery is tracked. Any infected neuron whose infection time was chosen to end at the current time step is moved from I to R. These newly recovered neurons must now be assigned a recovery time, i.e., the number of time steps for which that neuron will remain in the R class. Similar to how the infection times were chosen, we choose a number of time steps from a geometric distribution with a mean of t_{imm}/τ for each newly recovered neuron.

The final phase of each time step involves moving neurons from R to S. Any recovered neuron whose recovery time was chosen to be the current time step re-enters the susceptible class. Once all time steps have been realized, every fifth step is reported as output from the model so as to match the sampling frequency of the data.

Before attempting to fit the model to the data, we need to transform the raw output into an EEG signal. EEGs measure how the voltage in the brain changes over time; since each neuron state—resting (S), undergoing an action potential (I), and refractory (R)—has an associated average voltage, we can multiply the number of neurons in each state during each time step by these voltages to create an approximate EEG signal. The corresponding voltages are -70 mV in the resting state, 30 mV during an action potential, and -80 mV in during the refractory state [10].

2.4 Constructing, smoothing, and optimizing the likelihood surface

To fit this model to the data, we want to optimize the likelihood surface; however, for very noisy systems (as we have here) this is often impossible to compute directly. An alternative construct, called synthetic likelihood, can be used in such cases [14]. To start, a set of summary statistics—metrics on which to base the comparison between the model and the data—is chosen. For a given vector θ of parameter values, the idea is then to simulate the model n_{sim} times and compute a set $\{s_i\}_{i=1}^{n_{\text{sim}}}$ of vectors of summary statistics [14]. We then compute $\mu = \sum_{i=1}^{n_{\text{sim}}} s_i/n_{\text{sim}}$, $S = (s_1 - \mu, s_2 - \mu, \ldots, s_{n_{\text{sim}}} - \mu)$, and $\Sigma = SS^T/(n_{\text{sim}} - 1)$, and define the log synthetic likelihood as

$$\mathcal{L}_s = \frac{1}{2}(\tilde{s} - \mu)^T \Sigma^{-1}(s - \mu) - \frac{1}{2}\log|\Sigma|\,, \tag{9}$$

where the summary statistics are assumed to be multivariate normally distributed and $\tilde{s}$ is the vector of summary statistics computed from the data [14].

Our choice of summary statistics follows that of Ghorbanian et al. [8, 9]; we use the power spectrum, Shannon entropy, and sample entropy, as defined below.

In the field of neuroscience, there are several common frequency bands that are studied and frequencies outside of these bands are generally considered to be noise [8]. These bands are labelled as: delta (approximately 0.5-4 Hz), theta (approximately 4-8 Hz), alpha (approximately 8-13 Hz), beta (approximately 13-30 Hz), and gamma (approximately 30-60 Hz)[8]. Our summary statistics vectors contain five elements corresponding to the total power in each of these frequency bands.

Entropy is a measure of the information content contained in a signal, which is related to the signal's level of unpredictability [9]. The original (information theoretic) definition of entropy is that of Shannon entropy. Given a discrete random variable X, Shannon defined the entropy of X to be

$$-\sum_i P(x_i)\log_2(P(x_i)), \tag{10}$$

where P represents the probability mass function of X. Here, we take X to be the time series of voltages.

Other methods of calculating entropy have been devised since Shannon first introduced the idea. Ghorbanian et al. report obtaining a better fit to EEG data when they include both Shannon entropy and sample entropy in their summary statistics [9]. Sample entropy measures the unpredictability of a signal by calculating the probability that two segments of the time series similar (as measured by the Euclidean distance between the two segments) for m time steps will remain similar at the $m+1$ time step. We let n_m represent the number of pairs of time series segments of length m with Euclidean distance less than some cutoff value r, and n_{m+1} represent the number of pairs of time series segments of length $m+1$ with Euclidean distance less than that same cutoff value r. Then the sample entropy of the signal is given by

$$-\log\left(\frac{\tilde{n}_m}{\tilde{n}_{m+1}}\right), \tag{11}$$

where $\tilde{n}_m = n_m/((N-m)(N-m-1))$ and $\tilde{n}_{m+1} = n_{m+1}/((N-m-1)(N-m-2))$ [9]. Common choices for m and r are $m = 2$ and $r = \sigma(X)/4$ (where σ represents the standard deviation of the signal), so these are the values we use here [9].

To optimize the synthetic likelihood function, we compute a portion of the likelihood surface over a range of reasonable parameter values, remove any outliers (points farther than three standard deviations from the mean), smooth the surface to reduce the level of noise, and optimize the smoothed surface. In smoothing the surface, we use a generalized additive model (GAM); this is similar to a generalized linear model (GLM) except the response variable is expressed as a smooth function s of the predictor variables [13]. Since we have three model parameters, $s = s(\beta, t_{\text{inf}}, t_{\text{imm}})$ is a multivariate function; furthermore, our parameters do not inherently live on the same scale. These factors mean that s should be what is called a tensor product smooth [13]. To fit the GAM, we need to express s as a linear combination of basis functions. This is accomplished by fitting (univariate) cubic splines $s_\beta, s_{t_{\text{inf}}}, s_{t_{\text{imm}}}$ to each predictor variable, then taking a tensor product of the model matrices for these spline fits and using the resulting functions as the basis functions for the tensor product smooth s [13]. A least squares method is then used to determine the parameters of the GAM fit [13].

Once the likelihood surface has been smoothed, we use the Nelder-Mead algorithm to optimize it (this performs minimization by default, so we optimize over the negative log-likelihood surface). In general, given a starting point $P = (x_1, x_2, \ldots, x_n)$, this algorithm begins by computing $n+1$ points around P and using them as the vertices of a simplex [11]. The vertices then go through a series of reflections, expansions, and contractions, all the while "pulling" the

simplex in the direction of smaller likelihood values; the algorithm terminates once all vertices contract to the same point and that point is returned as the minimum of the surface [11].

2.5 Model comparison

After finding the optimal parameter values for the random, scale-free, and small-world network models across both data sets, we need to compute confidence intervals for those parameters. If $\hat{\mathcal{L}}_s\left(\hat{\beta}, \hat{t}_{\text{inf}}, \hat{t}_{\text{imm}}\right)$ is the minimum log synthetic likelihood, and $\mathcal{L}'_s$ is the minimum synthetic likelihood obtained by keeping one of the parameters fixed and optimizing over the remaining two, then the Likelihood Ratio Test says that

$$2\left(\hat{\mathcal{L}}_s - \mathcal{L}'_s\right) \sim \chi^2_1 \tag{12}$$

where χ^2_1 represents the chi-squared distribution with one degree of freedom [5]. So to find the 95% confidence interval for a given parameter, we fix that parameter to a range of values around its optimum, optimize the (negative) log synthetic likelihood with respect to the remaining parameters, and find the values of the fixed parameter at which the likelihood reaches $-\hat{\mathcal{L}}_s + \frac{\chi^2_1(0.95)}{2}$. Figure 6 shows this idea for the t_{inf} parameter in the random network model of the control data, and Figure 7 shows the same for the t_{inf} parameter in the scale-free network model of the epilepsy data.

Finally, to compare different network structure models within each data set, we use the Akaike information criterion (AIC). AIC is defined as

$$-2\hat{\mathcal{L}}_s + 2\mathcal{K}, \tag{13}$$

where $\mathcal{K}$ represents the number of parameters in the model; smaller AIC values indicate a better model of the data. To compare the models, we look at the difference between each model's AIC value and the minimum AIC value across all models.

3 Results, Discussion, and Concluding Remarks

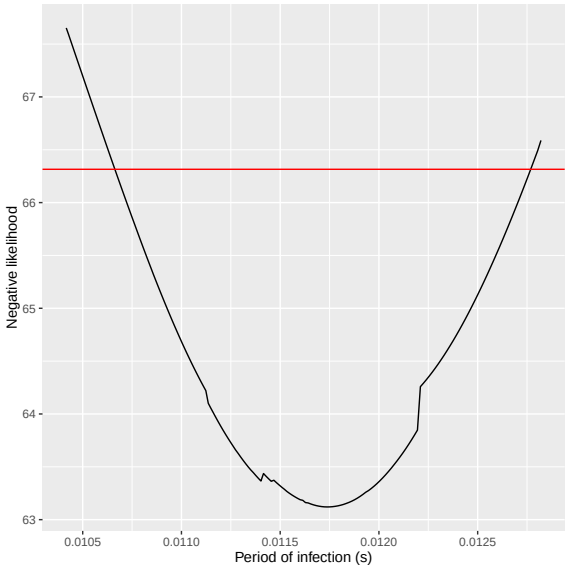

Figure 6: Computing the cutoff values for the 95% confidence interval of t_{inf} in the random network model of the control data; red line represents $-\hat{\mathcal{L}}_s + \frac{\chi_1^2(0.95)}{2}$.

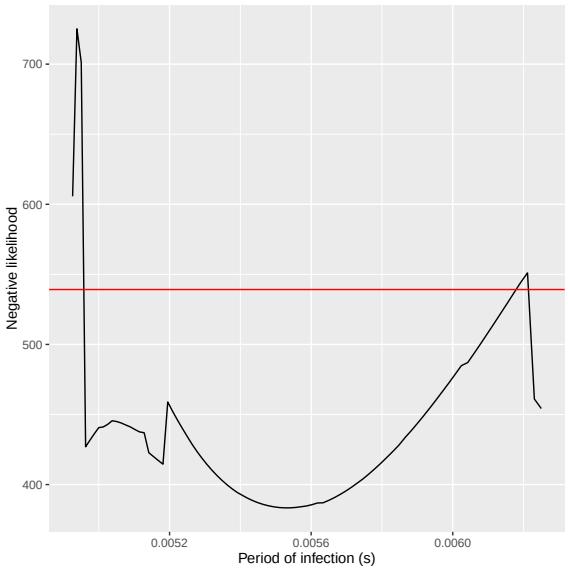

Figure 7: Computing the cutoff values for the 95% confidence interval of t_{inf} in the scale-free network model of the epilepsy data; red line represents $-\hat{\mathcal{L}}_s + \frac{\chi_1^2(0.95)}{2}$.

		Control		**Epilepsy**	
		Value	*95% CI*	*Value*	*95% CI*
Random	β	45.01	(43.30, 49.72)	169.80	(166.02, 172.80)
	t_{inf}	0.012	(0.0107, 0.0128)	0.00524	(0.00489, 0.00526)
	t_{imm}	0.167	(0.1415, 0.1713)	0.1115	(0.1096, 0.1130)
Scale-free	β	53.41	(44.88, 55.58)	174.63	(170.50, 178.16)
	t_{inf}	0.014	(0.0128, 0.0150)	0.00570	(0.005048, 0.006179)
	t_{imm}	0.165	(0.1431, 0.1721)	0.0910	(0.09091, 0.09259)
Small-world	β	62.07	(54.31, 66.77)	207.36	(197.16, 238.31)
	t_{inf}	0.015	(0.0126, 0.0155)	0.00601	(0.00594, 0.00604)
	t_{imm}	0.162	(0.1479, 0.1705)	0.0987	(0.0971, 0.151)

Table 1: Optimal parameter values and their associated 95% confidence intervals.

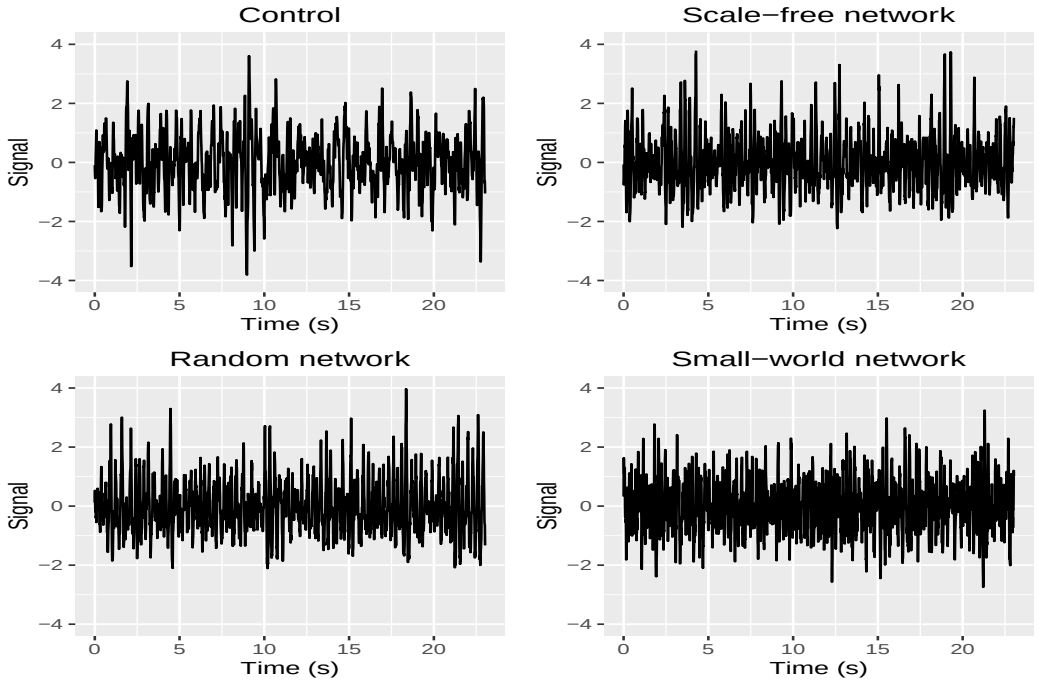

Figure 8: Control data and model output at associated optimal parameters.

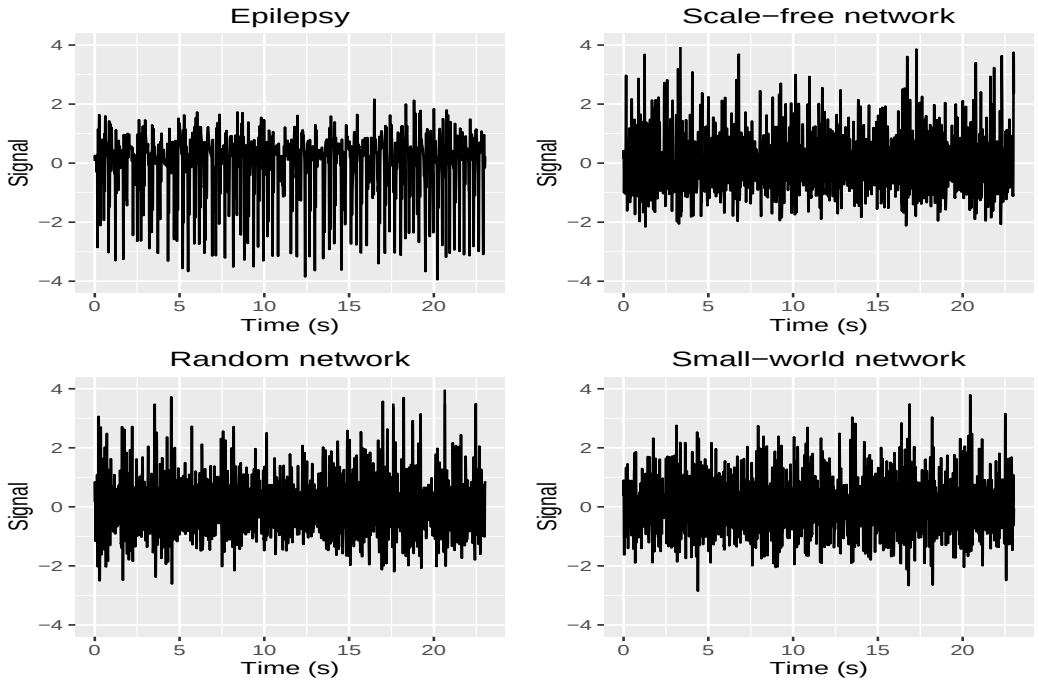

Figure 9: Epilepsy data and model output at associated optimal parameters.

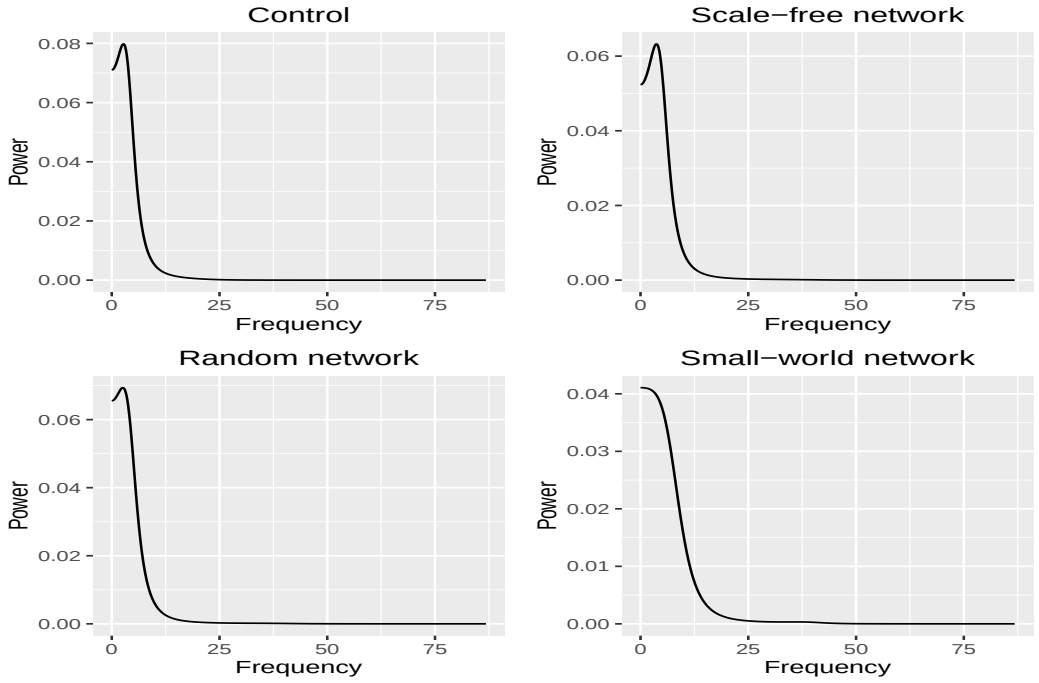

Figure 10: Power spectra for control data and model output at associated optimal parameters.

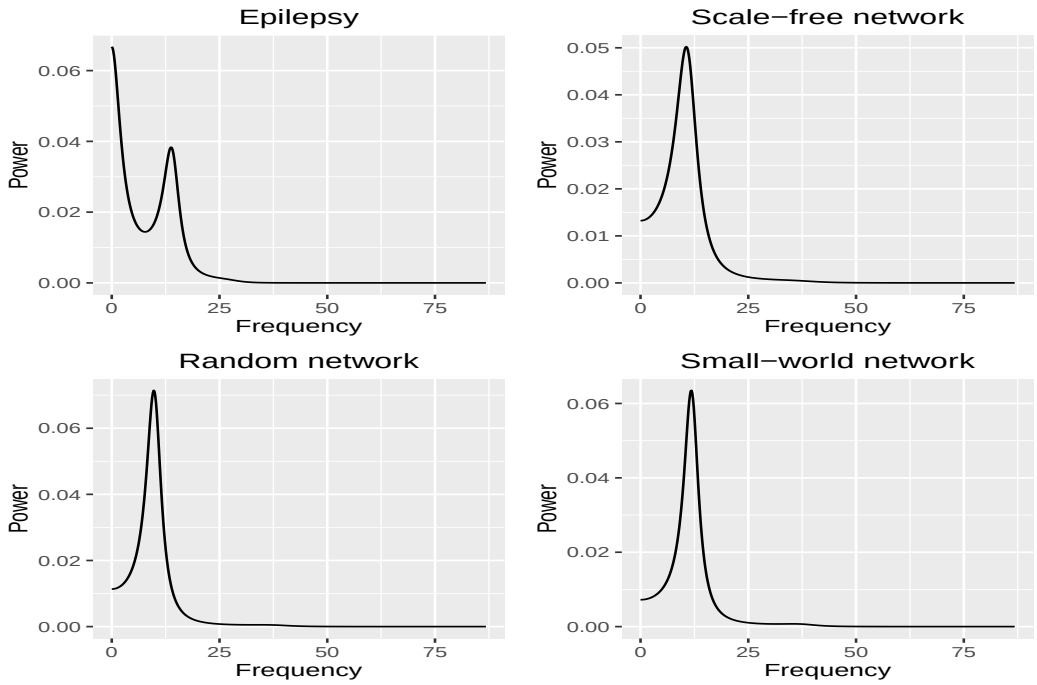

Figure 11: Power spectra for epilepsy data and model output at associated optimal parameters.

The optimal parameter values and their associated 95% confidence intervals are presented in Table 1. Figure 8 presents the control data along with the time series for each network type at the associated optimal parameters; a similar picture for the epilepsy data is presented in Figure 9. The power spectra for the control data and time series for each network type at the optimal parameter values are depicted in Figure 10, and in Figure 11 for the epilepsy data.

We set out to address two questions. First: Can the stochastic SIRS model built over certain network structures distinguish between epilepsy and non-epilepsy (control) EEG data? Second: Does the specific network structure affect how well the model fits the data? Looking at the random network model across both sets of data, none of the confidence intervals overlap with each other suggesting that there are significantly different sets of parameter values that model the control versus the epilepsy data. The same can be said of the scale-free network model. In the case of the small-world network, the confidence intervals for t_{imm} overlap meaning that parameter on its own might not be able to distinguish between the control and epilepsy data. Overall, though, the findings suggest that all three network structures are able to differentiate between the two sets of data.

To answer the second question, we compute the change in the AIC between models within each data set. For the control data, the scale-free network model

had the lowest AIC with the random network having an AIC difference of approximately 20.4 and the small-world network having an AIC difference of approximately 78.8; the difference in AIC between the small-world and random networks was about 58.4. Not only does this suggest that all three network models are significantly different from each other, but it also suggests that the scale-free network does the best job of modelling the control data. A similar pattern holds for the epilepsy data. The scale-free network model again has the lowest AIC value with the random network having a difference of 1596.0 and the small-world network having a difference of 2293.1; the difference in AIC between the small-world and random network models is approximately 697.1. So again, all three models seem to be significantly different from each other and the scale-free network best models the data. (A caveat: AIC assumes that the "best" model fits the data adequately, but none of the models were able to reproduce the second peak seen in the power spectrum of the epilepsy data.)

A few avenues exist for future research in this area. An important feature of small-world networks that was not addressed in this thesis is the rewiring probability (the probability with which each edge in the starting ring lattice is rewired). As this probability increases from zero to one, the resulting network transforms from a ring lattice to a random network [12]. Here, the rewiring probability was fixed at $p = 0.5$, but it would be interesting to explore the effects of adding this as an extra parameter in the small-world network model.

The summary statistics could be examined more closely to try to determine where the models and data agree/disagree the most. Doing so might suggest some changes to the model that could improve the fit to the data.

Another direction in which this work could be taken is to fit the models to larger data sets. In this thesis, we considered only a single control EEG and a single epileptic EEG. If the finding that a scale-free network model provides the best fit holds true over a much larger set of EEG signals, it could potentially give some insight into how the brain is organized.

4 References

[1] L. Acedo and J. A. Moraño. Brain oscillations in a random neural network. *Mathematical and Computer Modelling*, 57, 2013.

[2] R.G. Andrzejak et al. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical Review E*, 64(061907), 2001.

[3] J. Baladron et al. Mean-field description and propagation of chaos in networks of Hodgkin-Huxley and FitzHugh-Nagumo neurons. *The Journal of Mathematical Neuroscience*, 2012.

[4] A. L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286, 1999.

[5] B. M. Bolker. *Ecological Models and Data in R.* Princeton University Press, 2008.

[6] A. N. Burkitt. A review of the integrate-and-fire neutron model: I. homogeneous synaptic input. *Biological Cybernetics*, 2006.

[7] A. Clauset, C. Rohilla, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51, 2009.

[8] P. Ghorbanian et al. A phenomenological model of EEG based on the dynamics of a stochastic Duffing-van der Pol oscillator network. *Biomedical Signal Processing and Control*, 15, 2015.

[9] P. Ghorbanian, S. Ramakrishnan, and H. Ashrafiuon. Stochastic nonlinear oscillator models of EEG: the Alzheimer's disease case. *Frontiers in Computational Neuroscience*, 9, 2015.

[10] B. Kolb and I. Q. Whishaw. *Fundamentals of Human Neuropsychology.* Worth Publishers, 6th edition, 2009.

[11] J. A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7, 1965.

[12] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393, 1998.

[13] S. N. Wood. *Generalized Additive Models: An Introduction with R.* CRC Press, 2006.

[14] S. N. Wood. Statistical inference for noisy nonlinear ecological dynamic systems. *Nature*, 466, 2010.