Analysis of Localized Brain Activity

Olga Vsevolozhskaya
Department of Mathematical Sciences
Montana State University

May 8, 2009

A writing project submitted in partial fulfillment
Of the requirements for the degree

Master of Science in Statistics

APPROVAL

Of a writing project submitted by

Olga Vsevolozhskaya

This writing project has been read by the writing project director and has beeen found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the Statistics Faculty.

Date	Mark C. Greenwood Writing Project Director

Contents

Abstract

Electrodes implanted into the cerebral cortex of a macaque monkey allow one to observe a wide range of electrophysiological activity. Several methods exist for handling neurophysiological data, some of which rely on Fourier transformations, spectral power analysis or spatial-temporal modeling. I take a different approach where I consider neuronal signals emanating from cortical tissue as a sample of functions to be modeled. These functions are the consequence of a preliminary smoothing process, applied to observations taken discretely in time based on a millisecond frequency recording of the neuronal signals. The functional data analysis framework offers the required flexibility to reproduce the characteristics of the electrophysiological activity, and curve clustering is performed in an attempt to characterize a dynamic change in these activity patterns between locations. Analysis of functional maps, generated by considering locations with similar signals, by comparing distance matrices provides an idea of the expected time window over which one can observe a relative homogeneity of brain connectivity. The proposed methodology provides information on functional connectivity in the brain, establishes a framework to produce maps of distributed brain activity over a window of time, and addresses a question of possible similarity of the patterns of distribution across time.

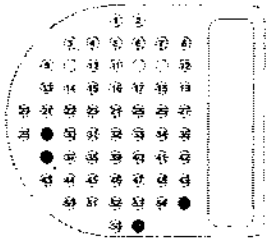
1 Introduction

The data in this study comes from an ongoing neurophysiology study investigating the neural processes that underlie visual perception. The recordings were taken by Dr. Charles Gray's lab at Montana State University. The main objective of this project is to measure neuronal activity in the cerebral cortex, and develop techniques to study the relationships between that activity and visually guided behavior. In order to study these perceptual and cognitive processes, a new type of electromechanical system was developed to implant 57 electrodes (channels) into an area of 1cm in diameter in the surface of the macaque's cerebral cortex. The recording chamber was hermetically sealed to reduce the chances of infection and help re-establish the normal balance of intracranial pressure (Gray et al. 2007). It also allows one to monitor the activities of multiple neurons in this region of the brain simultaneously. A printed circuit board grid (PCB grid), shown in Figure 1, provides connectivity between the electrodes and the control electronics, and also shows relative positioning of the electrodes in the surface of macaque's cerebral cortex. Some of the locations on the grid, that are represented by empty circles, were removed from the study because there were some technical problems during the mounting of the electrodes or electrodes never penetrated the cortex at those locations. The recordings were taken from a single macaque monkey, watching the original *Planet of the Apes* movie. The signals were sampled with a frequency of 1000 observations per second, then amplified (10K),

bandpass-filtered (0.6 kHz-6kHz), digitized (30 Khz/channel), and filtered using a lowpass filter of 1 Hz-100 Hz. Lowpass filtering allowed one to observe an ongoing low frequency background brain activity, representing what is called the local field potential (LFP). Each rhythm or band of frequencies in the local field potential can be associated with a particular behavior or mental state. For instance, Beta rhythm (12-30Hz) can be related to sensor-motor processing (Ba 2001; Senkowski 2005), and Delta rhythms (0-4Hz) are known to appear during sleep cycles (Steriade 2000). The analysis of similarities in the patterns of these rhythms is essential for understanding the underlying neurophysiological processes.

A common modeling strategy for neurophysiological data is to perform a Fourier transformation of the data and use Spectral power analysis for interpretation. Spatio-temporal modeling (Bowman et al., 2005) of localized brain activity may also be performed, which allows the production of spatially smoothed maps of distributed brain activity. A draw back of spatial modeling is that it assumes some structure of the connectivity inside the brain. For instance, it might assume that electrical measurements recorded from nearby locations should have a stronger relationship than locations spread further apart. Spatial models try to adjust for these types of relationships or spatial correlations. In the present paper, I consider an approach to explore/model the distribution of localized brain activity without any assumption of the underlying structure of connections, solely based on data-driven cluster analysis. I also address a question of a reasonable choice of time window over which one can expect to see a relative homogeneity in neuronal signals.

Figure 1: PCB grid

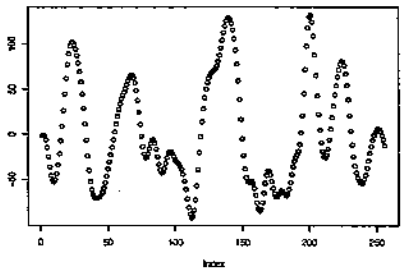


2 Methods

2.1 Functional Data Analysis

Functional data may be defined as data for which measurements were taken on a very fine grid of discretization, or as sparse data of a clearly functional nature. Figure 2 displays first 256 recordings from electrode channel 1 and appears clearly as a discretized curve. Because of the fineness of the temporal grid, one can consider recordings from each single electrode channel as a continuous curve. The first step in estimating this curve may involve some numerical approximation techniques or smoothing methods.

Figure 2: Neural Signals from Electrode Channel 1



Let y_{ij} denote an ith neuronal signal, $i = 1, ..., 5000$, received from the jth electrode channel, $j = 1, ..., 57$. We have a 5000 by 57 matrix of discrete observations in our data set which we can treat as a "noise" realization of 57 continuous smooth curves. As a first step we want to convert these discrete values to continuous functions of time, x_j, with values $x_j(t_i)$. We also want functions x_j to be smooth, meaning that they need to possess one or more derivatives. Notationally we can express this idea as

$$y_{ij} = x_j(t_i) + \varepsilon_{ij},$$

where the noise term ε_{ij} contributes a roughness to the raw data. Potentially, there are two sources of noise: discrete pointwise noise that is basically measurement error, and random variability in the curves, where each curve is a single realization from an underlying population with a true curve. Due to the lowpass filtering, we can exclude discrete pointwise noise, ε_{ij}, from the model. Several methods exist to estimate the function $x_j(t_i)$: kernel methods, Fourier series, spline-based methods, etc. I chose to work with B-spline cubic polynomial basis functions to estimate each function because this type of basis possesses good mathematical properties, such as continuous differentiability and integrability, relative to the types of observations encountered in this project.

2.1.1 Smoothing Splines

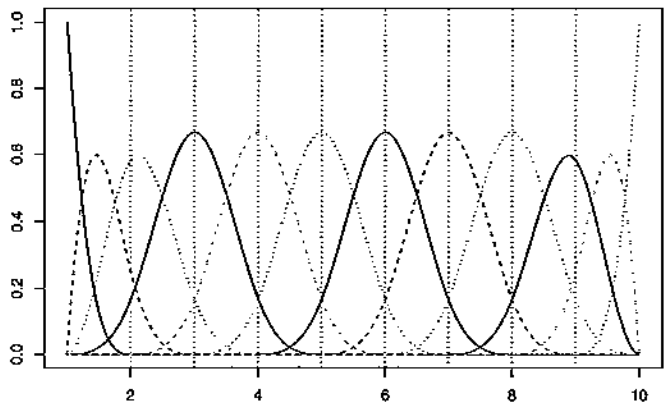
Functional data objects are usually constructed by specifying a set of mathematically independent basis functions, ϕ_k, and a set of coefficients, c_k, and then defining a weighted linear combination of these basis functions:

$$x(t) = \sum_{k=1}^{n} c_k \phi_k(t).$$

Periodic functions are usually approximated by a linear combination of sines and cosines through Fourier transformation. The most common choice of approximation system for non-periodic functional data is spline functions. They are constructed by smoothly joining together polynomials of specific order m at values τ_k, $k = 1, ..., n$, called knots. In practice it is very convenient

to use B-splines, the code for working with them is readily available in R in the fda package (Ramsay et al., 2008). Each basis function $\phi_k(t)$ in the B-spline basis is a spline function itself. The order of these basis spline functions in R by default is 4 which produces a continuous cubic polynomial that has continuous first and second derivatives on the interval $[\tau_k, \tau_{k+1}]$. These cubic polynomials are produced under the constraint that the functions must agree at the knots and the first and second derivatives must also agree. An illustration of B-spline basis functions defined on 10 knots is provided in Figure 3. The dotted vertical lines indicate the knot values. Four non-zero functions in between the vertical lines indicate an order 4 basis, and a linear combination of the three non-zero functions at each single dotted line produces a smooth cubic polynomial. The functions are not symmetric at the margins of the plot based on encountering the edges of the domain. To insure sufficient flexibility of the actual fit, I put a knot at each single time point t_i and define a B-spline basis over 256 observations, illustrated in Figure 4.

Figure 3: B-spline basis functions defined on 10 knots

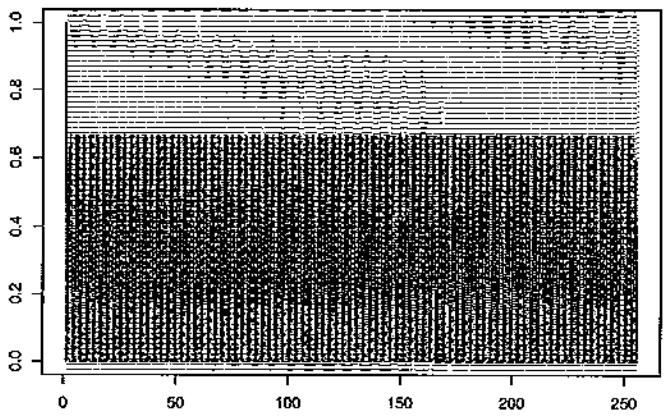


Once a B-spline basis is created, one can use a function, once again readily available in the fda package, to smooth the data and estimate the B-spline coefficients subject to a roughness penalty. The degree of smoothness is controlled by a smoothing parameter λ with smaller values meaning a more flexible fit, closer to a typical least squares estimate. I chose a value of λ to be 0.01 which produced only minimal smoothing in this application and a near perfect fit to the smooth discrete observations. This choice is based on the knowledge of prior lowpass filtering and a visual evaluation of the fit, with the example shown in Figure 5.

2.1.2 Measure of Proximity Between Two Curves

Let $\mathbf{x} = (x_1, ..., x_{57})$ be a set of curves pertaining to the 57 implemented electrodes. The goal is to partition $\mathbf{x}$ into subsets of curves with homogeneous neuronal signals, emanating from cortical

Figure 4: B-spline basis functions defined on 256 knots



tissue over a certain time window. To perform the partitioning one must define a measure of proximity between two curves, which appears to be a non-trivial task.

In classical multivariate clustering algorithms, the object to be partitioned arrives in the form of finite-dimensional vector and a classical norm can be used to define the distance between two vectors. Because in a finite dimensional Euclidian space, say R^p, there is an equivalence between all norms, the choice of the measure of proximity, in a mathematical sense, is not crucial. For instance, if we let $\mathbf{x}, \mathbf{y}$ be two vectors in R^p; then, the classical Euclidean norm is defined $\| \mathbf{x} - \mathbf{y} \|^2 = \sum_{j=1}^{p} (x_j - y_j)^2 = (\mathbf{x} - \mathbf{y})'(\mathbf{x} - \mathbf{y})$, and a whole family of norms can be deduced based on the Euclidian norm by using different positive definite matrices, $\mathbf{M}$, in the following way $\| \mathbf{x} - \mathbf{y} \|_{\mathbf{M}}^2 = (\mathbf{x} - \mathbf{y})'\mathbf{M}(\mathbf{x} - \mathbf{y})$. In particular, by a specific choice of $\mathbf{M}$ one can define the Mahalanobis distance.

Now, since we have a set of continuous curves, with an infinite number of observations per curve, we are considering an infinite dimensional space. In infinite dimensional space, the equivalence between norms fails and the choice of the preliminary norm becomes crucial. I tackled the problem by defining a measure of distance over a window of time a to b by

$$D_{ij}^2 = \int_{[a,b]} (x_i - x_j)^2 dt.$$

This distance measure satisfies the properties of the distance metric:

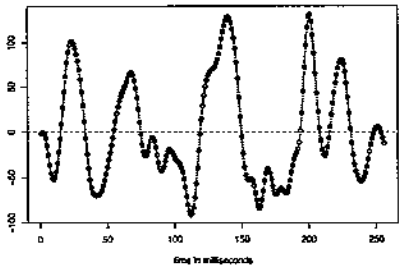
positive definiteness: $d(x, y) \geq 0$, and $d(x, y) = 0 \Longleftrightarrow x = y$

symmetry: $d(x, y) = d(y, x)$

triangle inequality: $d(x, z) \leq d(x, y) + d(y, z)$,

and is equivalent to the Euclidean norm. While not used here, functional data analysis opens

Figure 5: Estimated Smooth Curve



up the possibility of a derivative based semi-metric

$$D_{ij}^2 = \int_{[a,b]} (x_i' - x_j')^2 dt,$$

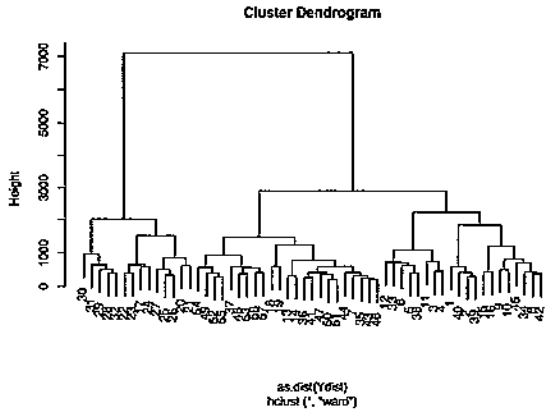
which does not necessarily satisfy the positive definiteness property. In particular, it is possible for the distance between two objects to be zero, $d(x, y) = 0$, with two objects not being equal to one another, $x \neq y$. As an example, consider two parallel lines. The values of their respective derivatives are identical, however the two lines are not the same. The computation of a semi-metric based on derivatives is more sensitive numerically. In particular, edge effects become a bigger issue than in the original functional estimates.

2.1.3 Curve Clustering

In the previous sections it has been shown that neuronal signals received from the electrode channels can be treated as realizations of continuous smooth curves. In this section, I describe a method of data-driven cluster analysis that attempts to define groups with a high degree of within-cluster homogeneity in brain functions.

Several clustering algorithms exist, including single linkage, complete linkage, etc. In the present paper, I consider the Ward's minimum variance clustering algorithm (Ward 1963) which is an agglomerative algorithm. In general, agglomerative algorithms start with each object in a separate class, and then amalgamate the"most similar" classes until all of the objects belong to a single group. Ward's method, in particular, uses sums of squared deviations from centroids of clusters to calculate the similarity measure between different groups. It starts the partitioning by assigning each single object to its own cluster. At this point, the sum of squared distances between the objects and the centroids of the clusters is 0. As clusters form, the centroids move away from the actual objects and the sum of the squared distances from the centroids increase. Each subsequential amalgamation is done in such a way so that the sum of squared deviations increases as little as possible. Figure 6 contains a dendrogram produced by Ward's clustering algorithm of the 57 estimated smooth curves over the first 256 milliseconds of recordings.

Figure 6: Ward's Method Dendrogram



2.1.4 Stopping Rules

The next dilemma one usually encounters in cluster analysis is the selection of the number of clusters or partitions to use in the final solution. Unfortunately, virtually all clustering procedures provide little, if any, information as to the number of clusters present in the data. Sometimes the choice of the number of clusters could be driven by knowledge of the data, but with the present data set we do not know how many and which groups might be expected. So, to determine the preferred number of clusters, I used a more formal method referred to as a "stopping rule". Stopping rules evaluate a measure, *G(c)*, of the goodness of partition into c clusters, and identify the value of c for which *G(c)* is optimized.

I chose to use the *G3(c)* index (Milligan, et al 1985) that determines the optimal number of clusters by evaluating the standardized sum, $D(c)$, of all within-cluster dissimilarities in a partition of the objects into c clusters. It is defined as

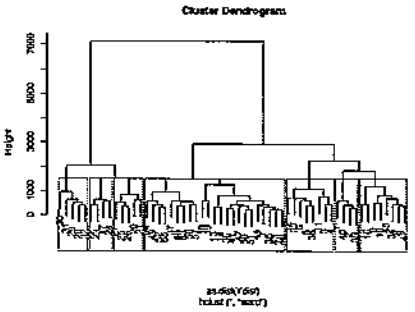
$$G3(c) \equiv (D(c) - min(D(c)))/(max(D(c) - min(D(c))).$$

The index was found to exhibit excellent recovery characteristics by Milligan (1985), and its minimum value across the hierarchy levels is used to indicate the optimal number of clusters.

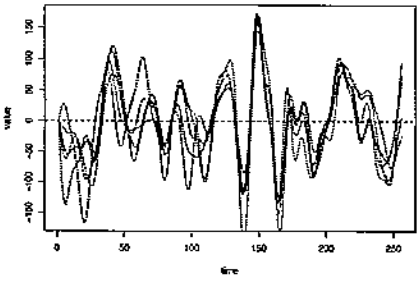
As an example, consider the dendrogram in Figure 6. To determine the optimal number of clusters over the first time window of 256 milliseconds I evaluated the *G3(c)* index for $c = 2, 3, ..., 10$. I capped the possible number of groups at 10 to maintain an easier visual evaluation of the partitioning. For the dendrogram in Figure 6, the values found were: $G3(2) = 0.135$, $G3(3) = 0.185$, $G3(4) = 0.143$, $G3(5) = 0.113$, $G3(6) = 0.098$, $G3(7) = 0.088$, and $G3(8) = 0.12$. It is clear that the value of *G3* is minimized for $c = 7$. Figure 2.1.4 illustrates the optimal cut of the dendrogram and shows curves grouped according to their cluster membership. One can essentially think that all the functions in a particular group came from the same population with some mean curve and the differences between them are due to sampling variability in the curves. Most of the curves from the same group tend to agree at the peaks but the timing of the peaks varies from cluster to cluster. Ignoring this timing variation in computing a cross

sectional mean function can result in an estimate of average pattern that does not resemble any of the observed curves. This might be a big issue in a functional data analysis that combines information across curves, however it is not our intention to estimate an underlying population mean curve. We rather want to provide a visual evaluation of the clustering method based on the similarity of the curves that were grouped into the same cluster.

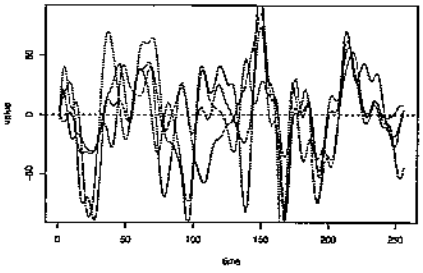
Dendrogram



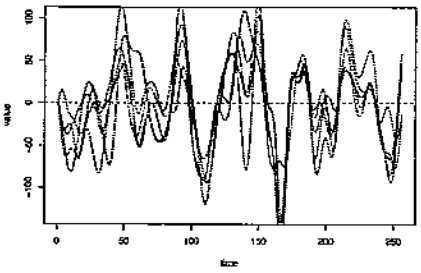
Cluster1



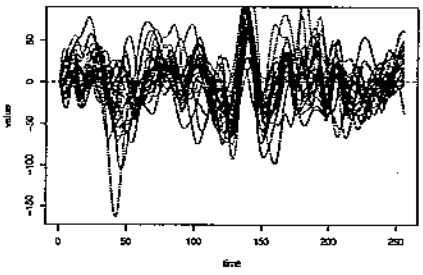
Cluster2



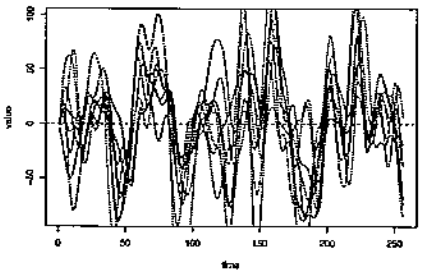
Cluster3



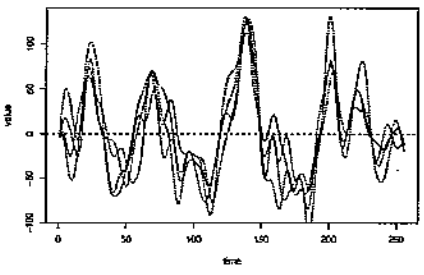
Cluster4



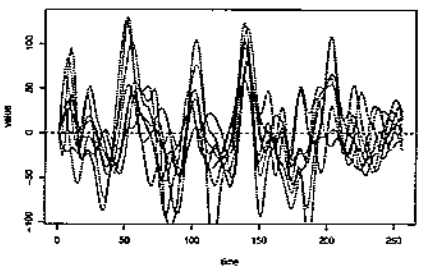
Cluster5



Cluster6



Cluster7



2.1.5 Visualization Methods

The ultimate goal of my functional data analysis was to identify the distribution of patterns of localized brain activity and visualize the results by producing cluster maps. Figure 2.1.5 presents the plot of the 57 locations on the printed circuit board grid. Locations are grouped into clusters and each cluster is depicted by a unique color and symbol. The data-driven cluster analysis attempts to define groups with a high degree of within-cluster homogeneity in brain function. However, cluster analysis generally does not quantify or test the degree of association between intracluster locations. To estimate dependence among clusters a non-metric multidimensional scaling (MDS) plot is produced along with the plot of the cluster maps (Venables at el., 2002).

Multidimensional scaling is a generic name for a body of procedures and algorithms that start with an ordinary proximity matrix and generate configurations of points in one, two, three or higher dimensional space such that distances between the points in that space best match those of the distance matrix. This task of embedding points in a space is non-trivial. For instance, it is clear that the two points can be placed on a line to match the dissimilarities between two objects. Configuration of three points on a two-dimensional plane can always be defined in such a way that interpoint distances exactly match the dissimilarities among three objects. In general, n points in a metric space can always be embedded in an $(n-1)$-dimensional space so as to exactly recreate the proximities among objects.

Now, we begin with a 57×57 distance matrix of dissimilarities d_{ij} with a 256-dimensional configuration of points $(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{57})$, where

$$\mathbf{x}_i = [x_{i1} x_{i2} ... x_{i256}]'.$$

It is obvious that we can not consider configurations in a 56-dimensional space. However, we can squeeze a high-dimensional point cloud into a small number of dimensions while preserving as well as possible the inter-point distances.

There are different forms of multidimensional scaling: classical form, Sammon's (1969) non-linear mapping, and a non-metric version. In the latter one the idea is to choose a configuration to minimize

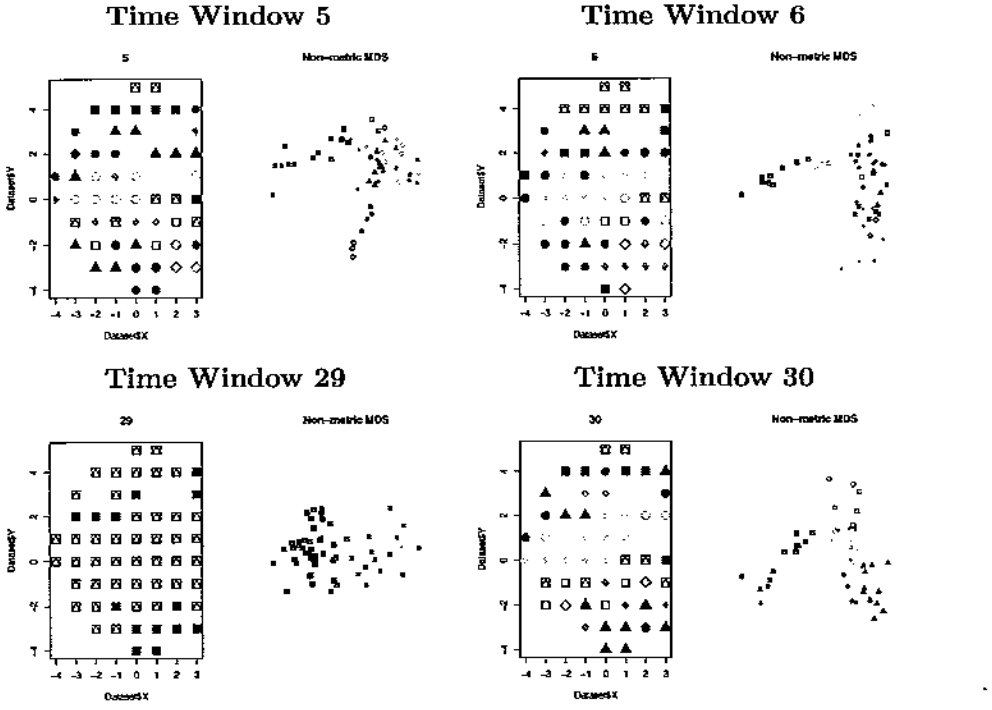
$$STRESS^2 = \sum_{i \neq j} [\theta(d_{ij}) - \tilde{d}_{ij}]^2 / \sum_{i \neq j} \tilde{d}_{ij}^2,$$

over both configuration of points and an increasing function θ. Location, rotation, reflection and scale of those configurations are all indeterminate which reflects a non-metric nature to the method. $\tilde{d}_{ij}$ is the measure of distance in the configuration space or Minkowski metric defined by

$$\tilde{d}_{ij} = \left(\sum_{k=1}^{m} |x_{ik} - x_{jk}|^r \right)^{1/r}.$$

Sammon (1969) defined the mean-squared error between the two sets of distances, E, which is similar to a $STRESS$ criterion from non-metric multidimensional scaling, so the two methods are similar, however Sammon's function puts much more stress on reproducing small distances accurately. We used a non-metric version of MDS to try to capture global, not local, structure of configurations; to produce few tight groups to estimate the dissimilarities among the clusters. The reason for this behavior is that $STRESS$ weights large distances more heavily than small ones.

For an easy visualization, we considered configurations in R^2, however it is hard to determine whether or not the interpoint distance in a configuration is a reasonably accurate reflection of the original proximities. It is quite possible that the minimal dimensionality for the configurations should be greater than 2. Overall, 37 plots were produced over time windows of 256 milliseconds and an overlap of 128 milliseconds. The size of the time window and overlap were chosen due to numerical efficiency of spectral computations not discussed in this paper. Also, it seemed unreasonable to consider a time window shorter than a quarter of a second since any significant divinations in the patterns of electrical signals might not have been detected with too few observations in each time window.



Some spatial configurations on the PCB grid are similar across time windows, as shown in the upper panel of Figure 2.1.5. Others, on the contrary, differ dramatically (lower panel of Figure 2.1.5). 17 out of 37 times 10 groups were chosen to be an optimal number of partitions according to the $G3$ index, which indicates that we probably need to increase the possible number of clusters to get a more appropriate representation of the spatial connectivity. Most of the MDS plots were quite difficult to interpret. As an example see Time Windows 5 and 6, that do not project well into a small number of dimensions. Normalized STRESS values in two dimensions for all time windows was over 0.10. Kruskal (1964) suggests a threshold of 0.05 for the STRESS of a "good" configuration. Normalized STRESS values of about 0.15 in three dimensions and about 0.05 in 9 dimensions indicate that perhaps, even for a non-metric solution, more than two dimensions are needed to provide an adequate representation of such a large number of channels.

2.2 Analysis of Similarities Across Time Windows

It's been mentioned above that the choice of temporal window was quite arbitrary and by and large based on the numerical efficiency of computations. Therefore, it is of interest to estimate a reasonable choice of time window over which one can expect to observe a relative homogeneity in the distribution of the localized brain activity. In the next section I am going to discuss a statistical approach to quantify temporal changes in the overall spatial-patterns of brain connectivity. In particular, I am going to evaluate the similarities in spatial connectivity configurations using a proximity measure based on the Mantel statistic, a well known statistical tool designed specifically to quantify similarities between distance matrices.

2.2.1 Mantel Statistic

The Mantel test was first developed in 1967 to correlate temporal and spatial distributions of cancer incidents. It provides a measure of association between distance matrices or more generally proximity matrices based on the degree of the relationship between two sets of variables taken at the same sampling locations. The Mantel approach consists of calculating the correlation between the values of two upper (or lower) triangle parts of the square symmetric distance matrices A_{ij} and B_{ij}. The statistic used to measure correlation between two matrices is the classical Pearson correlation coefficient:

$$r = \frac{1}{N-1} \sum_{i=1}^{N} \sum_{j=1}^{N} \left[\frac{A_{ij} - \overline{A}}{s_A} \right] \left[\frac{B_{ij} - \overline{B}}{s_B} \right]$$

where N is the number of elements in the lower or upper triangular part of the matrix, $\overline{A}$ is the mean for the A elements and s_A is the standard deviation of A elements. Note that the coefficient r measures only linear correlation, therefore if non-linear relationships exists, they may be missed. However, it is possible to allow for non-linear relationships by finding the Spearman correlation instead of Pearson.

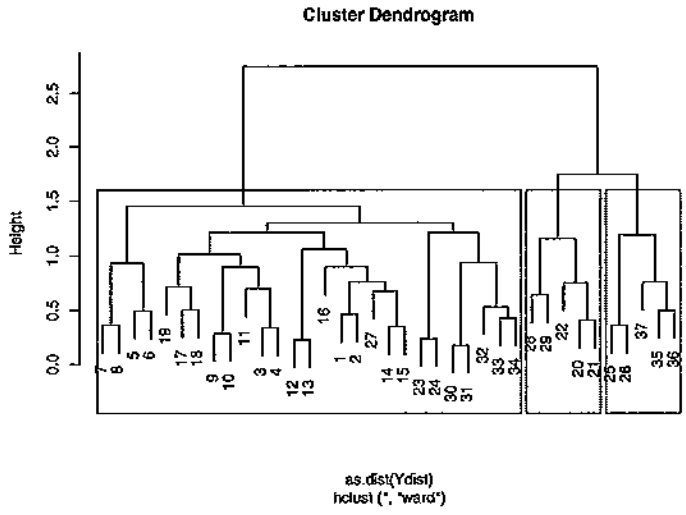
2.2.2 Measure of Proximity Between Time Windows

I have a set of 37 distance matrices of size 57×57 computed over 4608 milliseconds of recordings. The rows and columns of all matrices correspond to the same channels. The first step is to find the Mantel statistic by computing the Pearson correlation coefficient between the corresponding elements of the upper triangular matrices. Next, the correlation is converted into a distance (Hastie et al. (2009)) by defining a proximity measure between distance matrices of

$$D_{ij} = \sqrt{2(1 - Mantel)}.$$

Now, all of the pairwise comparisons of proximity matrices corresponding to 37 different time windows are organized into a new 37×37 distance matrix. The Ward's minimum variance method and the $G3$ index stopping rule are applied to the new distance matrix to cluster time windows over which one can observe relatively stable spatial connections. The value of $G3$ is minimized for $c = 3$. Figure 7 illustrates the optimal cut of the dendrogram. A multidimensional scaling plot is also produced to visualize these results.

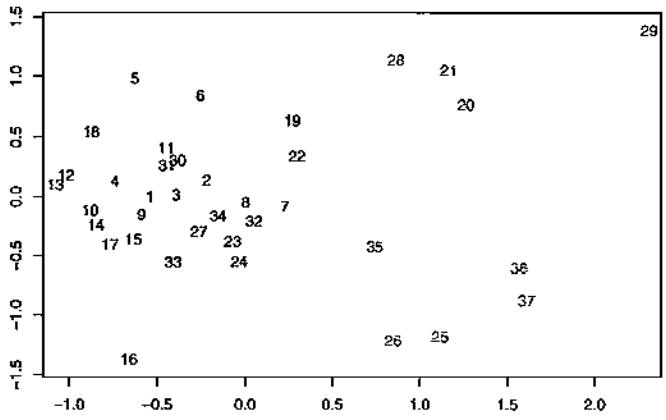
Figure 7: Dendrogram of Between Window Dissimilarities



2.2.3 Visualization of the Results

A two-dimensional configuration of time windows using a non-metric MDS method is shown in Figure 8. The 37 time windows in Figure 8 have been labeled by color to highlight the separation of the three clusters. These three clusters constitute three separate patterns of spatial brain connectivity. A normalized STRESS value of 0.1736 indicates that probability more than two dimensions are needed to provide an adequate representation of time window configurations. All of the first 19 time windows were put into the same group indicated by the black color which displays a similar pattern of the overall brain connectivity over these time windows. Then we observe a slight deviation, over time windows 20, 21, and 22 from the overall pattern of the spatial brain connection observed over the first 2432 milliseconds. These time windows represent the second type of brain connectivity pattern marked by the red color. Afterwards, over time windows 23 and 24, we return to the "black" pattern witnessed previously. The time windows 25 and 26 were put into a 3rd group which represents yet a third pattern of the spatial brain connectivity labeled as "green". The remaining time windows jump from pattern to pattern in the following fashion: 27 "black", 28, 29 "red", 30-33 "black", and 35-37 "green". This "jumpiness" probably indicates that the monkey was not "focused" over that time period. To further understand the temporal distribution of these data it would be helpful to know the specific visual images the macaque monkey was viewing and the mental state she was in at the time the recordings were made.

Figure 8: Non-metric MDS mapping of Time Windows



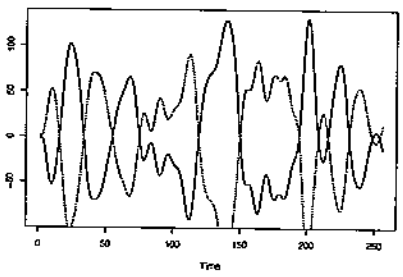
3 Limitations of Methods

The time domain or functional data analysis approach used does not accommodate time/phase delay or negative correlations between a pair of channels. Both are thought to be acceptable for showing connections in local field potential. For example, consider Figure 9. The red line estimates the LFP function for the first 256 milliseconds of recordings from Channel 1. The black line is a mirror image of the red one, and was artificially produced to illustrate the idea of perfect negative correlation. The dynamics of change of these two curves should be considered to be identical, however an integrated squared difference between the two is obviously non-zero. In future projects, we would like to incorporate a possible time delay of up to 15 milliseconds, as well as possible negative correlation, into a measure of proximity between two curves. We also would like to analyze more "exciting" data sets with more channels and measurements that were taken under certain experimental conditions. Finally, we would like to explore the semi-metric based on derivatives to estimate a measure of proximity between two curves.

4 Discussion

In this study we considered an approach to explore/model spatio-temporal changes in localized brain activity. Functional data analysis was found to be a reasonably useful approach to estimate electrical signals recorded from cortical tissue. We defined a measure of proximity between two curves and performed cluster analysis to estimate a spatial distribution of localized brain activity. Next, assuming there exists a certain dependency across time windows, we performed cluster analysis on the next level of hierarchy by defining a distance measure between functional maps based on the Mantel statistic. We summarized the results by producing plots of distributed

Figure 9: Illustration of two neuronal signals that should be considered identical to one another



brain activity by groups on original space and in a multidimensional scaling plot. The proposed analysis may serve as a first step to understanding spatial-temporal connectivity of the brain.

5 References

1. Bowman, F. et al. (2005) Spatio-temporal modeling of localized brain activity, *Biostatistics*,558-575, doi:10.1093/biostatistics/kxi027.

2. Ferraty, F., Vieu, P. (2006) *Nonparametric Functional Data Analysis*, Springer.

3. Gordon, A. D. (1999) *Classifications*, Chapman.

4. Gray, C. (2007) Multichannel Micromanipulator and Chamber System for Recording Multi-neuronal Activity in Alert, Non-Human Primates, *Neurophysiology*, 527-536, doi:10.1152/jn.00259.2007

5. Hastie, T., et al. (2009) *The Elements of Statistical Learning*, Springer.

6. Hegde, A., et al. (2006) On Spatio-Temporal Dependency Changes in Epileptic Intracranial EEG: A Statistical Assessment, *Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE*, 6711-6714.

7. Hennig, C. (2007). fpc: Fixed point clusters, clusterwise regression and discriminant plots. R package version 1.2-3.

8. Jain, A. et al.(1988) *Algorithms for Clustering Data*, Prentice Hall.

9. Milligan, G. et al. (1985) An examination of procedures for determining the number of clusters in a data set, *Psychometrika*, 50(2), 159-179.

10. Ramsay J. O., et al. (2008). fda: Functional Data Analysis. R package version 2.1.1

11. Ramsay, J.O, Silverman, B.W,(2005) *Functional Data Analysis*, Springer.

12. Venables, W., Ripley, B. (2002) *Modern Applied Statistics with S*, Springer.