

Classifying Brain States Using Causal Brain Networks

Kamal Kamalaldin

I. DEFINITION

A. Project Overview

This project explores the possibility applications of unsupervised learning on classifying different brain states in a rat's brain. The topic is in the neuroscience field, and the motivation behind this project is to help detect epilepsy in the human brain. Although there have been approaches to solving this problem by applying neural networks to EEG signals, and a recent competition has been launched on Kaggle that deals with epilepsy prediction, we take a different approach to this problem by relying on a causality model that is not often used in the field of machine learning [1]. This causality model, Sugihara Causality, is used as a means to establish a measure of brain communication patterns [2]. Using this causality model, we were able to build a series of graphs each of which (theoretically) represents a snapshot of brain communication. Since these graphs are highly dimensional, we use PCA to attempt to reduce its complexity before clustering. we then analyze different clustering possibilities and finally observe if clustering gives an intelligible insight into classifying epileptic and non epileptic brain states.

The dataset used in this project has been collected privately by Dr. Zsolt Borhegyi at MTA-TKI , MTA-ELTE-NAP B-Opto-Neuropharmacology Group in Hungary. The original data consisted of EEG recordings of a rat's brain. The processed (graph series) version of the data can be downloaded from <https://drive.google.com/file/d/0BzyCB-i-aKDWU2pBYkVoOV81S3M/view?usp=sharing>. The research conducted in this paper was also funded by the NSF as part of grant 1359275 at the University of Colorado Colorado Springs.

B. Problem Statement

Given a mesh of EEG signals that have been translated into a brain region communication graph series, can we cluster the graphs into meaningful segments that can help us find epilepsy

states. To find out, we plan to use unsupervised clustering in the form of Gaussian Mixture algorithm. We also employ principal component analysis (PCA) to help reduce the dimensionality of the data for faster processing time.

C. Metrics

The unsupervised learning used here along with the high dimensionality of the dataset used in this project bring with them a unique difficulty in terms of defining a clear scoring method. In the scikit-learn library, there are only two methods that can be used to give a score to clusters (assuming no knowledge of *true* labels of each data point): Silhouette Coefficient and Calinsky-Harabaz Index. Both scores ascertain the quality of clustering by measuring how dense, well defined, and separated the clusters are. We are forced to use these metrics due to the nature of the data we are given, as we don't know the true labels of the data points. However, we do know the specific times of when epilepsy was induced, and since epilepsy is *one* state in the brain, then we can use the time indices for epilepsy to be the one cluster we perform an F1 metric on. The F1 score is calculated as

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

$$precision = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

$$recall = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

where positive is epilepsy and negative is non-epilepsy.

II. ANALYSIS

A. Methods

1) Data Collection: EEG data was collected from an 4x8 endodermal electrode array at a frequency of 1000 Hz (31 channels, one channel malfunctioned, Fig. 1). The data recording lasts for 500 seconds, during which epileptic seizures were evoked using 4-aminopyridine and EEG data was recorded from the electrode array. The spiking voltage of each recorded electrode is used as a signal and is referred to as a channel (Fig. 2). Each channel has a continuous data series that spans the 500 seconds.

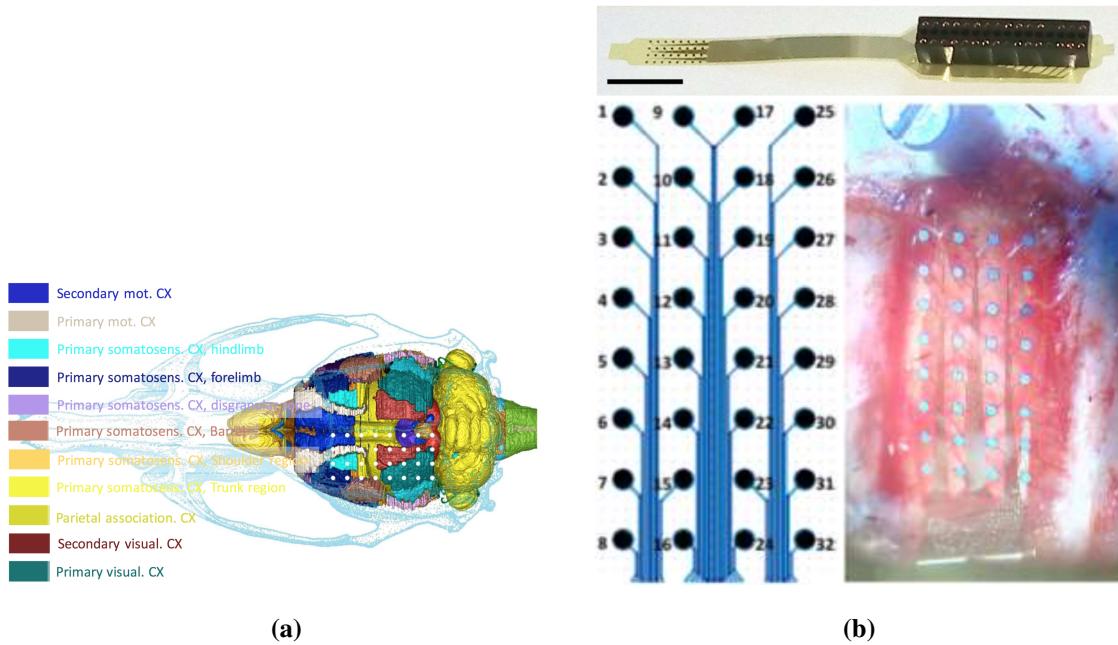


Fig. 1: (a) A composite drawing showing the brain in the scull. The 3D reconstruction of the brain has been made using the maps of the Paxinos atlas, and the localization of cortical areas are indicated by different colors. White points indicate the position of the recording sites of the membrane electrode. Names for the cortical areas are also shown (based on Hjornevik *et al.* and Paxinos *et al.* [3] [4]). (b) Photograph of a membrane electrode shows the construction on the top, the numbering (bottom right) and the surgical implantation (bottom left) is also shown. Electrode 1 malfunctioned during recording.

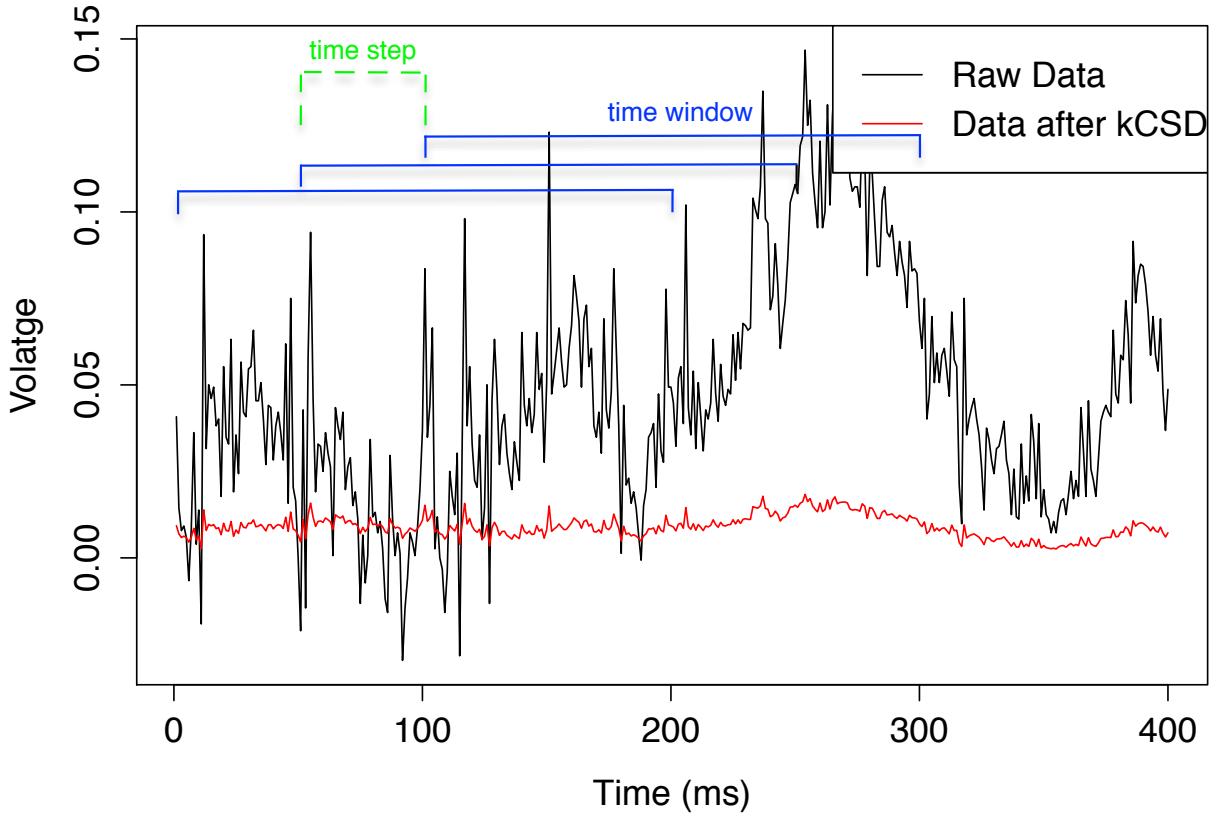


Fig. 2: Sample data from electrode 1, in both its raw form and then in its form after kCSD transformation. Applying kernel Current Source Density extracts the true relationship of neural activity, and in theory eliminates disturbances that might be caused by signal interference. Three sample blue lines are drawn to demonstrate the sliding windows that were used to calculate Sugihara Causality between brain regions. In each time window, the causality was calculated between each brain region, and the time window was moved by an amount *time step*. Different configurations were used for the time window and time step, as reported in table I. Here, the configuration shown has time window as 200 and time step as 50.

TABLE I: Configurations used for the sliding window when calculating Sugihara causality that was used as a measure of brain communication. These configurations inform our model of the granularity of brain communication. Since we do not know exactly how quickly information is being passed down from region to region, arbitrary choices were made to test how well each performs and make a more informed decision in future research. Because we have different time steps, there are a different number of graphs (data points) in each dataset.

Dataset name	Time window	Time step	Number of Data points
dataset 50	200	50	10190
dataset 200	200	200	2547
dataset 250	2000	250	2038

2) *Data Preprocessing:* Kernel Current Source Density (kCSD) method was used on the grid of channels to account for possible electrical interference with the direct measurement (Fig. 2) [5]. The measured potentials produced by kCSD arise as the linear combination of the transmembrane currents, which is a more direct and localized quantity to measure the neural activity. Therefore current source density distribution was calculated by the kCSD method and used for the analysis instead of the original EEG signals.

Afterwards, in order to reduce the data size that is operated on, we lumped channels from the same brain regions together by averaging them (Fig. 3). In addition to reducing dimensionality, this process also puts emphasis of causality on functional brain regions instead of a local cluster of neurons.

3) *Sugihara Causality Measurement:* After lumping the kCSD signals from the same region, we perform Sugihara Causality calculation on the regions by using a sliding windows method (Fig. 2). Sugihara Causality is discussed in detail in the appendix, and any discussion or reference can be obtained there. Because we do not know how granular of a time frame the brain uses to communicate (and how that affects the EEG signal we record), we use three different time scales as reported in table I. This produces 3 datasets with similar characteristics. The important thing to keep in mind is that after this step, each causality measure that resulted from the Sugihara Causality analysis is a measure of how much one region is affecting the other, and we use that as a way to describe brain communication and information flow. Each causality measure is therefore an *edge* of communication from one region to another, meaning that each brain region is a node and each causality measure is a weighted edge. Figure 4 shows a sample graph constructed from joining all the causality measures from a specific time window. This is essentially a (theoretical) snapshot of how each brain region was affecting other brain regions in that time window.

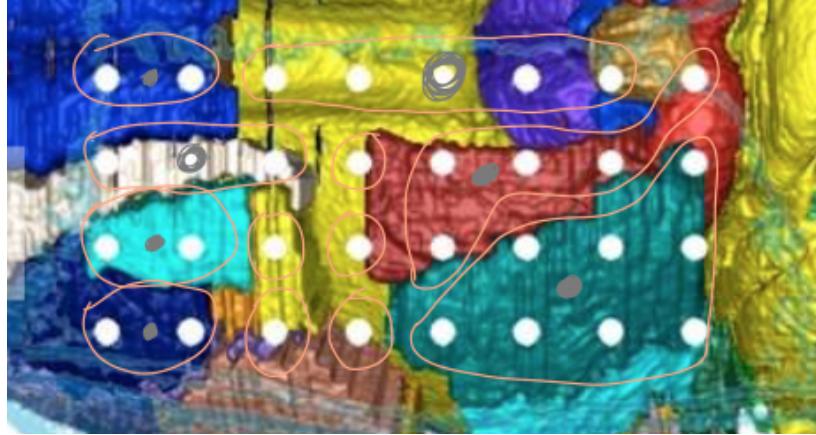


Fig. 3: The distribution and lumping of the brain regions in the brain. A total of 12 region channels were constructed from the initial 31 local channels. The schematic is based on rat brain atlas mapping.

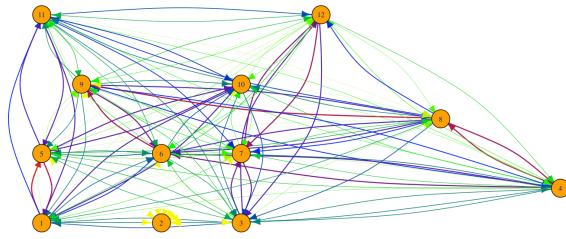


Fig. 4: Sample graph constructed from calculating the Sugihara causality between brain regions. Lumped brain regions correspond to the gray circular mapping in Fig. 3. Edge colors represent the strength of the causal relationship. From weakest to strongest: Yellow, Green, Blue, Red. The edge weight ranges from 0 to 1.

B. Data Exploration and Visualization

The data we produced is a series of complete, directed graphs (all the same topology; each node is a brain region as seen in figure 1; each edge is a causality/communication measure) and their edge weights that we refer to as ρ scores (ρ is the convention name for the score of a Sugihara Causality analysis [see appendix]). Since we processed the data with different time steps, there are a different number of graphs available in each dataset (Tab. I). For each

graph, there are 132 ρ scores, all of which range between 0 and 1 (decimals). Each graph was transformed into a vector where each vector is a data point, and all vectors make a dataset. This transformation was required for the data to be entered into sklearn's Gaussian Mixture model function properly.

In all graphs, most of the ρ scores appear to be of high values, suggesting high connectivity between brain regions and therefore a high density graph (Fig. 5b). This indicates that there is constant information flow between regions in the brain. Furthermore, we also notice that many of the ρ scores to be similar within a pair of nodes (Fig. 6). For example, when looking at regions 1 and 2 within the first 200 milliseconds, they appear to be equally communicating to one another (Fig. 5a). This could be indicative of either bidirectional causality or unidirectional forcing, both of which can be detected by the Sugihara model.

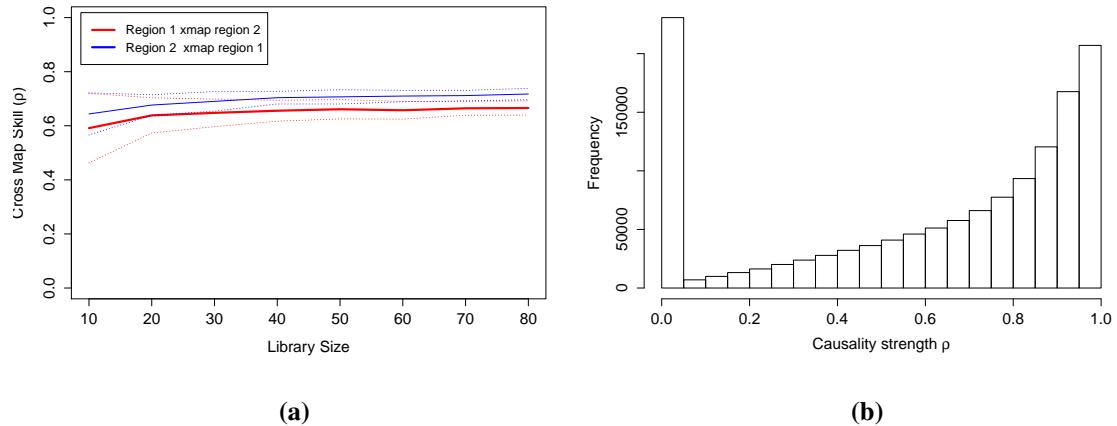


Fig. 5: (a) A sample graph plotting the convergent cross mapping skill (ρ) between region (nodes) 1 and 2 during the first 200 ms of the experiment. The skill mapping channel 1 to channel 2 is very similar to the one mapping channel 2 to channel 1. This might infer either bidirectional causality or unidirectional forcing. A similar pattern (close ρ value between pairs) was found for most of the pairs of edges. (b) The distribution of CCM skill (ρ) during the entire experiment. Many relationships appear to be causal in the brain, with equally as many being non-causal throughout the experiment. Causality was calculated from signals of lumped regions after calculating the CSD. Cross mapping was done on every pair of regions with library size of 80, and each pair has two causality directions. Sliding windows of 200 ms were used, with a sliding step of 50 ms.

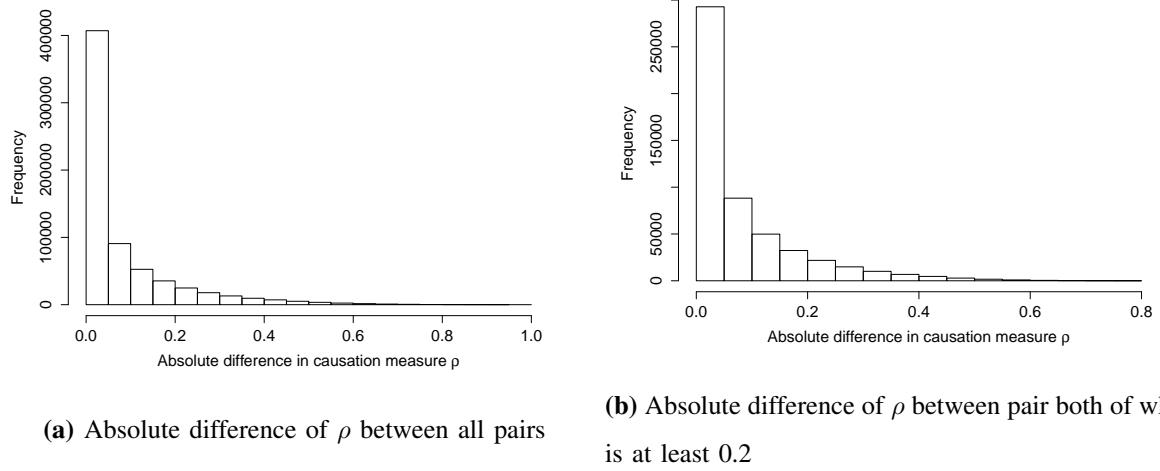


Fig. 6: Difference of causality values within a pair is small, even when accounting for non-existing relationships where both causalities are below 0.2. This similarity between directions of causality in a pair could imply a bidirectional relationship between most regions, or could alternatively imply a unidirectional forcing (synchrony) phenomenon. Sliding windows of 200 ms were used, with a sliding step of 50 ms. A library size of 80 was used.

C. Algorithms and Techniques

The machine learning algorithms we use in this project are all centered around unsupervised learning, specifically clustering using an altered version of K-means called the Gaussian Mixture model. K-mean algorithms require as input a number n of clusters to find, and a set of data points on which the clusters are to be found. The algorithm then works by choosing n random points in the data set and setting them as centers of the clusters. Each data point that is not a center then becomes in the cluster of the nearest center. The centers are then incrementally reassigned to be the mean of the points in the clusters. This method converges in finite time to a solution, and most of the time (depending on the random initialization) finds good clusters that are well separated and in which each cluster has data points that share common features.

Here, we choose the Gaussian Mixture algorithm as implemented by the Sci-kit learn python library instead of the simple K-means algorithm. Gaussian Mixture Model (GMM) clustering can be thought of as a K-means algorithm that is generalized to account for the covariance of the data and centers of the Gaussian distributions. Because of this feature, while K-means algorithm only clusters data points in spherical shapes, the GMM model allows clusters to take many shapes as described below under the name covariance type. For these reasons, the GMM would be a more suitable algorithm to use in our case. GMM has a few parameters that could

be tuned, the most important of which is the covariance type (full, tied, diagonal, or spherical) which dictates how the cluster edges are expanded and how the clusters grow. This parameter will be fine tuned using grid search.

Furthermore, we use Principal Component Analysis (PCA) to reduce the dimensionality of the dataset. This reduction is important because we have 132 features, and reducing them would result in a simpler analysis of the dataset. PCA works by finding the n axes of most variance in the dataset and transforming the dataset into a space of only these axes. This essentially describes the dataset in a new, lesser set of dimensions than the original dimensions.

Another algorithm that is used is Grid Search, used for fine tuning the Gaussian Mixture model. The grid search algorithm is much simpler than the previous algorithms, since all it does is try to create a model with different parameters and scores each model using a given scoring function. The grid search object then ranks the models based on their scores, and can be used to retrieve the best parameters (the ones that produce the highest score).

D. Benchmark

Since this method has never been attempted before (using Sugihara Causality graphs to identify epilepsy), providing a succinct and complete benchmark is a difficult task. Furthermore, the clustering here is used to explore whether or not there are any extractable information from causal networks in the brain, and whether the brain communicates differently when it is suffering from epilepsy compared to when it is operating normally.

Keeping this in mind, we rely on the simple F1 score to provide a simple reliability measure of how well the clustering is separating and predicting the epilepsy cluster. Since we know that epilepsy doesn't happen often in our dataset (only 20 seconds of epilepsy related events in our 500 seconds dataset), we can assign the cluster with the least data points as the epilepsy one, and then measure how exclusive the epilepsy cluster is to the times when epilepsy was *actually* happening by measuring an F1 score. Since our epilepsy happens for about 4% of the dataset, then the baseline score we expect should be 96% of accuracy and 0.99 F1 score. These scores are attained if all the data points were labeled as one (non-epilepsy) cluster.

III. METHODOLOGY

A. Data Preprocessing

Some of our edge weights had some Not A Numerical values, and those were altered to 0, which is the lowest ρ score possible. Once that error was corrected and the data has been put in the graph series format, separated into the three different datasets, we perform Principal Component Analysis (PCA) to reduce the dimensionality of the datasets. This both gives us insight into what components of the graph are logically important for the model and helps decrease the time required to run our model. The results of PCA was used to create three different datasets from each dataset: a dataset that covered 80% of the variance, a dataset that covered 90% of the variance, and a dataset that covered 99% of the variance. This was done to test how dimensionality reduction would affect the final clustering efforts. Furthermore, there were some edges that showed a consistent weight of 0 (for example, all edges from node 2 to node 1 were of weight 0), and having that redundant data is not helpful to distinguishing states. PCA would help us remove such unvarying features from the datasets.

If the scoring for the reduced datasets with only 80% of the variance is similar to the scoring of the reduced datasets with 99% of the variance, then we use the less dimensional reduced datasets even though they give us a less accurate version of the original information. The original dimensionality of the datasets was 132 features per data point, and a list of the PCA datasets and their different variance levels with component numbers are reported in table II. After the principal component analysis was done on the datasets, the components required to reach the threshold levels of 80, 90, and 99 percents were recorded in table III (Fig. 7).

TABLE II: Dataset names for the result of PCA analysis. Three different reduced datasets are produced each of which keeps a certain percentage of the original variance. Since we started with three datasets and created three new ones out of each by applying PCA with different levels of components, we arrive at 9 total reduced datasets.

	80% variance	90% variance	99% variance
dataset 50	reduced 50 80	reduced 50 90	reduced 50 99
dataset 200	reduced 200 80	reduced 200 90	reduced 200 99
dataset 250	reduced 250 80	reduced 250 90	reduced 250 99

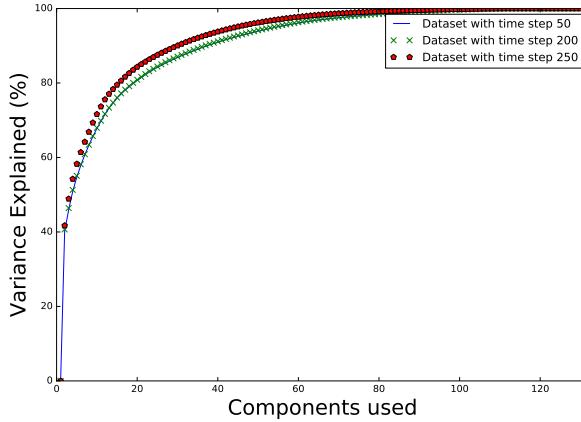


Fig. 7: Principle Component Analysis conducted on the three datasets reveals that the overall variance of the data can be maintained by using only a subset of the edges of the graph rather than all of them. The dataset with a time step of 250 ms needs less components to explain the variance than the other two, while the datasets with 200 and 50 ms time steps showed an almost exact level of variance explained with the number of components. This suggests that the 250 ms dataset is not complex enough, which might indicate that 2000 ms is too big of a time frame to record the granular communication speed of the brain.

TABLE III: Different component numbers are needed for each dataset to maintain a certain level of variance explained using PCA for dimensionality reduction. These values for components required for reaching a score standard were used to produce reduced forms of the datasets as noted in table II. These reduced forms of the datasets were then analyzed in figure 8.

	80% variance explained	90% variance explained	99% variance explained
Components in dataset 50	20	38	87
Components in dataset 200	20	37	87
Components in dataset 250	16	30	76

B. Implementation

Once the dimensionality of the data was reduced with PCA, a Gaussian Mixture clusterer was initiated for each dataset from table II, and all number of clusters from 2 to 40 were tried and tested by the Silhouette Coefficient and Calinsky-Harabaz Index scorers (Figs. 8, 9).

The complete ordered list of procedures was:

- Removing NaN's with 0's if any exist in the edge weights
- Splitting the dataset into three datasets depending on the time step used

- Vectorizing the graphs so that each graph is a data point
- Running PCA analysis on the three datasets with varying degree of variance kept, creating 9 total datasets
- Clustering the PCA-reduced datasets with cluster numbers from 2 to 40.
- Computing Silhouette and Calinsky scores on the clusters in the previous step

The procedures were implemented successfully, but computing the silhouette and Calinsky scores was a bottleneck as it required calculating the distances between each data point. This could possibly have been avoided by pre-calculating the distances between all data points and passing them to the silhouette and Calinsky scoring functions, a feature available in the sci-kit learn library.

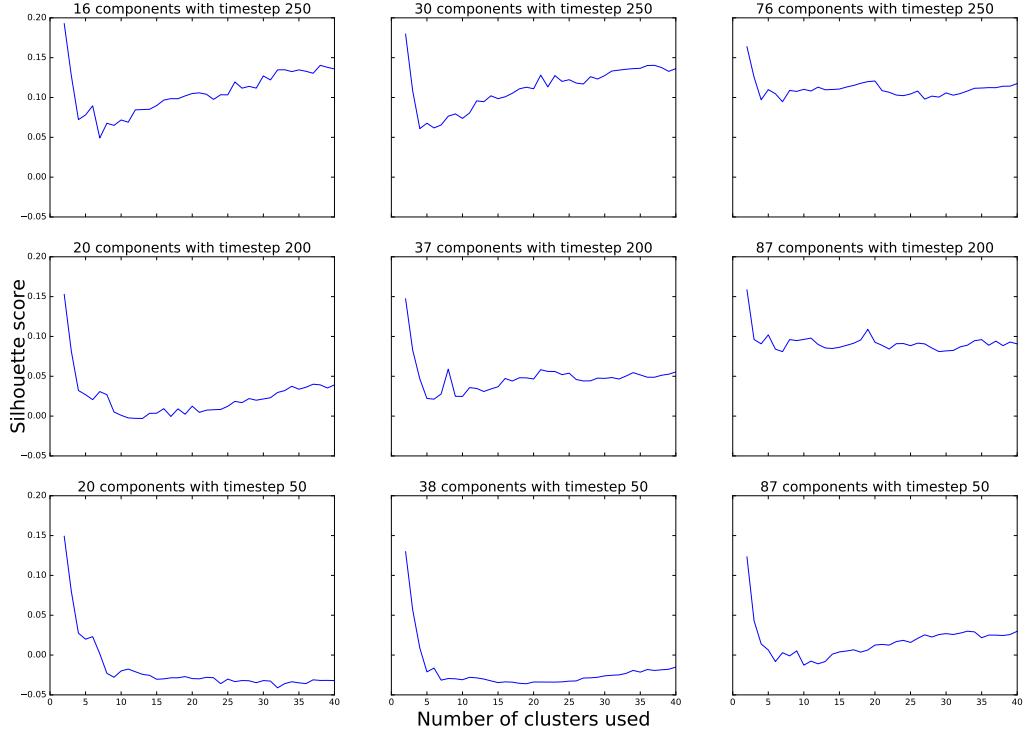


Fig. 8: Silhouette scores for clustering with different cluster numbers. Scores close to 1 suggest clusters that are dense and well separated; scores close to -1 suggest clusters that overlap and are not dense. The best score is consistently when there are only two clusters, and dips greatly when creating more clusters but begins to rise after 30 clusters. With the dataset 250 and dataset 50, PCA with less components (16 and 20 components respectively) showed a better silhouette score than the other, higher component reduced datasets. However, dataset 200 showed a better silhouette score when more components were used. Especially with a rather thin margin of difference in scores between components, it is difficult to conclude whether having more PCA reductionist approach is beneficial.

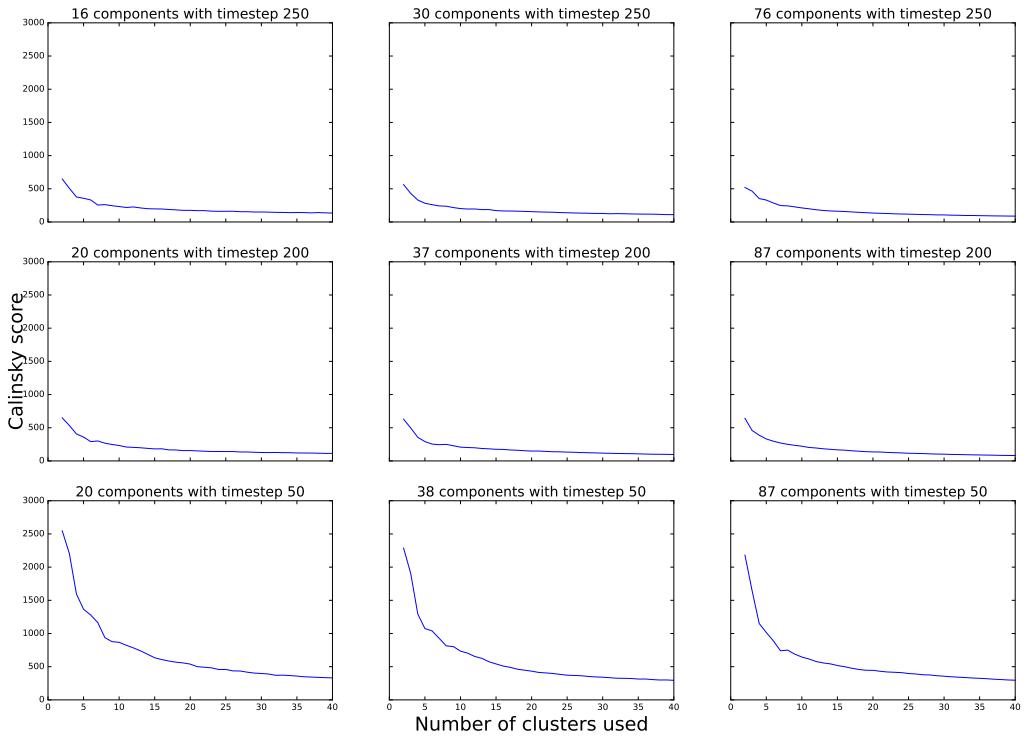


Fig. 9: Calinsky-Harabaz scores for clustering with different cluster numbers. Higher scores mean more separated and dense clusters, which is positive. The score dips greatly when creating more than two clusters, but begins to rise after 30 clusters. The silhouette scores show that having more components when reducing with PCA results in better clustering scores. The scores also show that scores for the reduced data that keep 99% of the variance (right column) are all similar in score and trend.

C. Refinement

From figures 8 and 9, it can be seen that the best clustering number is two, and it happens with dataset *reduced 50 80* (dataset that is made with time step of 50 and transformed with PCA). We therefore fine tuned a Gaussian Mixture model with the *reduced 50 80* dataset using Grid Search as implemented by the scikit learn library. The silhouette score was used as a scoring function for the grid search performance measurement. The parameters tuned with grid search was the *covariance type* (with a choice of tied, diagonal, sphere, or full), and the best parameter was 'tied' with a silhouette score of 0.254. This is an increase of 0.05 in the silhouette score, which is a 25% improvement.

IV. MODEL EVALUATION AND VALIDATION

The results so far indicate that the rat’s brain showed the presence of two distinct brain states, as we can cluster the brain’s communication network into two clusters. This could be good news, since we can assume that the two brain states are *epilepsy state* and *normal state*. This simplifies our choice of metric to evaluate the model since we can use a simple F1 measurement by assuming the less frequent brain state is the epileptic state, and record the less occurring cluster as the epileptic cluster; we then test the time indices of this cluster with the time indices we were given for when epilepsy was evoked. Since we know that the epilepsy happened between seconds 447-457 and 250-260, we set true labels to be 0 for all data points that were in that time period, and 1 for all other data points. We then set the less frequent cluster to be of label 0 (epilepsy cluster) and the more frequent cluster to 1 (non-epilepsy cluster). Our evaluation metric then becomes the F1 score of the predicted labels against the true labels that we produced given the knowledge of our experiment.

Using the fine-tuned model, we achieve an F1 score of 0.788, which seems acceptable at first, but is rather disappointing when considering that the blind benchmark discussed in section II-D has a score of 0.99. This is mainly due to the fact that the epilepsy state is infrequent, and so a clustering of non-epilepsy state as epilepsy creates a great error. This indicates that our model does not perform better than blindly guessing that there is only one state in the brain. This does not align with our hopes and expectations for the clustering solution.

V. CONCLUSION

A. Free-Form Visualization

To analyze what our model is doing, we plot an annotated time strip that shows what each time window of causality was clustered to (Fig. 10). From this graph, it can be seen that there is no correlation between what we observe in the real-life experiment and our clustering on the graph series constructed from Sugihara Causality measures.

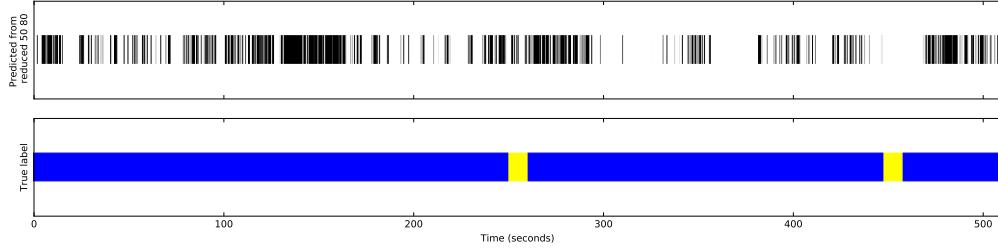


Fig. 10: Time strips that show the epilepsy time track and the results of the respective clustering done on the Sugihara Causality graphs from those time windows. Blue corresponds to true non-epileptic state; yellow to true epileptic state; white to predicted epileptic state; and black to predicted non-epileptic state. There are only two clusters in the top bar. The clustering results does not show any correlation with the actual epilepsy induction time series.

B. Reflection

This project transformed 32 EEG electrode signals into 12 regional neuronal activity signals, and created a (theoretical) communication network from those signals using Sugihara Causality. These communication graphs were analyzed by unsupervised learning using the Gaussian Mixture model for the goal of gaining some insight into the epileptic state, and test whether unsupervised clustering is able to discriminate between brain states from the communication networks. From our initial results, it is clear that our approach is not successful in distinguishing brain states in a rat's brain. However, the ground work has been laid for future explorations in the topic of communication networks in the brain.

C. Improvement

Since various improvements can be made to this work, we break it down by stage of implementation.

1) *Methods:* It would have been enriching to our knowledge of our true labels if we had access to exactly what the rat was doing during the experiment. A video capture of the rat, or recorded structured interventions to the rat's behavior could have provided more opportunity to label the data with ground truth. For example, we could have shone light into the rat's eyes, and recorded that as a separate brain state; we could have put the rat on a hamster wheel and recorded his running as a separate brain state; eating food could be a brain state; and sleeping is another possible brain state. All these interventions to the rat's environment could enable us to

gain more labels rather than the simple epilepsy, non-epilepsy labeling system we worked with. Although this complicates our approach, it would allow us access to better scoring metrics other than the silhouette and Calinksy-Harabaz scores, where the labels have to be known.

2) Data Processing: Apart from collecting the data and applying Sugihara Causality on the dataset (which took 2 days), calculating the current source density using kCSD was by far the most time consuming process. Even though we do not discuss the details of how that was done in this project (we provide already processed kCSD data), it would have been nice to see a GPU implementation of the kCSD, which took advantage of CPU parallelism but not GPU parallelism. This would massively improve the speed of the implementation and allow for testing more various approaches in the Sugihara Causality step.

3) Cluster Scoring with Higher Dimensions: Due to the high dimensionality of the dataset we worked with, it was difficult to visually measure how well the clustering with different number of clusters took place. Beyond three dimensions, it is impossible to visualize the dataset with labels, and although the PCA reduction resulted in a greatly reduced dimensional feature, it was not enough to allow us to visualize the clusters visually.

ACKNOWLEDGMENT

Special thanks goes to Dr. Somogivari and Dorrottaya R. from Winger RCP, Budapest for providing the neuronal data used in this paper. Furthermore, Dr. Peter Erdi is thanked for his patient guidance of this research topic which was instrumental for the production of this paper.

APPENDIX

The concepts of abstract correspondence, correlation and interpreting causation has been discussed in philosophical literature at least as early as Berkley's and Locke's arguments on human perception [6] [7]. Until now, the debate focused on what constitutes a causative effect and how such an effect might be discerned. From philosophy, the debate has moved to empirical science, where different models of causality have been proposed, none of which has been declared the true standard. Causality is often mistaken for correlation because the relationship between the two is not always clear. For example, falling down the stairs could be correlated with breaking a limb, but is falling down the stairs caused by breaking a limb, or is breaking a limb caused by falling down the stairs? One might argue that a person wouldn't break a limb without falling down the stairs, but what if a limb was stressed enough that it broke while a person was normally walking down the stairs, causing the person to lose control of their leg and therefor fall? In this sense, causality is a murky subject that is often considered to be highly subjective and at times nondetachable from the context of the situation. With that being said, there have been many attempts to address the concept of causality in mathematics in the form of the causality of continuous signals on one another. For example, if one could attain the stock prices for hot dogs and hot dogs buns, one would be able to hypothetically see a correlation between the sales of hot dogs and hot dog buns. It is also clear that one can describe a logical causation between the two stock prices, since the products are complementary to one another: if the price for one goes down, the price for the other goes down as well. For the case when substantial fluctuations happen in the stock price for both products, how would one determine which product caused the change in the price of the other? Surely the price drop was caused by *something*, so which of the products caused the other? This is where causality models come into play—this is an example of the use of causality models in the field of Economy.

A particular causality model, Granger Causality (GC), has been widely used in application in the econometric fields [8], and has been the de facto model when causality is concerned. The way Granger Causality works is beyond the scope of this paper, but suffice it to say that Granger Causality behaves best in linear, stochastic systems. However, this causality model carries its own limitations, as, even with extensions to non-linear systems, it has generally not been seen capable of inferring causality in deterministic systems where feedback loops and nonlinearity are a defining feature. New models of causality have been introduced to attempt to go beyond these

limitations. Dynamic Bayesian Networks and, more recently, the Convergent Cross Mapping (CCM) are some such models. The CCM model, also called Sugihara Causality model, relies on the convergence of distance of nearest neighbors in the shadow manifold of pairs of variables [2]. To understand what all this means, we must first discuss manifolds (also called attractors).

A. Manifolds

Given three time series, a manifold of those three series is a new, three-dimensional representation of the state of the three variables at each time point. If the time series are somehow related to a specific concept, then this concept is called a *system*, and the manifold then represents the state progression of the system. To illustrate this, a common example is used, namely the Lorenz Attractor. The Lorenz system is made up of three time series that are governed by the following differential equations

$$\frac{\partial x}{\partial t} = \sigma(y - x) \quad (1)$$

$$\frac{\partial y}{\partial t} = x(\rho - z) - y \quad (2)$$

$$\frac{\partial z}{\partial t} = xy - \beta z \quad (3)$$

where σ , ρ , and β are constants [9]. With an initial condition and some time steps, we can produce three time series like in figure 11a. Given these three time series, a manifold M can be constructed by creating a new time series from the combination of these three series such that each time point t in the manifold, M_t , is described as

$$M_t = (x_t, y_t, z_t).$$

Of course, since this manifold has information about all the three previous time series, it is more highly dimensional than the time series themselves. In this case, since the manifold is of three time series, it is three dimensional (Fig. 11b). Furthermore, since the manifold completely encapsulates the information about the system, the time series x , y , and z can be generated if one has access only to the manifold of the system.

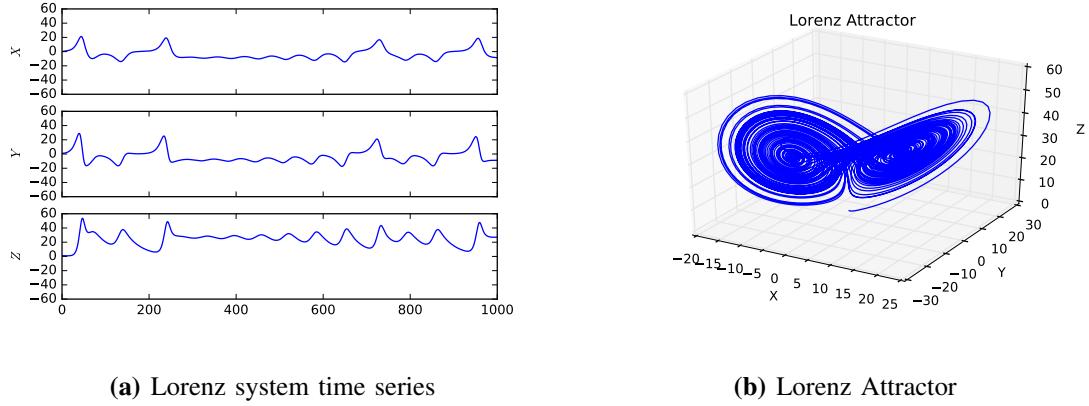


Fig. 11: A Lorenz system with initial conditions $x = 0$, $y = 1.0$, $z = 1.05$, $\sigma = 10$, $\rho = 28$, $\beta = 2.667$, and time differential of 0.01. The series represent interlinked components in a system, and the manifold is made up of all the three series combined. The manifold is constructed by mapping the combination of x , y , and z at each unique time point to a coordinate in an n -dimensional space, where n is the number of components— $n = 3$ in this case. The resulting mathematical n -dimensional object is called a manifold, and can be used to describe the state of the system in terms of complexity, flux, and stability. In this given example, it can be seen that the system gravitates towards two main "loops", and therefore the system is *attracted* to those points in the three dimensional space.

B. Shadow Manifolds

A shadow manifold of time series ω is an E dimensional reconstruction of E delayed signals of ω . For example, if our ω variable is actually the *sin* function, then to construct E delayed signals of ω we would calculate

$$\begin{aligned}\omega_1(t) &= \sin(t) \\ \omega_2(t) &= \sin(t - 1) \\ \omega_3(t) &= \sin(t - 2) \\ &\vdots \\ \omega_E(t) &= \sin(t - (E - 1))\end{aligned}$$

and then to construct a shadow manifold of ω we would construct a manifold of ω_1 , ω_2 , ..., ω_E (Fig. 12). Since the scalar by 1 increment seems too general, the signals are delayed by

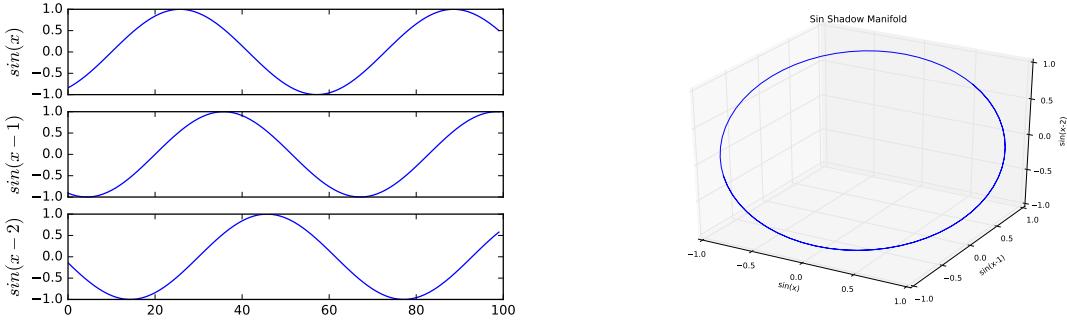


Fig. 12: A shadow manifold of a simple sin function. The delayed signals are created by delaying the series by a scalar constant τ (in this case $\tau = 1$ and $E = 3$ so the function is delayed by 1, 2, 3). From these delayed signals, a manifold is constructed by mapping each time point of all the delayed signals to an E -dimensional manifold called a shadow manifold.

multiples of τ , where τ is approximated by field knowledge or mathematical analysis. Therefore, the formal way to describe shadow manifolds is that they are signals delayed by a scalar multiple of $E\tau$ such that shadow the manifold of ω , M_ω , is described as

$$\begin{aligned} M_\omega &= f(\omega_1(t), \omega_2(t), \omega_3(t), \dots, \omega_E(t)) \\ &= f(\omega(t), \omega(t - \tau), \omega(t - 2\tau), \dots, \omega(t - (E - 1)\tau)) \end{aligned}$$

By using a theorem called Takens' embedding theorem, it can be shown that each shadow manifold of a variable is a projection of the dynamic system's manifold, M , that preserves the topology of M [10], [11], [12]. For example, in a dynamic system like the Lorenz Attractor where the dynamics of each variable is affected by the other variables in the system, it can be said that each variable subscribes to the overall dynamic of the system. Therefore, the state of one variable could be used to infer the state of another variable if they are dynamically linked. A visual demonstration of this can be seen in Figure 13 where the shadow manifold of each of the variables in the Lorenz system maintains part of the topology of the original whole system. This important property of shadow manifolds imply that there is a one-to-one relationship between a shadow manifold and the original system manifold. More importantly, this also implies that there is a one-to-one relationship between the shadow manifold of one variable and each of the other shadow manifolds in the system. Therefore, the relationship between each shadow manifold is strong. A great demonstration of this concept can be viewed in the online (Science) version of

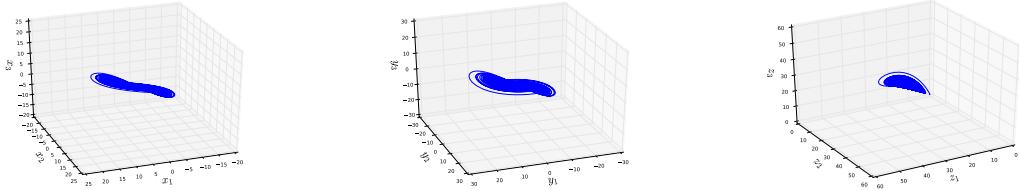


Fig. 13: The shadow manifolds of variables x , y , and z in the Lorenz system. Each of the shadow manifolds retains part of the topology of the original system, as proven by Takens' theorem [12].

the original paper by Sugihara [2] where a video is present to elegantly describe this concept (for ease of access, the reader can also refer to a youtube video called "Takens' theorem in action for the Lorenz chaotic attractor" [13]).

Using this relationship between shadow manifolds, the Sugihara Causality is calculated as the distance between points of similar time point on each shadow manifold. For variables x and y , the causality at time t is calculated as a measure of the time difference between closest neighbors of y_t . This concept is called cross-mapping, and is illustrated in the supplemental materials on the original paper, and the youtube video referred to previously [2], [13]. The last portion of the Sugihara model is the Convergent in Convergent Cross Mapping, which refers to the convergence of the causality value (cross mapping score) when the library of reference points in x and y , L , is considered. The library L simply refers to the specific range of x and y values to be used in the cross mapping calculation. For example, L could be 20, which means only 20 time-synced data point observations of x and y are taken into consideration, or it could be 100, which means 100 x and y variables are considered. The Sugihara model expects for a causal relationship to *converge* to a value as L increases. This implies that L needs to be sufficiently large to allow an observation of convergence. This convergence is the test used to determine Sugihara Causality, named after its author who describes it as a required but not complete definition of causality [2]. This approach is the first step towards more general and applicable causality models since Granger Causality. Since the introduction of Convergent Cross Mapping, it has been shown to be successfully predictive in biological [14], [15], [2], [16], [17] and cosmological [18] applications while showing weaknesses in others [19].

Extensions to and amalgamations of the CCM model are beginning to surface in literature. Clark *et al.* proposed an extension to CCM that relies on measuring the smoothness of the

mapping (also called flow) function ϕ , thereby reducing the L length requirement[20]. Wismller *et al.* proposed a Mutual Connectivity Analysis framework for the "analysis and visualization of non-linear functional connectivity in the human brain from resting state functional MRI" [21] which relies heavily on CCM. Tajima *et al.* use the fundamental idea of state space reconstruction to find two measures. The first is *Complexity* which is the best embedding dimension for a certain signal (embedding dimension at which the cross mapping is saturated). The second is *directionality*, the difference in cross map skill or embeddedness between two a pair of signals. With those two measures, they show that the brain exhibits different complexities during conscious and unconscious states. Here, we explore the application of CCM in estimating the causality between neuronal regions by constructing a network of pairwise causality. We then analyze features of such networks during normal and epileptic seizure periods, and use this in conjunction with machine learning to observe if Sugihara Causality can unearth modes of communication in the brain.

REFERENCES

- [1] Kaggle. (2016) Predict seizures in long-term human intracranial eeg recordings. [Online]. Available: <https://www.kaggle.com/c/melbourne-university-seizure-prediction>
- [2] G. Sugihara, R. May, H. Ye, C.-h. Hsieh, E. Deyle, M. Fogarty, and S. Munch, "Detecting causality in complex ecosystems," *science*, vol. 338, no. 6106, pp. 496–500, 2012.
- [3] T. Hjornevik, T. B. Leergaard, D. Darine, O. Moldestad, A. M. Dale, F. Willoch, and J. G. Bjaalie, "Three-dimensional atlas system for mouse and rat brain imaging data," *Frontiers in neuroinformatics*, vol. 1, p. 4, 2007.
- [4] G. Paxinos, C. Watson, P. Carrié, M. Kirkcaldie, and K. Ashwell, "Chemoarchitectonic atlas of the rat brain," 2009.
- [5] J. Potworowski, W. Jakuczun, S. Łski, and D. Wójcik, "Kernel current source density method," *Neural computation*, vol. 24, no. 2, pp. 541–575, 2012.
- [6] J. Locke, *An essay concerning human understanding*, 1841.
- [7] G. Berkeley, *A treatise concerning the principles of human knowledge*. Philadelphia: JB Lippincott & Company, 1874.
- [8] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, no. 3, pp. 424–438, 1969.
- [9] E. N. Lorenz, "Deterministic nonperiodic flow," *Journal of the atmospheric sciences*, vol. 20, no. 2, pp. 130–141, 1963.
- [10] G. S. Paul A. Dixon, Maria J. Milicich, "Episodic fluctuations in larval supply," *Science*, vol. 283, no. 5407, pp. 1528–1530, 1999.
- [11] E. R. Deyle and G. Sugihara, "Generalized theorems for nonlinear state space reconstruction," *PLoS One*, vol. 6, no. 3, p. e18295, 2011.
- [12] F. Takens, *Detecting strange attractors in turbulence*. Springer, 1981.
- [13] F. Villatoro. (2012) Takens' theorem in action for the lorenz chaotic attractor. [Online]. Available: <https://www.youtube.com/watch?v=6i57udsPKms>

- [14] E. R. Deyle, M. Fogarty, C.-h. Hsieh, L. Kaufman, A. D. MacCall, S. B. Munch, C. T. Perretti, H. Ye, and G. Sugihara, “Predicting climate effects on pacific sardine,” *Proceedings of the National Academy of Sciences*, vol. 110, no. 16, pp. 6430–6435, 2013.
- [15] X. Wang, S. Piao, P. Ciais, P. Friedlingstein, R. B. Myneni, P. Cox, M. Heimann, J. Miller, S. Peng, T. Wang, H. Yang, and A. Chen, “A two-fold increase of carbon cycle sensitivity to tropical temperature variations,” *Nature*, vol. 506, no. 7487, pp. 212–215, 02 2014.
- [16] J. C. McBride, X. Zhao, N. B. Munro, G. A. Jicha, F. A. Schmitt, R. J. Kryscio, C. D. Smith, and Y. Jiang, “Sugihara causality analysis of scalp eeg for detection of early alzheimer’s disease,” *NeuroImage: Clinical*, vol. 7, pp. 258–265, 2015.
- [17] E. H. van Nes, M. Scheffer, V. Brovkin, T. M. Lenton, H. Ye, E. Deyle, and G. Sugihara, “Causal feedbacks in climate change,” *Nature Climate Change*, vol. 5, no. 5, pp. 445–448, 2015.
- [18] A. A. Tsonis, E. R. Deyle, R. M. May, G. Sugihara, K. Swanson, J. D. Verbeten, and G. Wang, “Dynamical evidence for causality between galactic cosmic rays and interannual variation in global temperature,” *Proceedings of the National Academy of Sciences*, vol. 112, no. 11, pp. 3253–3256, 2015.
- [19] J. M. McCracken and R. S. Weigel, “Convergent cross-mapping and pairwise asymmetric inference,” *Physical Review E*, vol. 90, no. 6, p. 062903, 2014.
- [20] A. T. Clark, H. Ye, F. Isbell, E. R. Deyle, J. Cowles, G. D. Tilman, and G. Sugihara, “Spatial convergent cross mapping to detect causal relationships from short time series,” *Ecology*, vol. 96, no. 5, pp. 1174–1181, 2015.
- [21] A. Wismüller, X. Wang, A. M. DSouza, and M. B. Nagarajan, “A framework for exploring non-linear functional connectivity and causality in the human brain: Mutual connectivity analysis (mca) of resting-state functional mri with convergent cross-mapping and non-metric clustering,” *arXiv preprint arXiv:1407.3809*, 2014.