Trinity University

Digital Commons @ Trinity

Mathematics Faculty Research   Mathematics Department

7-2021

Analysis of EEG Data Using Complex Geometric Structurization

Eddy A. Kwessi
Trinity University, ekwessi@trinity.edu

L. J. Edwards

Analysis of EEG Data Using Complex Geometric Structurization

E. A. Kwessi
ekwessi@trinity.edu
Department of Mathematics, Trinity University,
San Antonio, TX 78212, U.S.A.

L. J. Edwards
ljedward@uab.edu
Department of Biostatistics, University of Alabama at Birmingham,
Birmingham, AL 35294, U.S.A.

Electroencephalogram (EEG) is a common tool used to understand brain activities. The data are typically obtained by placing electrodes at the surface of the scalp and recording the oscillations of currents passing through the electrodes. These oscillations can sometimes lead to various interpretations, depending on, for example, the subject's health condition, the experiment carried out, the sensitivity of the tools used, or human manipulations. The data obtained over time can be considered a time series. There is evidence in the literature that epilepsy EEG data may be chaotic. Either way, the Embedding Theory in dynamical systems suggests that time series from a complex system could be used to reconstruct its phase space under proper conditions. In this letter, we propose an analysis of epilepsy EEG time series data based on a novel approach dubbed complex geometric structurization. Complex geometric structurization stems from the construction of strange attractors using Embedding Theory from dynamical systems. The complex geometric structures are themselves obtained using a geometry tool, the α-shapes from shape analysis. Initial analyses show a proof of concept in that these complex structures capture the expected changes brain in lobes under consideration. Further, a deeper analysis suggests that these complex structures can be used as biomarkers for seizure changes.

1 Introduction

A common way to understand brain functions and brain-related diseases is to place electrodes on a subject's head and record the electrical activities that result. These activities, when represented, are often oscillations that change in shape, frequency, and range. They can be analyzed to see if meaningful information can be extracted that gives a clue about the brain

state of the subject at a certain time or in a certain region of the brain. The accuracy of the recording highly depends on the instrument used, the region of interest (ROI), the experimenter's experience, the time of the day, the subject's discipline during the recording and many other factors. This is to say that uncertainty in the accuracy of the information recorded is always present. Even when the information is being recorded on the same subject, on the same day, on the same ROI but at different intervals, changes are bound to occur. Electroencephalograms (EEG), a term coined by Berger (1929), are electrical activities recorded on humans or animals that display prominent oscillatory behavior subject to important changes during various behavioral states. These changes show a high degree of nonlinearity in the signals that may be important. Indeed, in the field of biomedical signal processing (analysis of heart rate variability, electrocardiogram, hand tremor, EEG), the presence or absence of nonlinearity often conveys information about the health condition of a subject. In particular, EEG signals are often examined using nonlinearity analysis techniques or by comparing signals that are recorded during different physiological brain states (e.g., epileptic seizure). The differences observed during these analyses can be due to either genuine differences in dynamical properties of the brain or differences in recording parameters. The EEG is often analyzed as times series, and there are many methods for analysis of times series in the literature. The methods can be grouped into two categories: univariate measures and multivariate measures. (For a good review on the topic, see Carney, Myers, & Geyer, 2011.)

Among the methods that have been touted as more efficient at providing insight into the real dynamics of EEG is the famous Embedding Theorem (Takens, 1981). This theorem has been instrumental in understanding how to reconstruct the true dynamics of systems based on the times series measured on these systems. Essentially, this reconstruction theory, in lay terms, suggests that a times series measured over a sufficiently long period of time contains enough information to reconstruct the phase space in which the associated system normally evolves. This shows that there are other intricate subunits that influence the changes observed in the measured variables that are represented in the time series. This theorem was used, for example, by Grassberger and Procaccia (1983) to propose a measure called correlation dimension, which was in turn instrumental to Lehnertz and Elger (1998) in predicting epilepsy seizures. Despite all the methods proposed predicting, there is no agreement on which method constitutes the best tool at extracting the most meaningful information that could be useful for the physician for the prediction of seizure-like diseases. Moreover, with the increasing use of the concepts of chaos and complexity in health sciences, it is becoming more difficult to distinguish their adequate application. For example, there has been evidence of chaos in EEG data (Destexhe & Babloyantz, 1986; Destexhe, 1992; Destexhe, Sepulchre, & Babloyantz, 1998), and since chaotic

systems are inherently complicated, they may look complex. Complex systems may also look chaotic. Distinguishing these two notions is important in applications, especially in health sciences. Henceforth, we adopt a terminology along the lines of Rickles, Hawe, and Shiell (2007) for understanding complexity and chaos.

In this letter, we propose a new method for analyzing EEG times series data, which we call complex geometric structurization (CGS). The complex nature of the method stems from the fact that we have multiple subunits interacting together, resulting in a rich collective behavior feeding back into the behavior of individual parts. The method is inspired by the Embedding Theorem (Takens, 1981) for the construction of a geometric structure whose volume can be evaluated from shape analysis technique. The volume of this geometric structure behaves as a key statistic akin to a biomarker for the phenomenon or ROI of interest. Using data-driven approaches to study brain pathologies is now an active field of study due to improvement in life expectancy across the world with its cohorts of problems such as brain disorders (Zheng, Fushing, & Ge, 2019; David, Machado, Ińacio, & Valentin, 2020). Moreover, the push to use data- and methodological-driven approaches to brain pathologies is also evidenced by the numerous grants offered by the National Institutes of Health and private foundations such as the Michael J. Fox Foundation for Parkinson Disease and the Bill and Melinda Gates Foundation, just to name some.

The remainder of the letter is organized as follows. In section 2, we briefly review how to use the Embedding Theorem to construct strange attractors; in section 3, we review important notions of statistical morphometry; in section 4, we introduce the complex geometric structurization method; in section 5, we discuss some applications of the CGS method on real data; and in section 6, we discuss the pros and cons of the proposed method in different contexts.

2 Understanding Embedding Theory

As its name suggests, a dynamical system is one whose variables evolve over time. Its phase space is a geometric representation of the trajectory of its variables over time. The values taken by the system's variables at an instant describe the system's states. To understand how to reconstruct the phase space of a dynamical system based on observations (times series) of one of its variables, we need to revisit the Embedding Theory (Takens, 1981), which is essentially a high-dimension transformation of the time series. Consider the n-dimensional space $\mathbb{R}^n$. We recall that a manifold M in the space $\mathbb{R}^n$ is a topological space that locally looks like a Euclidean space near each point. "Topology" here means that bending is ignored. For example, the surface of the globe is a topological manifold in the space $\mathbb{R}^3$. Now consider a dynamical system with a system's state $x(t)$ lying on a manifold
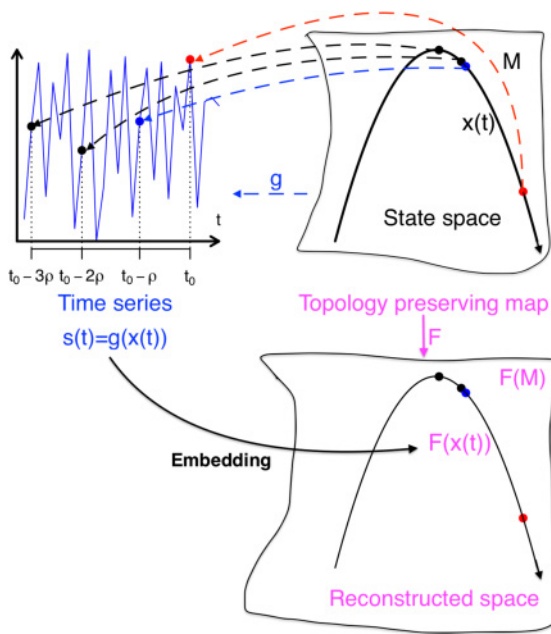
Figure 1: An illustration of the embedding mechanism.

M of $\mathbb{R}^n$. Let ρ be a sampling interval, and let the time series $s(t) = g(x(t))$ be given as a one-dimensional observation of the system dynamics through an observation function g. The Embedding Theory states that for almost every smooth function g, the delay coordinate map defined as $F : \mathbb{R}^n \to \mathbb{R}^m$ with $F(x(t)) = [s(t), s(t - \rho), \ldots, s(t - (m - 1)\rho)]^T$ is an embedding, that is, it is a one-to-one immersion of the state-space attractor with dimension d when $m > 2d$. In other words, the result states that $F(x(t))$ is a representative of $x(t)$, even if the true state-space M has not been observed (see Figure 1). The quantity m is referred to in the literature as the embedding dimension and ρ as the time delay (or lag). The Embedding Theory is predicated on the observation that a time series observed over a long period of time may show an internal structure. In fact, considering a time series $s(t) = g(x(t))$, we observe only an incomplete picture of $x(t)$ since $s(t)$ is a scalar. However, if we observe it for a long time, a more precise description will emerge, which will help in understanding its dynamics. In practice, most of the focus is given into how to find appropriate values for the time delay ρ and the embedding dimension m (see section A.2 in the appendix). We observe that this type of reconstruction technique has been applied successfully in the case of epilepsy EEG data (Destexhe & Babloyantz, 1986; Destexhe, 1992; Destexhe et al., 1998).
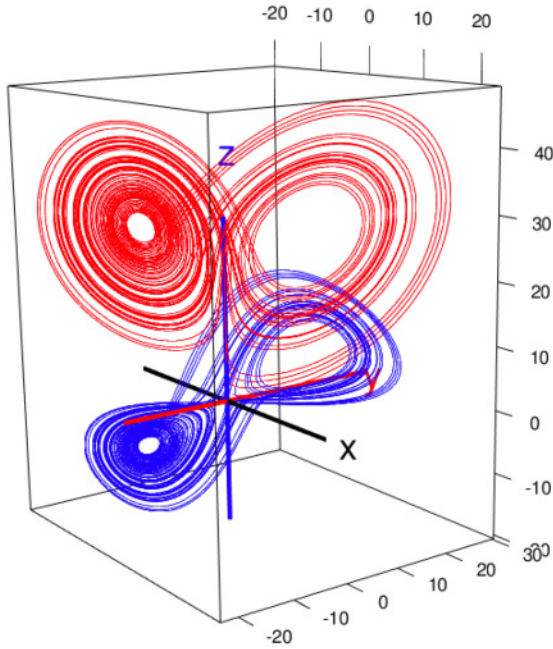
Figure 2: Lorenz attractor (red) with its reconstructed counterparts (blue) and plotted in the same system of coordinates. We can observe the topological equivalence between the original phase space and its reconstructed counterpart.

2.1 Examples.

2.1.1 Dynamical Systems. The Embedding Theory can be used to reconstruct the Lorenz, a famous attractor often mentioned in dynamical systems. Lorenz's (1963) system of differential equations is given as

$$
\begin{cases}
\dot{x} = s(y - x) \\
\dot{y} = rx - y - xz \\
\dot{z} = xy - bz
\end{cases}.
$$

It is known, for example, that the Lorenz system is chaotic for $s = 10$, $r = 28$, and $b = 8/3$. Figure 2 is a depiction of this attractor for these parameter values plotted in the space $x = x(t)$, $y = x(t - \rho)$, and $z = x(t - 2\rho)$. The time step used is $\Delta t = 0.005$ for an interval time of $[0, 75]$ for the Lorenz system. We also note that the time lag or delay is $\rho = 31\Delta t$ and was found using either the autocorrelation (ACF) or average mutual information (AMI; see section A.2).

2.1.2 Real Data. There have been many applications of Takens's Embedding Theory with real data. For example, in Fisher, Talathi, Cadotte, and Carney (2009) and Carney et al. (2011), an attractor is constructed from a publicly available (http://www.meb.unibonn.de/epileptologie/science/physik/eegdata.html) epilepsy data set, which we call EDATA for simplicity. The data consist of five sets A, B, C, D, and E. Each contains 100 single-channel EEG segments of 23.6 seconds, each selected after visual inspection for artifacts (such as acoustic and electrical shielding, separation of earth ground for laboratory, interconnectivity of devices on the same phase and ground centrally and locally) and has passed a weak stationarity criterion. Sets A and B were obtained from surface EEG recordings of five healthy subjects with eyes open and closed, respectively. Data were obtained in seizure-free intervals from five patients in the epileptogenic zone for set D and from the hippocampal formation of the opposite hemisphere of the brain for set C. Set E contains seizure activity, selected from all recording sites exhibiting ictal activity. Sets A and B have been recorded extracranially, whereas sets C, D, and E have been recorded intracranially.

All EEG signals were recorded with the same 128-channel amplifier system, using an average common reference (omitting electrodes containing pathological activity (C,D, and E) or strong eye movement artifacts (A and B)). After 12-bit analog-to-digital conversion, the data were written continuously onto the disk of a data acquisition computer system at a sampling rate of 173.61 Hz. Bandpass filter settings were 0.53 Hz to 40 Hz (12 dB/oct.; see Andrzejak et al., 2001. Figure 3 is an illustration of the data set EDATA. Each row represents one time series from sets A, B, C, D, and E, respectively. Clearly, the time series in the seizure set E has a pronounced amplitude synonymous with more brain activities. In Figure 4, we represent the reconstructed phase spaces based on the time-selected times series from sets A to E. We note that the axes are $x = x(t)$, $y = x(t - \rho)$, $z = x(t - 2\rho)$ where $\rho = 1\Delta t$, with $\Delta t = \frac{1}{fs} = 5.76$ ms.

3 Statistical Morphometry

Given a set of points in a two- or three-dimensional space, statistical morphology (or morphometrics) amounts to finding an appropriate geometric characterization of the variability of the shape and size of the set of points. This characterization includes volumes and surface area, surface roughness, deviation from convexity, porosity, and permeability. We note that in general, the set of points may not be convex in the strictest sense, so there is a need for a method that relaxes the convexity restriction. The notions of α-convexity and α-shape represent alternatives that relax the strict convexity assumption. These new α-shaped objects are even more flexible than α-convex objects in that α now controls the spatial scale of the estimator.
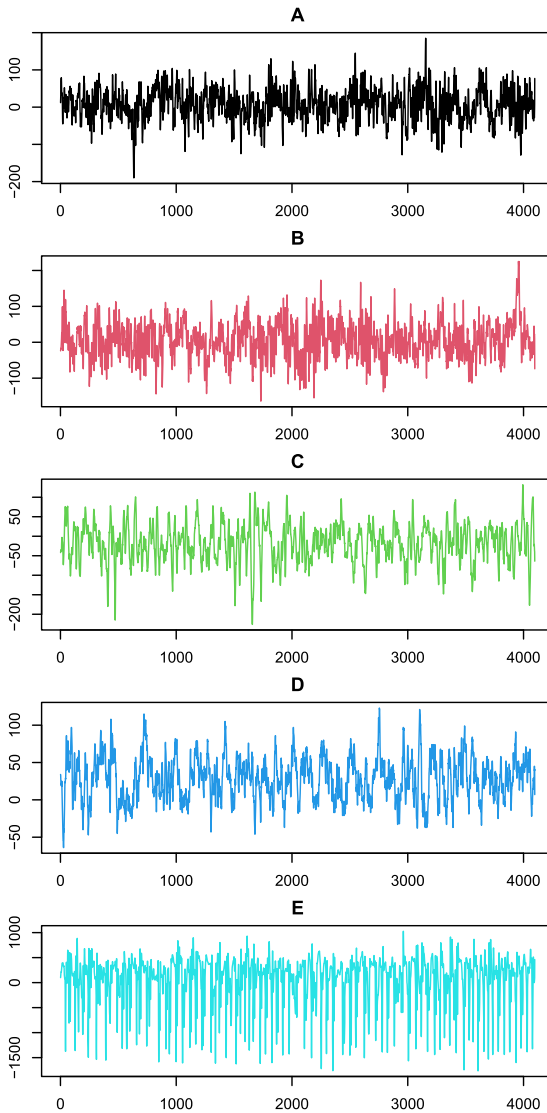
Figure 3: This figure represents one time series selected at random from each set A to E. The amplitude is much more pronounced in set E, which represents seizure-prone patients.

Note that α is a unitless quantity. In fact, as α decreases, the α-shape shrinks and more space appears among the sample points, whereas as α increases, the α-shape object converges to the convex hull of the sample. In other
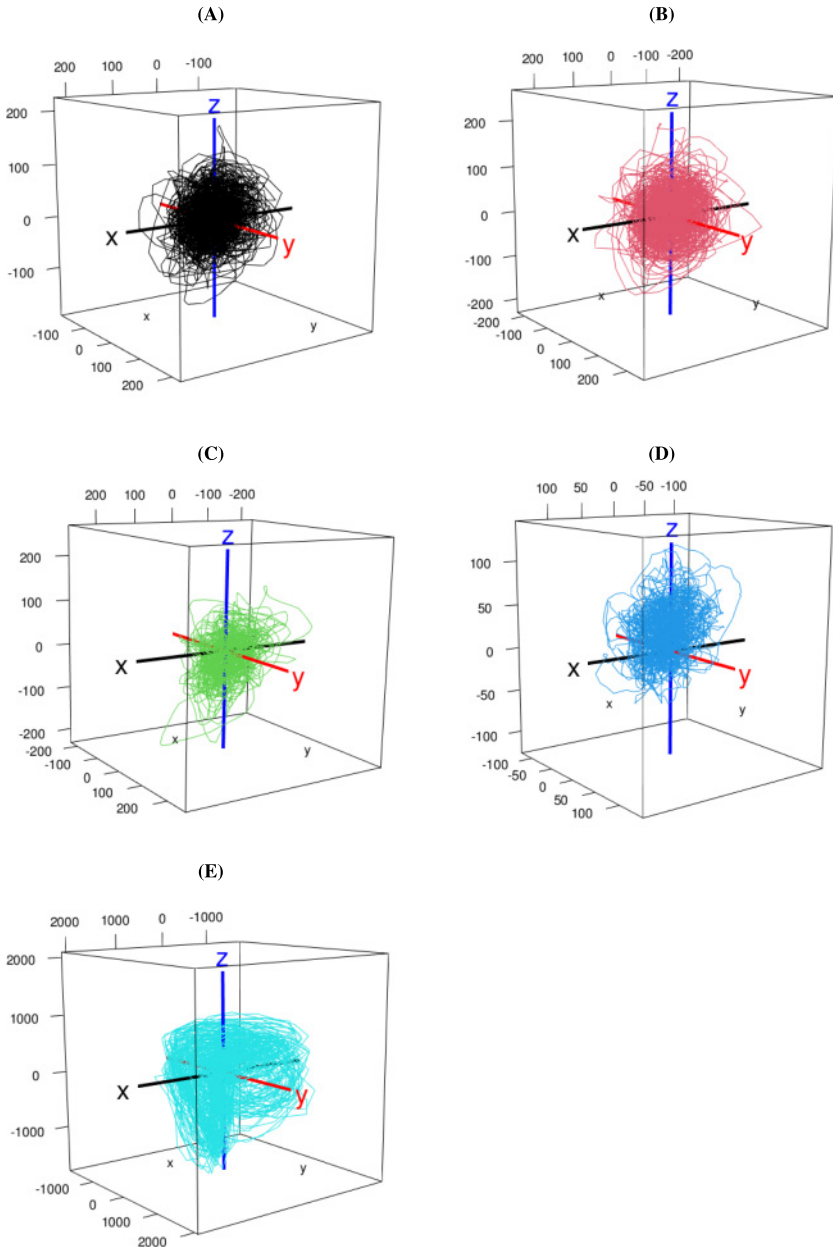
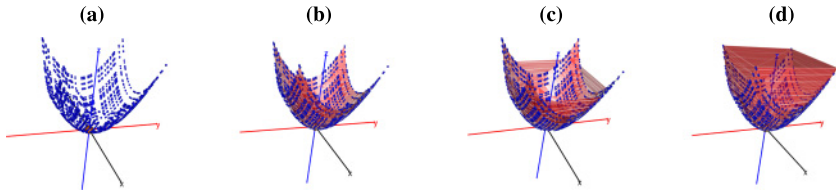Figure 4: Reconstructed phase-space diagram, restricted to the space *xyz*.

Figure 5: α-shape object (red) of S $(z = x^2 + y^2)$ for $\alpha = 0$ (a), $\alpha = 0.5$ (b), $\alpha = 0.8$ (c), and $\alpha = 2$ (d).

words, in α-shaped objects, α controls the amount of porosity between the sample points. These objects have been used in various fields for the characterization of biological systems (Lafarge, Pateiro-Lopez, Possolo, & Dunkers, 2014; Gardiner, Behnsen, & Brassey, 2018). We mention the pivotal work of Edelsbrunner and Mücke (1994) where the main algorithm for the construction of α-shaped objects can be found.

3.1 Example. In this example, we illustrate the α-shape construction in two and three dimensions, based on a random sample of data taken from the original object S.

We consider 2500 points in 3D, obtained from object S, which is the object delimited by the curve, and $z = x^2 + y^2$, where x and y are random points selected in the interval $[-1, 1]$. In Figure 5, we construct the α-shape object (red) for $\alpha = 0$ (a), $\alpha = 0.5$ (b), $\alpha = 0.8$ (c), and $\alpha = 2$ (d).

4 Complex Geometric Structurization

In view of the apparent shape that can be observed from the reconstructed phase space, the question of how to compare these complex structures arises. In other words, this question is related to the question of how to compare strange attractors. Among the methods proposed, we can mention the work of Grassberger and Procaccia (1983) and its many variants. The method we propose stems from the observation that there seems to be a complex geometric structure whose shape changes from healthy patients to seizure-prone ones, so we will rely on the notion of α-shape to construct the complex structure related to each situation. We use an algorithm to construct our complex geometric structure (CGS). The motivation comes from the fact that given a time series, if it has been observed long enough, it carries the signature of the original phase-space diagram in which the true system it originated from evolves. If we have many such time series from the same system, we should be able to capture enough information about the true phase-space diagram. The large number of these time series should be enough to eventually eliminate noise or undesired artifacts from the reconstruction. We then expect each reconstructed phase diagram to be

topologically equivalent to any other, therefore forming a structure compact enough to be the revolving center of all other reconstructed phase-space diagrams. In general, the dimension of the space in which the true system evolves is greater than three, making it impossible to visualize with the naked eye. Our empirical approach is to consider only a cross-section of the reconstructed phase space in dimension 3 in this case. The reason for this choice is two-fold: we can actually visualize a part of the true phase-space diagram, and we can use the existing method to evaluate the volume of the structure in lower dimensions. This is to say that the volume is preserved in reconstruction. In doing this, we are trying to find a measurable identifier or markup for this group of time series that will vary from group to group and from individual to individual.

4.1 Explanation of the Method. Given N time series, we use the Takens reconstruction technique to obtain the embedding dimension m_n (using ACF) and time delay (or sampling interval) ρ_n for $n = 1, 2 \ldots, N$ (using the method of false nearest neighbors). Let $m = \min\{m_1, \ldots, m_N\}$ and $\rho = \min \{\rho_n, n = 1, 2, \ldots, N\}$. If $m \geq 3$, then for each time series, we obtain the complex structure $\text{CGS}_{\alpha(n)}$. Let $\alpha = \min\{\alpha(n) : \text{Vol}_{3D}(\text{CGS}_{\alpha(n)})$ is maximized\}. We use the α-shape technique to obtain the volume of the CGS_α in 3D. This step is crucial since we choose to represent the complex structure using only the first three delay coordinates $x_1 = x(t)$, $x_2 = x(t - \rho)$, and $x_3 = x(t - 2\rho)$ even if the actual space has dimension $m > 3$. The motivation for this selection is that the volume of the CGS based on any combination of three coordinates would not be significantly different from any other volume obtained from any other combination of three coordinates. The reasoning behind the choice of α is that the volume of the complex structure is bounded by the volume of its convex hull as α increases, so we select the smallest alpha that maximizes the volume (see section 4.3.3 for an illustration). This leads to the following algorithm:

1. Use all the times series collected on patients in different groups.
2. For each time series, reconstruct a 3D "strange attractor" from the first three delay coordinates.
3. Use the α-shape technique to construct a CGS related to all strange attractors.
4. Define a measure related to each CGS that that can be statistically analyzed. This could be the volume, the surface area, or the center, for example.
5. Repeat this procedure for all replicates (if there are any) and thus obtain new data.
6. (a) For comparison: Use a statistical test to see if there is a significant difference between measures of different groups. This could be a parametric or a nonparametric test.

(b) For Prediction: Use the data obtained for training in machine learning (see Kwessi & Edwards, 2020, preferred to standard generalized linear model models because of the absence of a specific model) and test them on potential new data.

4.2 Comments.

1. The CGS terminology stems from the fact the structures obtained after reconstruction do not have classical geometric shapes. Rather, their shapes are complex in nature.
2. We propose to use all time series available for a particular region of the brain, at a specific instant—for example, before seizure, during seizure, or after seizure. It is common to use one time series (see Fisher et al., 2009). We note that repeated measures on the same subject do not yield the same values, and thus in the reconstruction, there could be individual times series whose excursions in the phase space are wider than others, yielding a bigger geometric structure in size and volume. In this case, individual times series could be used, and the subsequent large volumes could be trimmed if necessary. This is particularly important if one is interested, for example, in obtaining a richer data set to analyze.
3. The embedding dimension m often obtained is greater than 3, so in the absence of a visualization mechanism for data in spaces of dimensions higher than 3, we propose to focus on only the first three delay coordinates.
4. Using the R-package Alphashape3d, measures such as volumes or surface areas can be obtained for the α-shape object under consideration.
5. As we saw in the example above, the choice of α is critical in the α-shape construction. We propose to select the minimum value of α for which the volume is maximized.
6. Finally, we note that in practice, the experimenter can used data from chaotic or complex systems. One can check for chaos in data by calculating Lyapunov exponents. If the data are sensitive to initial conditions and chaotic, the attractor would be referred to as a "strange attractor" (see Celso, Ott, & Yorke, 1987). Even if the data are sensitive to initial conditions but nonchaotic, we would still keep the same "strange attractor" denomination because in this case (the Lyapunov exponents are nonpositive), the attractor obtained is still strange (see Celso, Ott, Pelikan, & Yorke, 1984; Paladin & Vulpiani, 1987). It also noteworthy to observe that nonchaotic systems that are sensitive to initial conditions are sometimes referred to in the literature as complex systems (see the comparative review between chaotic and complex systems by Rickles et al., 2007).
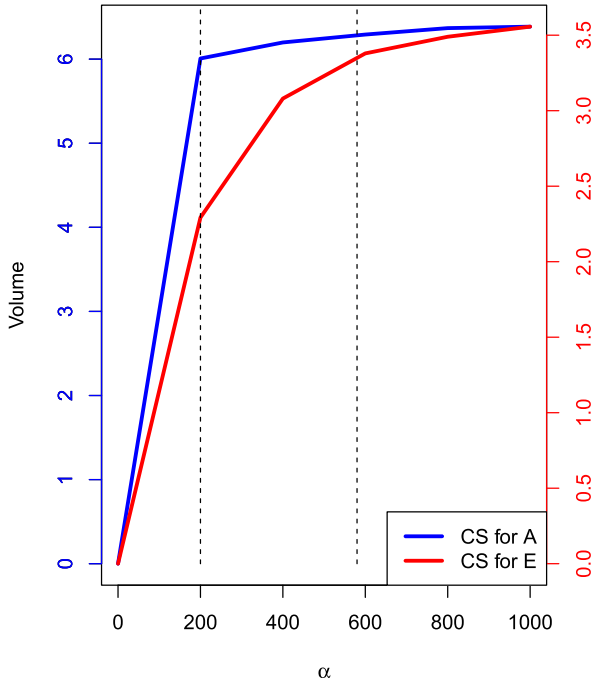
Figure 6: Evolution of the volume CGS as function of α.

4.3 Technical Considerations. In this section, we examine some technical considerations to keep in mind when implementing this method.

4.3.1 Choice of α. The choice of α is critical in this process. In what follows, we show that the optimal value of α is smallest that maximizes the volume of the CGS. Moreover, if one is interested in comparing CGS among groups, it would be adequate to select a common value of α, for example, as the largest α value, among the values that maximize the CGS for each group. For instance, using the EDATA set, we obtained $m = 10$ and $\rho = 1\Delta t$, with $\Delta t = \frac{1}{f_s} = 5.76$ ms. In Figure 6, we observe that for a value of $\alpha \approx 200$, the volume is maximized for subset A, whereas the maximum is reached for subset E at $\alpha \approx 580$, so we choose $\alpha_{\text{optim}} = 580$ as the optimal value of α if we want to compare the two CGSs.

4.3.2 Choice of the Embedding Dimension. We note that the embedding dimension is closely related to the time lag. We are not going to discuss which should be estimated first. However, we note that there are two methods for estimating the time lag and they do not always yield the same embedding
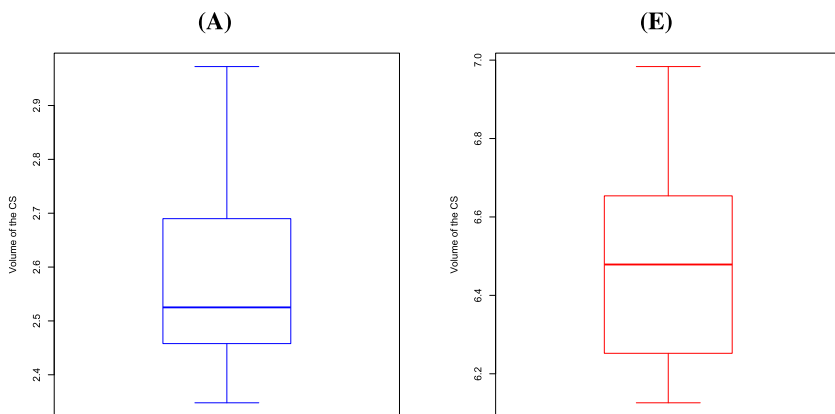
Figure 7: Evolution of the volume the CGS for the 120 different combinations of three delay-coordinates for subsets A and E.

dimension. We propose to select the embedding dimension as the smallest value of the two embedding dimensions when they are different.

4.3.3 Choice of the Delay-Coordinates. In section 4.3.2, we mentioned that we will select the first three delay coordinates to construct the complex structure. However, it is worthwhile to consider the question of whether the volume would change if a different combination of delay coordinates is used. Our choice is based on the conjecture that it does not matter which combination of delay coordinates is selected. To emphasize that point, we discuss the case of subsets A and E. In these subsets, the embedding dimension found is $m = 10$, so there are $\binom{10}{3} = 120$ possible different combinations of three delay coordinates, selected among 10. For each combination, we construct the corresponding CGS and assess whether the volume changes significantly across all the different CGS's. Figure 7 shows the volume of the CGS by combination of three delay coordinates for subsets A and E. The box plots suggest that the distribution of volumes for each set is reasonably concentrated around its median (thick red and blue lines) with a small range and no outliers. This is to say that taking the first three delay coordinates seems reasonable despite minor variations otherwise.

The variation of the volume as a function of the combination of delay coordinates appears not to be significant. Even for seizure data like subset E, the variation in volume appears to be mild, which is our impetus for conjecturing that a similar observation could be made for other data sets. Obviously we cannot predict what will happen for all data sets; ultimately that endeavor would require a mathematical proof that may go beyond the scope of this letter.
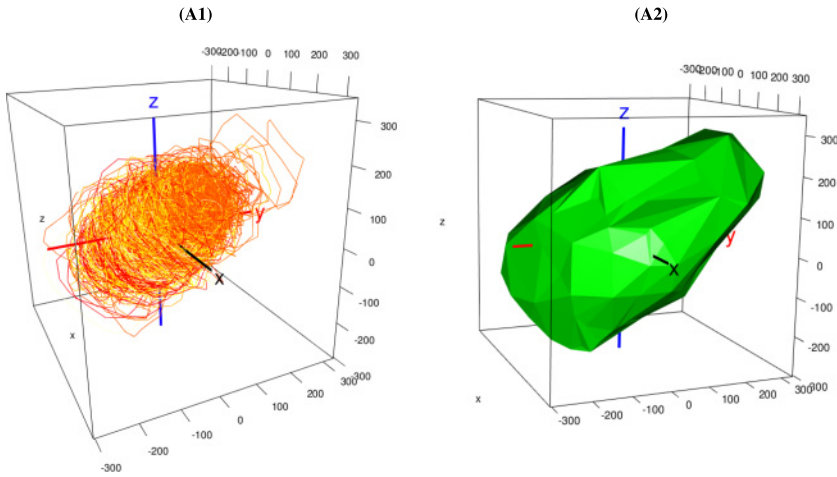
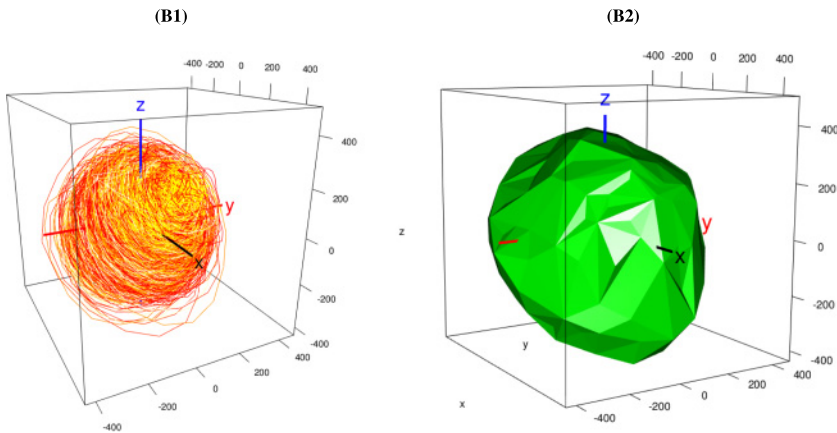Figure 8: Complex geometric structure for set A in EDATA.

Figure 9: Complex geometric structure for set B in EDATA.

5 Applications

5.1 Analysis of EDATA. In Figure 4, we showed a representation of the reconstructed phase-space diagram for one time series in the space $x = x(t), y = x(t - \rho), z = x(t - 2\rho)$ and for each set A to E. In Figures 8 to 12, we construct the phase-space diagram and the complex structures for the EDATA set for all 100 time series in each subset. We observe that for each set, we obtain a compact structure whose volume we can now evaluate (see panels A1 to E1). We selected $\alpha = 580$ for each case, because of all sets A
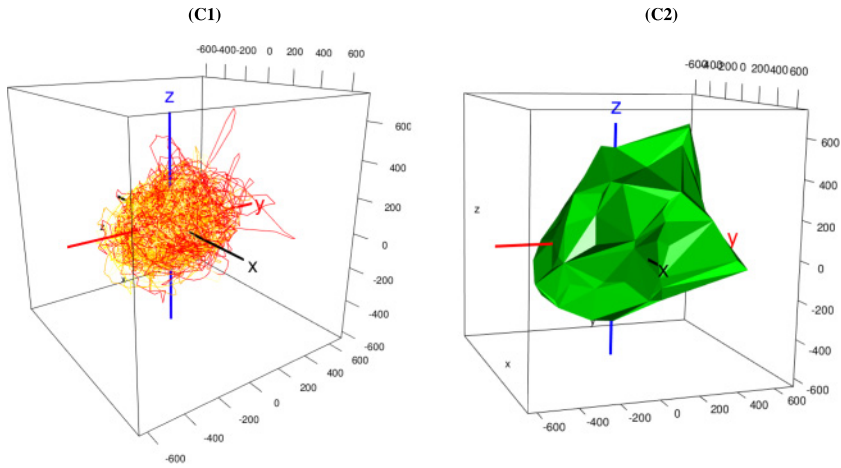
(C1)

(C2)

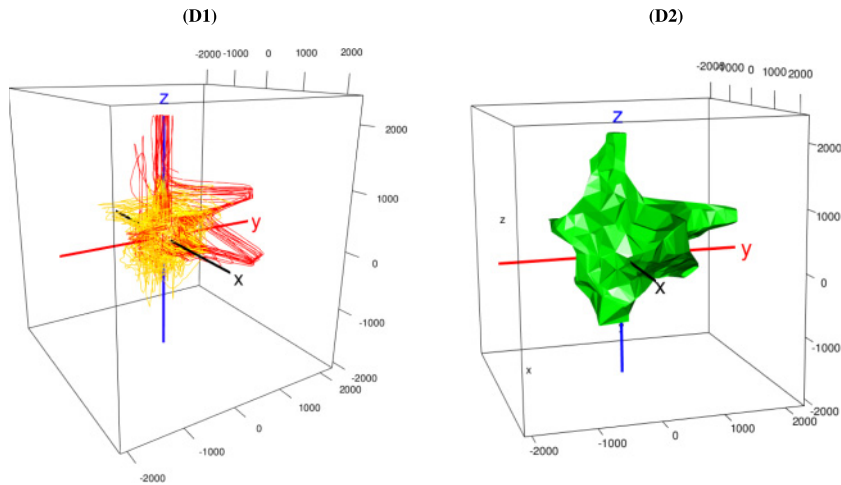Figure 10: Complex geometric structure for sets C in EDATA.

(D1)

(D2)

Figure 11: Complex geometric structure for set D in EDATA.

to E, 580 is the value of alpha that maximizes the volume of the complex structure of E, the largest of all of them (see the selection criterion given above; see Figure 6). The CGS structure for each set is then obtained (see panels A2 to E2). We observe that the CGS for set E appears to be bigger than all other CGSs, a sign of larger excursions in the phase space and therefore synonymous with intense brain activity during seizure.

Pooling the data may seem to inflate the volume of the CGS. Instead we could construct the complex structure related to individual channel and

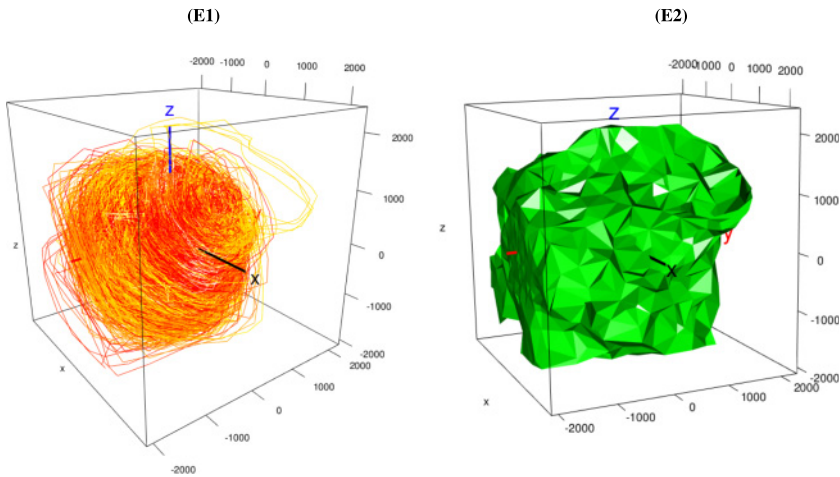**(E1)**   **(E2)**

Figure 12: Complex geometric structure for set E in EDATA.

evaluate their volumes. This will give richer data to analyze, where outliers can be removed. Figure 13 shows the box plot of the volume of the complex structure for each channel. The volume for subset E is very large compared to A to D and obscures the distribution for subsets A to D. As a result, in the right panel in Figure 13, we remove subset E so that the distributions of subsets A to D can be more clearly seen.

Figure 13 is further evidence of what already observable in the raw data in Figures 3 and 4. Sets A to D have comparable amplitude at the time-series level confirmed by the fact that the volumes of CGS are in similar range. The amplitude at the time series level of set E is much higher than that of sets A to D, which is also confirmed by a larger volume at CGS level. This is further evidence that during seizure, these volumes increase substantially when compared to seizure-free intervals. More important, we can extract meaningful statistics from these data as suggested in Table 1.

5.2 Analysis of Auditory and Visual Cortex of the Brain under Auditory and Visual Tasks, and Rest. In this example, we discuss data collected for the Brain Core Initiative at the University of Alabama-Birmingham. These data were collected on 20 individuals in two regions of their brain, the auditory and visual cortex (see Figure 14). The patients were subject to three "tasks": auditory, visual, and rest. Measurements of brain activity was obtained as EEG times series after removing unnecessary artifacts. For this data set, the optimum value of α is $\alpha_{optim} = 300$. We consider the following subsets: auditory cortex-auditory task, auditory cortex-visual task, auditory cortex-rest, visual cortex-visual task, visual cortex-visual task, visual cortex-rest.
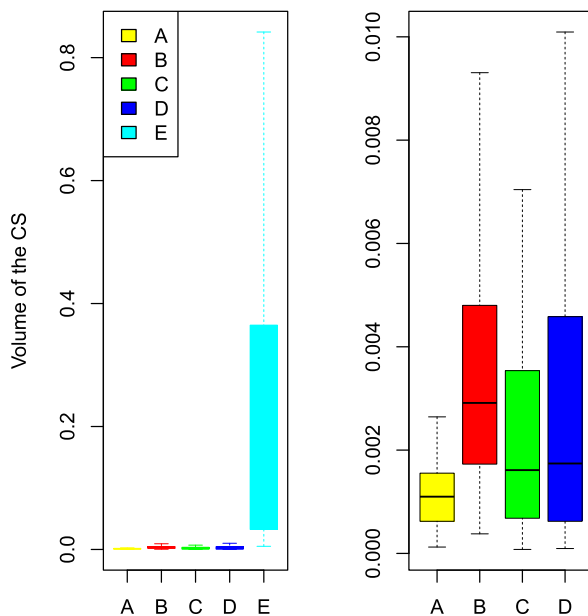
Figure 13: Box plots for the CGS for subsets A to E in EDATA, obtained from individual times series.

Table 1: Table of Volumes of CGS for Each Set A–E.

	Min	Q_1	Median	Q_3	Max	Mean	SD
A	12.30	63.63	115.87	154.17	317.88	109.84	66.08
B	37.89	63.63	173.50	291.31	479.06	364.25	266.63
C	7.73	68.38	161.30	262.70	351.60	262.10	333.70
D	9.45	62.78	174.10	4569.00	17,800.00	692.30	206.54
E	516.3	3,277.30	9,279.40	36,270.50	1,34,300.00	26,422.50	33,682.48

Note: The values are of order 10^{-5}.

5.2.1 Macrolevel Analysis. Here, all the time series are used. Figure 15 represents the CGS for each of the subsets above.

In a comparison of the first and second rows, the volumes seem to differ by cortex. Comparing the first, second, and third columns, the differences in volumes of tasks are, respectively, 0.449, 0.5538, and 3.296. These numbers show similarity between the auditory and visual tasks, but they both differ from rest. The increase of volume during rest may not be counterintuitive. In fact, a vast literature links resting brain activities to underlying high
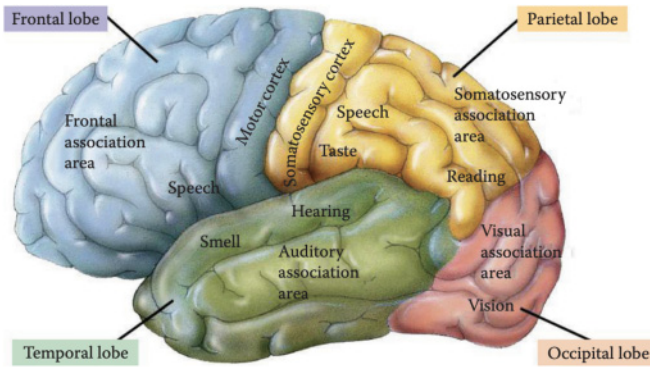
Figure 14: Functional diagram of the brain lobes. Image credit (Chen, 2011).

Auditory–Auditory Task

Vol=7.8181

Auditory–Visual task

Vol=7.8911

Auditory–Rest

Vol=11.449

Visual-Auditory Task

Vol=8.2875

Visual–Visual task

Vol=7.3373

Visual–Rest

Vol=14.745

Figure 15: Complex geometric structures for the above subsets.

cognitive processes such as moral reasoning, self-consciousness, remembering past experiences, or planning for the future (Bruckner, Andrews-Hanna, & Schacter, 2008).

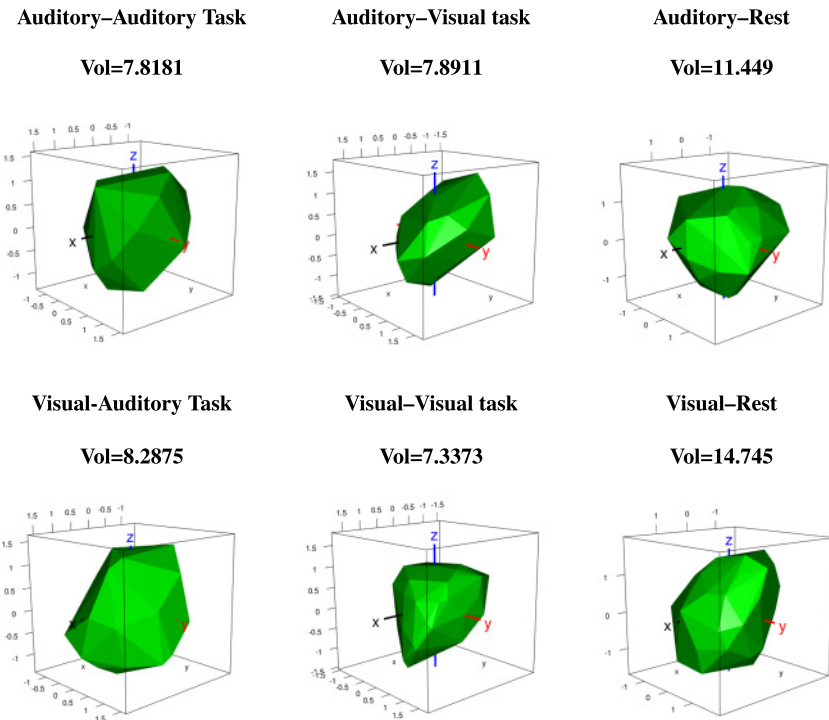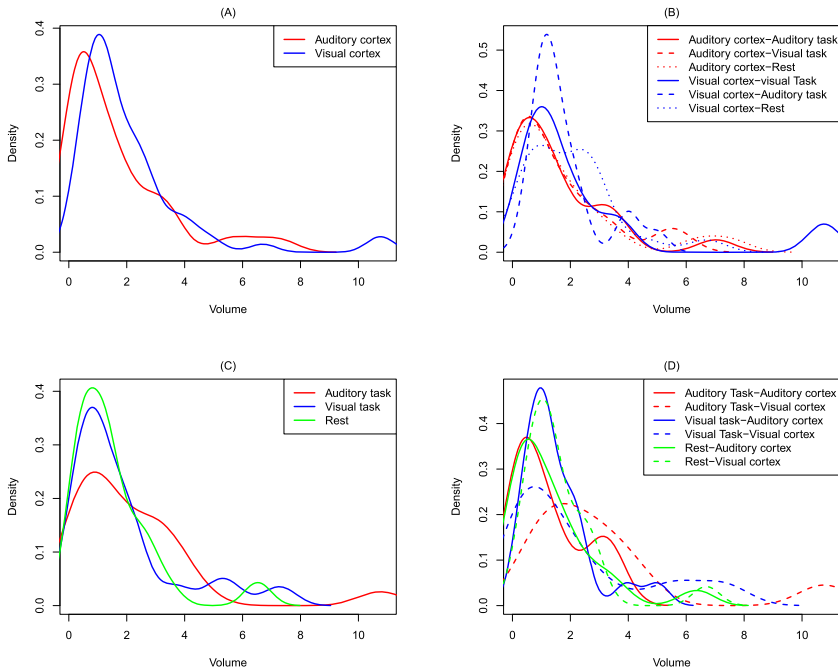Figure 16: Density plots for the volume of CGS obtained from individual times series.

5.2.2 *Microlevel Analysis.* If we use individual time series, since they represent individual patients, we can obtain their CGS by cortex and by task. This would enable having a richer data set and a more in-depth comparison. In Figures 16 and 17, we plot the density and box plot of the volumes of the CGS by cortex and/or by task. The plots in panel A represent the plots of CGS by cortex, those in panel B the plots of cortexes by task, those in panel C the plots by task, and those in panel D the plots tasks by cortex. At first glance, it seems as if a beta or log-normal distribution would provide a good fit to these densities.

We note there are 20 patients and three tasks per cortex for a total of 120 observations. When grouping by cortex, we labeled the observations 1 through 60 per cortex. When grouping by task, we labeled the observations 1 through 40, and 1 through 20 when grouping by task within cortex. The numbers shown in the box plots in Figure 17 represent the labels for the corresponding outlying observations. Plots A not suggest a difference between the auditory and visual cortexes, since the notches overlap for the most part. It shows that there are two common outliers (37 and 57) in both cortexes. This means that there are two patients whose brain activities are more pronounced than others in both cortexes, causing wide excursions in
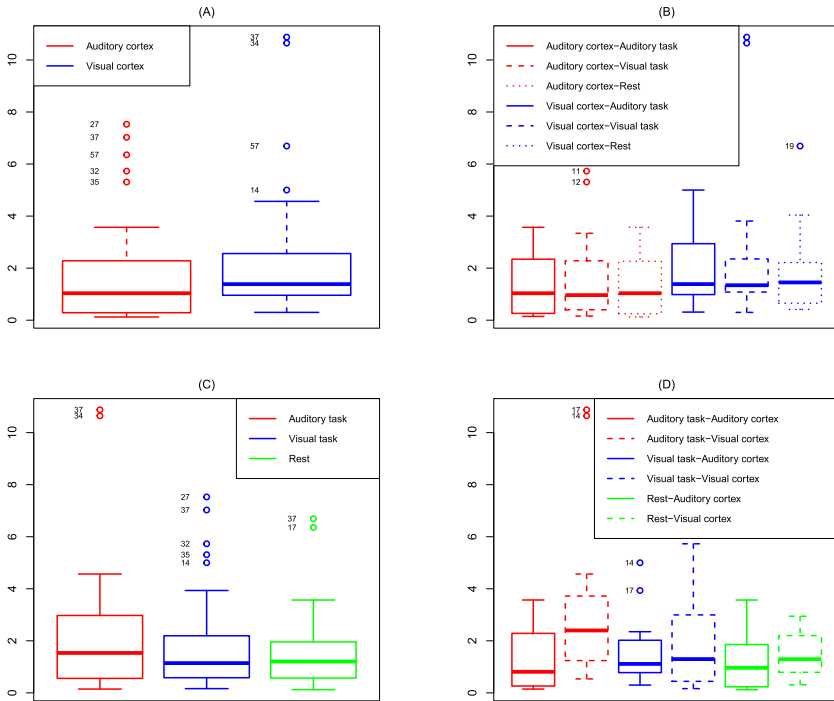
Figure 17: Box plots for the volume of CGS obtained from individual times series.

the phase space and therefore large volumes of their CGS. Plots B and C suggest a similarity between the tasks since all the notches do overlap. Plots D suggest that the difference observed in box plot A is due to the difference between the auditory tasks in both cortexes. Plots B basically confirm the observations in plots A that all the tasks are similar. A more in-depth analysis is needed to make more meaningful conclusions.

In Tables 2 and 3, we report the intrinsic discrepancy, that is, the minimum between $d_{KL}(F, G)$ and $d_{KL}(G, F)$, where $d_{KL}(F, G)$ is defined as the Kullback-Leibler (KL) directed divergences between the probability distributions F and G (see section 8.4). So the smaller the intrinsic discrepancy, the bigger the difference between F and G.

To make a meaningful use of Tables 2 and 3, a reference point is needed. However, the densities above allow us to hypothesize that the volumes are not normally distributed. A nonparametric test such as a Wilcoxon-Mann-Whitney would be necessary to compare, for example, the volumes of the two cortexes, controlling for the tasks performed. Such a test yields a *p*-value of 0.021 small enough to suggest a statistically significant difference

Table 2: Statistical Analysis of the Intrinsic Discrepancy between the Cortexes.

	Auditory Cortex	Visual Cortex	p-Value
Auditory cortex	0	0.160	0.021

Table 3: Statistical Analysis of the Intrinsic Discrepancy between the Tasks.

	Auditory Task	Visual Task	Rest	p-Value
Auditory task	0.000	0.104	0.029	0.451
	—	(0.98)	(0.98)	
Visual task	0.104	0.000	0.080	
	(0.98)	—	(0.98)	

between the CGS volumes for auditory and visual cortexes. It also suggests that the intrinsic discrepancy 0.160 obtained above is a sign of a statistical significant difference between the two cortexes. Likewise, the nonparametric Kruskall-Wallis test is used to compare the tasks, and it yields a p-value of 0.451, which is large, suggesting that the volumes of the CGS are not significantly different. The p-values for the pairwise Wilcoxon rank sum test (in parentheses) with Bonferroni correction are also large. This could be further interpreted as the brain activities of individuals during the auditory task are not significantly different from their brain activities during the visual task. Now, whether these differences or lack thereof are corroborated at the biological level remains to be proved.

5.2.3 Comparison with a Cross-Correlation Function. In this section, we compare our method with the cross-correlation function (CCF); see section A.5. In the heat maps below, represented are distances $\rho(X, Y)$, where X and Y are the EEG data (times series) collected on the 20 patients in different regions of the brain when subjected to the three tasks. The heat gradient represents the values of $\rho(X, Y)$. Figure 18A represents the comparison between auditory and visual cortexes. Its symmetric nature is indicative of the similarity between the two cortexes. This is in agreement with Figure 17A. From Figure 18B, we observe that the heat maps are very similar. In particular, there are four areas that are redder than the majority and indicative of the strongest correlation. A classification by a k-means algorithm confirms the existence of these two clusters (see Figure 19B). This is similar to saying that there is no significant difference between the tasks. The same conclusion can be made with box plots in Figure 17C, where the outliers in the volume of CGS match precisely the regions with high correlation. There is not a huge difference between the first three heat maps in Figures 18C and 18D, confirming that there is no significant difference
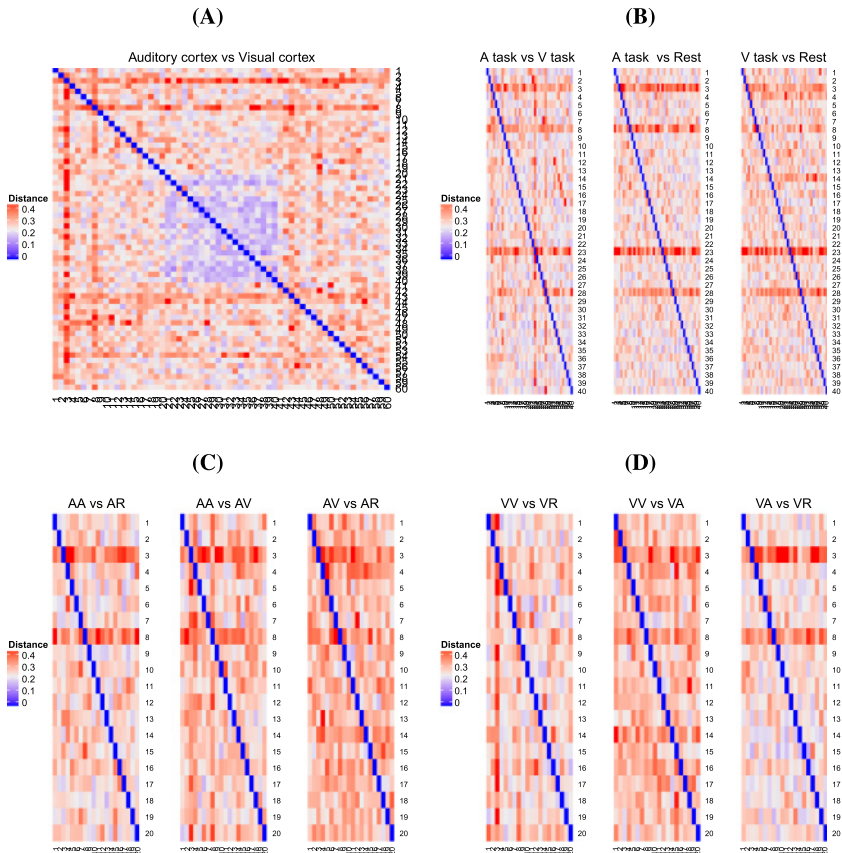
Figure 18: Plot of $\rho(X, Y)$ between the auditory and visual cortexes.

between the tasks within the auditory cortex, an observation also made from Figure 17B. The same observation can be made about the last three maps in the visual cortex.

6 Discussion

We offer some items for discussion as a result of our findings:

1. The method we are proposing has the potential to discriminate brain activities in different parts of the brain. Further, changes in volume could be an indication of changes of activity level in the part of the brain under study. This predictive potential could be used to predict or monitor potential brain disorder.
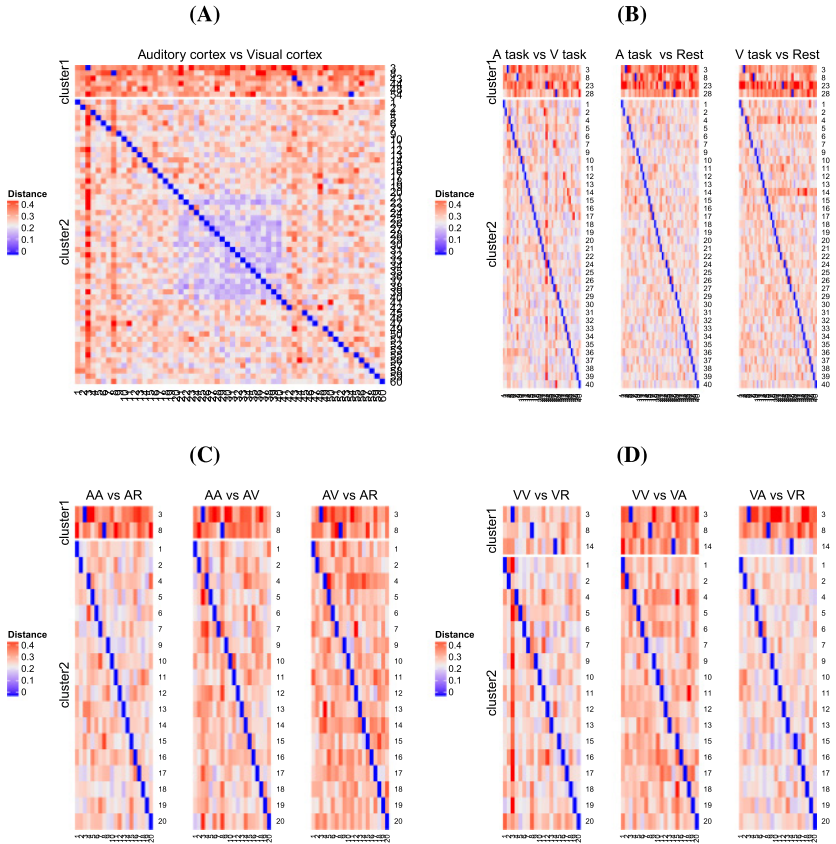
Figure 19: Plot of $\rho(X, Y)$ between the auditory and visual cortexes by cluster. The heat gradient represents the values of $\rho(X, Y)$.

2. This method has the potential to be extended to experiments in which EEG data are suitable, such as epilepsy, Alzheimer's, attention deficit disorder, learning disabilities, anxiety disorders, fetal alcohol syndrome, autism, chronic pain, insomnia, and dyslexia, for example (Kumar & Bhuvaneswari, 2012).

3. Preliminary analysis of local field potential (LFP) data suggests that this method can be extended beyond EEG. In fact LFP differs from EEG in that the electrodes are inserted in the brain tissue rather than at the surface of the scalp. However, we did not obtain the authorization to release the results on the LFP data at this time.

4. There have been studies suggesting that group averaging of neuroimaging data is not clinically relevant. One possible explanation for

this issue is the lack of focus on individuals. The method we propose can remedy that by providing a measure that is individual focused.

5. We have not addressed the weak stationarity of the EEG data used here and how much stationarity, or lack thereof, plays a role in the method we are proposing. The data set EDATA was checked for weak stationarity, but the data collected for the Brain Core Initiative were too few for stationarity. This overall is an issue in multiple studies where samples tend to be small.

6. One of the drawbacks of the "distance" above is that its definition may change, and therefore this could affect the interpretation of results. In fact, given two times series X and Y, the distance could also be defined as $\rho(X, Y) = 1 - \max_{1 \leq \tau \leq m_l} \{|CCF(X, Y, \tau)|\}$ or by $\rho(X, Y) = \text{mean}\{CCF(X, Y, \tau)\}$ with $CCF(X, Y, \tau) = \frac{E[(X_{t+\tau} - \overline{X}) \cdot (Y_t - \overline{Y})]}{\sqrt{E[(X_t - \overline{X})^2] E[(Y_t - \overline{Y})^2]}}$, where $X_{t+\tau}$ is the time-shifted version of X_t, τ is the time lag separating the two times series X and Y, and $\overline{X}, \overline{Y}$ are the respective means of the times series X and Y. (See Arbabshirani, Havlicek, Kiehl, Pearlson, & Calhoun, 2012, for more information.)

7 Conclusion

In this letter, we have proposed a method called complex geometric structurization to help analyze EEG signals. The method is based on the Embedding Theory of dynamical systems and shape analysis. The method works well on epilepsy data. The method has the ability to discriminate individual EEG and also discriminates groups of EEG signals. More important, the method performs better when compared to cross-correlation function. It is important to note that the results and the method, though empirical, offer an important proof of concept relating time series and the volume of their reconstructed complex in three dimensions that need to be explored further by validating them with more solid mathematical concepts. If successful, this method could be an important addition to the literature of EEG signal analysis and can be used to explore other brain pathologies such as sleep disorder, schizophrenia, or Alzheimer's. Its discrimination potential also can be used to improve our understanding of functional brain connectivity in general.

Appendix

A.1 α-convex and α-shape.

Definition 1. *A set $A \subseteq \mathbb{R}^m$ is said to be α-convex, for $\alpha > 0$ if*

$$A = C_\alpha(A) = \underset{B}{\cup} B_\alpha(x), \tag{A.1}$$

where $B = \{B_\alpha(x) : A \cap B_\alpha(x) \neq 0\}$, $B_\alpha(x) = \{y : \|y - x\| \geq \alpha\}$, *and* $\| \cdot \|$ *is a norm in* $\mathbb{R}^m$.

The quantity $C_\alpha(A)$ in equation A.1 is called the α-convex hull of A. Now suppose we have a sample $S_n = \{X_1, \dots, X_n\}$ obtained from an object S in $\mathbb{R}^m$; then an estimator of S is $C_\alpha(S_n)$ if S is assumed α-convex. Estimators of α-convex objects are constructed from a 2D arc of circles in and 3D spherical caps in.

Definition 2. *Let $A \subseteq \mathbb{R}^m$ be α-convex. An α-shaped estimator of A is an estimator of its α-convex hull $C_\alpha(A)$ obtained by approximating an arc of circles with polygonal curves in 2D and spherical caps with polyhedral surfaces in 3D.*

A.2 Estimating the Time Delay and Embedding Dimension.

A.2.1 Estimating the Time Delay. There are two popular methods for estimating the time delay ρ: the autocorrelation function (ACF) and the average mutual information (AMI). Indeed, consider N measurements of a time series $s(t)$. Then the sample ACF is defined as

$$\rho(t) = \frac{\sum_{n=1}^{n} (s(n+t) - \bar{s})(s(n) - \bar{s})}{\sum_{n=1}^{n}(s(n) - \bar{s})^2}, \text{ with } \bar{s} = N^{-1} \sum_{n=1}^{N} s(n). \qquad (A.2)$$

The time delay is chosen as the $\rho = \min_{t>0}\{\rho(t) < 0\}$.

Now define the AMI as

$$I(t) = \frac{1}{N} \sum_{n=1}^{N} \mathbb{P}_r(s(n), s(n+t)) \log_2 \left(\frac{\mathbb{P}_r(s(n), s(n+t))}{\mathbb{P}_r(s(n)) \, \mathbb{P}_r(s(n+t))} \right), \qquad (A.3)$$

where $\mathbb{P}_r(s(n))$ and $\mathbb{P}_r(s(n), s(n+t))$ are, respectively, the probability of observing $s(n)$ and the probability of observing $s(n)$ and $s(n+t)$. The time delay is estimated as the first local minima of $I(t)$.

A.3 Estimating the Embedding Dimension.
The most popular method for estimating the embedding dimension m is the so-called false nearest neighbors technique (Kennel, Brown, & Abarbanel, 1992).

A.4 Kullback-Leibler Divergence Function. Given two distributions F and G of a continuous random variable, with respective density functions f and g, the Kullback-Leibler divergence function is defined as

$$d_{KL}(F, G) = \int_{-\infty}^{\infty} f(x) \log \left(\frac{f(x)}{g(x)} \right) dx. \qquad (A.4)$$

This quantity measures the degree to which *F* diverges from *G*. We will use it to assess the difference between the densities of the different cortex per task and thus the difference between their brain activities (see section 5.2).

A.5 Cross-Correlation Function. Given a time lag τ and two time-series $X = \{X_t\}$ and $Y = \{Y_t\}$, the CCF is defined as

$$CCF(X, Y, \tau) = \frac{\mathbb{E}[(X_{t-\tau} - \overline{X})(Y_t - \overline{Y})]}{\sqrt{\mathbb{E}[(X_{t-\tau} - \overline{X})^2]\mathbb{E}[(Y_{t-\tau} - \overline{Y})^2]}},$$

where $\overline{X}$ and $\overline{Y}$ are the respective mean of the time series X and Y. The CCF is then calculated over range of temporal lags and a "distance" $\rho(\cdot, \cdot)$ is defined as the maximum absolute CCF over the interval $[1, m_l]$ (for a maximum lag value of m_l to be selected):

$$\rho(X, Y) = \max_{1 \leq \tau \leq m_l} \{|CCF(X, Y, \tau)|\}.$$

Note that difference defines the similarity between the two time series.

Acknowledgments ────────────────────────────────

References ────────────────────────────────

Andrzejak, R. G., Lehnertz, K., Mormann, F., Rieke, C., David, P., & Elge, C. E. (2001). 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*, 06107. doi:10.1103/PhysRevE.64.061907.

Arbabshirani, M. R., Havlicek, M., Kiehl, K. A., Pearlson, G. D., & Calhoun, V. D. (2012). Functional network connectivity during rest and task conditions: A comparative study. *Human Brain Mapping*, *34*(11), 2959–2971.

Berger, B. (1929). Über das elektroenkephalogramm des menschen. *Arch. Psychiatr. Nervenkr*, *87*(1), 527–570.

Bruckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: Anatomy, function, and relevance to disease. *Ann. N.Y. Acad. Sci.*, *1124*, 1–38.

Carney, P. R., Myers, S., & Geyer, J. D. (2011). Seizure prediction: Methods. *Epilepsy and Behavior*, *22*, S94–S101.

Celso, G., Ott, E., Pelikan, S., & Yorke, J. A. (1984). Strange attractors that are not chaotic. *Physica D: Nonlinear Phenomena. Elsevier BV*, *13*(1–2), 261–268.

Celso, G., Ott, E., & Yorke, J. A. (1987). Chaos, strange attractors, and fractal basin boundaries in nonlinear dynamics. *Science*, *238*(4827), 632–638.

Chen, P. (2011). Principles of biological science. http://bio1152.nicerweb.com/.

David, S. A., Machado, J. A. T., Inácio, C. M. C., & Valentin, C. A. (2020). A combined measure to differentiate EEG signals using fractal dimension and complex structurization, a process of building complex geometrical structure. *Communications in Nonlinear Science and Numerical Simulation*, *84*, 1–13.

Destexhe, A. (1992). *Nonlinear dynamics of the rhythmical activity of the brain*. PhD diss., Université Libre de Bruxelles, Brussels, Belgium.

Destexhe, A., & Babloyantz, A. (1986). Low-dimensional chaos in an instance of epilepsy. *Proc. Natl. Acad. Sci.*, *83*, 3513–3517.

Destexhe, A., Sepulchre, J. A., & Babloyantz, A. (1998). A comparative study of the experimental quantification of deterministic chaos. *Physics Letters A*, *132*, 101–106.

Edelsbrunner, H., & Mücke, E. P. (1984). Three-dimensional alpha shapes. *ACM Transactions on Graphics*, *13*(1), 43–72.

Fisher, N. K., Talathi, S. S., Cadotte, A., & Carney, P. R. (2009). Epilepsy detecting and monitoring. In S. Tong & N. V. Thakor (Eds.), *Quantitative EEG analysis methods and clinical applications* (pp. 141–165). Norwood, MA: Artech House.

Gardiner, J. D., Behnsen, J., & Brassey, C. A. (2018). Alpha shapes: Determining 3D shape complexity across morphologically diverse structures. *BMC Evolutionary Biology*, *18*(184). doi:10.1186/s12862-018-1305-z.

Grassberger, P., & Procaccia, I. (1983). Measuring the strangeness of strange attractors. *Physica D. Nonlinear Phenomena*, *9*(1–2), 189–208.

Kennel, M., Brown, R., & Abarbanel, H. (1992). Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical Review A*, *45*(6), 3403–3411.

Kumar, J. S., & Bhuvaneswari, P. (2012). Analysis of electroencephalography (EEG) signals and its categorization: A study. *Procedia Engineering*, *38*, 2525–2536.

Kwessi, E. A., & Edwards, L. J. (2020). Artificial neural networks with a signed-rank objective function and applications. *Communication in Statistics-Simulations and Computations*. doi:10.1080/03610918.2020.1714659.

Lafarge, T., Pateiro-Lopez, B., Possolo, A. & Dunkers, J. P. (2014). R implementation of a polyhedral approximation to a 3D set of points using the α-shape. *J. Stat. Software*, *54*(4), 1–19.

Lehnertz, K., & Elger, C. E. (1998). Can epileptic seizures be predicted? Evidence from nonlinear time series analysis of brain electrical activity. *Physical Review Letters*, *80*, 5019–5022.

Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, *20*(2), 130–141.

Paladin, G., & Vulpiani, A. (1987). Anomalous scaling laws in multifractals objects. *Physics Reports*, *156*(4), 147–225.

Rickles, D., Hawe, P., & Shiell, A. (2007). A simple guide to chaos and complexity. *Journal of Epidemiology and Community Health*, *61*(11), 933–937.

Takens, F. (1981). *Lecture Notes in Mathematics: Vol. 898. Detecting strange attractors in turbulence dynamical systems and turbulence* (pp. 366–381). Berlin: Springer.

Zheng, J., Fushing, J., & Ge, L. (2019). A data-driven approach to predict and classify epileptic seizures from brain-wide calcium imaging video data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *17* (6).