Anomaly Detection in Brain Connectivity Structure: An Application to Epilepsy

by

Andrew Sweet

B.A., Artificial Intelligence, University of Sussex, 2008

M.Sc., Computer Science, University College London, 2009

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science
in Electrical Engineering and Computer Science
at the Massachusetts Institute of Technology

February 2013

Signature of Author: ______________________________
Department of Electrical Engineering and Computer Science
February 1, 2013

Certified by: ______________________________
Polina Golland
Associate Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by: ______________________________
Leslie A. Kolodziejski
Professor of Electrical Engineering and Computer Science
Chair, Committee for Graduate Students

Anomaly Detection in Brain Connectivity Structure: An Application to Epilepsy

by Andrew Sweet

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of
Master of Science

Abstract

In this thesis, we study approaches for detecting anomalous regions in brain connectivity networks estimated from resting state fMRI. We are motivated by the problem of localizing diseased regions to be resected in pre-surgical epilepsy patients. Our goal is to investigate the potential of these non-invasive connectivity approaches to augment and even replace the clinical gold standard for localization, which requires invasive implantation of electrodes onto the surface of the brain. We focus on adapting an existing method that detects anomalies from a small set of large candidate regions in a population of patients. The main contribution of the work is to develop this method for our application, so that it can efficiently identify anomalies from a large set of small candidate regions in a single epilepsy patient. We find that standard statistical approaches identify regions that overlap reasonably well with electrode recordings of abnormal activity, but are sensitive to manual parameter selection. Our method matches this performance, but has the advantage of automatically determining its corresponding parameters. While localization is not generally accurate enough to consider replacement of invasive electrode implantation, the method discovers potentially diseased regions that may better guide electrode placement.

Thesis Supervisor: Polina Golland
Title: Associate Professor of Electrical Engineering and Computer Science

Acknowledgments

First and foremost, I want to acknowledge my supervisor, Polina Golland, for her patience and advice, not only on this thesis, but throughout my time at MIT. Furthermore, I extend my gratitude to the member's of Polina's group, who have made this time an enriching experience. Special appreciation goes to Georg Langs, who I enjoyed working with when I first arrived, and Archana Venkataraman who has been invaluable in the last few months. I hope that I have made at least a small contribution to the group in my time here. I also want to thank Steve Stufflebeam for introducing me to this project. Surgical planning is a scary business, but learning a little about the science and practice behind epilepsy has been a fascinating experience. My thanks also go to Naoro Tanaka and Hesheng Liu for spending the time to provide me with the epilepsy data used in this thesis.

On a personal note, I am grateful to my friends and family for supporting my decision to come to MIT. The institute has impressed me to no end and I've met some exceptional people here that I hope to stay in contact with throughout my life. Special thanks goes to Emily Xi Lin for providing tea and company in our lengthy study sessions, and the residents of Pleasant House, who have made Cambridge feel like a second home.

Contents

Abstract iii

Acknowledgments iv

List of Figures ix

1 Introduction 1
1.1 Epilepsy & The Epileptogenic Zone 1
1.2 Scope of the Research 4
1.3 Outline of the Thesis 4

2 Background 5
2.1 Intracranial EEG 5
2.2 Resting State fMRI 7
2.2.1 Statistics of Functional Correlations 11
2.2.2 Statistics of Graph Theoretical Features 12
2.2.3 Generative Models of Functional Connectivity 14

3 Methods 17
3.1 Generative Model of Anomalous Regions 17
3.2 Mean Field T-Algorithm 21
3.3 Mean Field $T\tilde{F}$-Algorithm 26
3.4 Summary of the Algorithms 30

4 Experiments 31
4.1 Synthetic Experiments 31
4.1.1 Data 31
4.1.2 Evaluation 33
4.1.3 Results 33
4.2 Epilepsy Experiments 36
4.2.1 Data & Pre-processing 36
4.2.2 Evaluation 37

4.2.3 Coarse Parcelation Results 38
4.2.4 Fine Parcelation Results 43
4.3 Summary 45

5 Discussion **47**
5.1 Contributions 47
5.2 Limitations & Future Work 48

Bibliography **51**

List of Figures

1.1 Regions that characterize the epileptogenic zone 2

2.1 Intracranial EEG measurements of epileptogenic regions 6
2.2 Functional MRI measurements of epileptogenic regions 8
2.3 Correlations between the fMRI time series of cortical regions 10
2.4 Computing region statistics of correlations 12
2.5 Computing region statistics of graph features 14
2.6 A generative model of anomalous regions 15

3.1 Directed graph of our generative model 20
3.2 Directed graph of anomalous regions and connections 21

4.1 Two mixtures of Normals used for synthetic experiments 32
4.2 Performance of inference algorithms on ideal synthetic data 34
4.3 Performance of inference algorithms on real synthetic data 35
4.4 Coarse and fine cortical parcelations 37
4.5 Rigid alignment of post-implantation CT with pre-implantation MRI . . 38
4.6 Projection of pial results to the dural surface 39
4.7 Parameter sweep for the coarse parcelation 40
4.8 Epileptogenic zone detection results for the coarse parcelation 42
4.9 Parameter sweep for the fine parcelation 44
4.10 Epileptogenic zone detection results for the fine parcelation 46

5.1 Epileptogenic zone detection results using the mean posterior 50

Chapter 1

Introduction

In this chapter, we introduce the application and scope of this thesis. In Section 1.1, we briefly describe epilepsy as a disorder, the importance of surgical resection of epileptogenic areas and how these areas are imaged for surgical planning. We then clarify the goals of the research in Section 1.2, and provide an outline of the thesis in Section 1.3.

■ 1.1 Epilepsy & The Epileptogenic Zone

Epilepsy is a chronic disorder symptomatically characterized by seizures. These seizures can be defined as transient manifestations of abnormally excessive neuronal activity, or spiking, in the gray matter of the cerebral cortex [19]. The International League Against Epilepsy defines two main categories of seizure: *generalized* seizures, which begin simultaneously across the cortex; and *partial* seizures, which begin in one or a few focal regions and spread to affect a larger areas of the brain [26]. While the basic mechanisms of these seizures are understood to include molecular and cellular abnormalities, it is clear that epilepsy is a function of large populations of synchronously active neurons and abnormalities in widespread brain networks [23, 25].

Epilepsy is a relatively common neurological disorder with a prevalence rate of around 0.5% [19]. Initial treatment involves prescription of anticonvulsant medication, which prevents seizures for around 80% of patients [11]. When medication is ineffective and the patient experiences partial seziures, the most appropriate course of action is surgical resection of the area causing the seizures, which is known as the epileptogenic

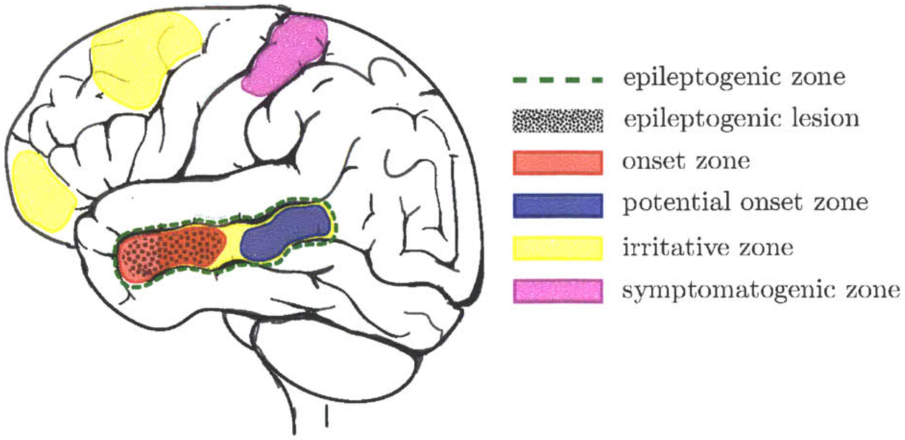

Figure 1.1: The epileptogenic zone is characterized by several regions. The most important are the epileptogenic lesion, when it is observable, and the seizure onset zone. Resection of these regions may not always eliminate seizures because of other potential onset zones. The irritative zone may include these potential onset zones, but may also include other non-epileptogenic regions. The symptomatogenic zone is the area of eloquent cortex that the seizure activity spreads to first and typically does not overlap with the epileptogenic zone.

zone [12]. Accurate localization of this area promises to reduce the amount of tissue to be removed, thus limiting potential damage to brain function.

The epileptogenic zone is specifically defined as the minimal area of cortex that must be resected or completely disconnected to eliminate seizures [28]. During presurgical evaluation, it is characterized by several areas, which are illustrated in Figure 1.1.

In some patients, the epileptogenic zone is fully characterized by the epileptogenic lesion, which is a macroscopic lesion either caused by a proximal epileptogenic region or is epileptogenic itself. This is convenient because magnetic resonance imaging (MRI) is non-invasive and can be used to obtain a high resolution scan of the lesion.

Although a lesion is almost always present, it may be microscopic and not observable from conventional MRI [28]. In these difficult cases, functional measurements of the brain are used to identify regions that exhibit the abnormal spiking behavior indicative of seizure activity. This spiking may be *ictal*, occurring at the onset of a seizure, or *interictal*, occurring between seizures.

The onset zone is the area of cortex that exhibits the initial ictal spiking of a seizure. In some cases, this is identical to the epileptogenic zone. However, complete resection of the onset zone does not always eliminate seizures, due to the suspected presence of potential onset zones that continue to initiate seizures after surgery. The onset zone can be coarsely localized using scalp electroencephalography (sEEG), which is non-invasive, but only detects spiking that synchronizes over a relatively large area of the cortex [1]. Intracranial EEG (iEEG), where electrodes are implanted directly onto or into the brain, can refine this localization, but requires invasive surgery that increases risk to the patient's health.

The irritative zone is the area of cortex that exhibits interictal spiking and is usually more extensive than the epileptogenic zone. It may include potential seizure onset zones that are not observable from ictal measurements, but it may also include areas that are only peripherally related to seizure events. Only magnetoencephalography (MEG), sEEG and iEEG have the necessary temporal resolution to reliably detect interictal spiking, but these modalities are limited by their relatively low spatial resolution [28]. Functional MRI (fMRI), which has a lower temporal resolution, but higher spatial resolution, may also be used to localize the irritative and onset zones, but its reliability is unclear [33].

Finally, the symptomatogenic zone is the area of eloquent cortex that, when activated, produces the initial ictal symptoms. For example, a patient may initially exhibit seizure related tremors in their left arm, indicating that the symptomatogenic zone is contained within the right motor cortex. Typically, the symptomatogenic zone does not overlap with the epileptogenic zone, but is simply the area of eloquent cortex that the seizure activity spreads to first.

■ 1.2 Scope of the Research

Our goal is to investigate the accuracy of using brain connectivity anomalies derived from resting state fMRI for localization of epileptogenic regions. The proposed approach offers two potential advantages for surgical planning in epilepsy. First, fMRI may provide additional information alongside sEEG and MEG for coarse localization, which is used to determine where iEEG electrodes should be placed. If connectivity analysis reliably detects regions that overlap with neural spiking, then other detected areas may be suitable candidate locations for electrodes. Second, non-invasive localization using fMRI has the potential to ultimately replace the invasive procedure of iEEG.

Most resting state fMRI studies are exploratory in nature, and are rarely compared against anything resembling ground truth. By contrast, we evaluate results with respect to iEEG spiking labels and design a principled approach to such evaluation. Furthermore, existing connectivity analysis techniques typically restrict attention to a small number of large regions. But since epileptogenic regions may be small, we must adapt existing methods to scale well with large numbers of small regions.

■ 1.3 Outline of the Thesis

The remainder of this thesis is organized as follows. In Chapter 2, we introduce the measurements used in this work and review existing approaches of brain connectivity analysis, taking special consideration of methods that detect anomalous connections and regions in patients with neurological disorders. In Chapter 3, we adapt a previously demonstrated model of anomalous region detection in a population of patients to the problem of anomaly detection in a single patient and present efficient inference algorithms that scale well as the number of regions grows. In Chapter 4, we compare the algorithms on synthetic data to evaluate performance and empirical run time, and also compare detection accuracy on clinical data with existing approaches. In Chapter 5, we highlight the contributions of the thesis and suggest future research directions that may help to improve epileptogenic zone localization.

Chapter 2

Background

In this chapter, we introduce the measurements, typical pre-processing steps and brain connectivity analysis techniques. In Section 2.1, we present intracranial EEG as the gold standard for epileptogenic zone localization and the challenges associated with using it for validation of anomaly detection in the cerebral cortex. In Section 2.2, we introduce resting state fMRI as a measurement of brain connectivity and discuss pre-processing steps. Furthermore, we review analysis techniques that discover regions associated with deviations from healthy brain connectivity in patients with neurological disorders.

■ 2.1 Intracranial EEG

EEG Electrodes can be used to measure electrical activity in the brain caused by voltage fluctuations [35]. These electrodes are typically placed on the scalp and provide a noisy measurement of neural activity. For pre-surgical evaluation of some epilepsy patients, sEEG measurements are used to provide a coarse localization of the onset zone. In order to refine this localization, a craniotomy is performed and iEEG electrodes are implanted onto the surface of the brain. Due to the location of these electrodes, iEEG measures electrical activity in the cortical gray matter more accurately than scalp EEG or MEG, especially in regions far from the exterior of the head, such as the inferior temporal and frontal lobes [8]. However, for similar reasons, even iEEG provides poor measurement of subcortical activity. The measurements also have a limited spatial resolution, with a typical distance between electrodes of around 10mm.

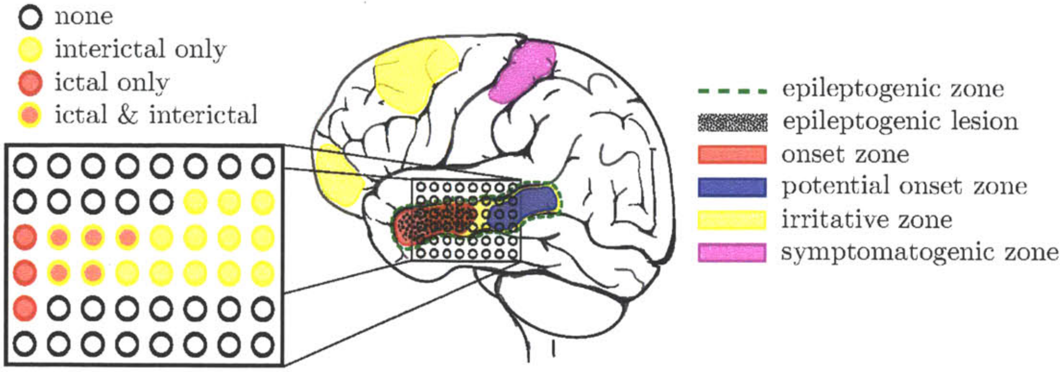

Figure 2.1: In iEEG, electrode arrays and strips are placed over a limited area of the cortical surface that is suspected to contain the onset zone. An epileptologist labels electrodes that exhibit abnormal ictal and interictal neural spiking.

Electrical activity in epilepsy patients is recorded for roughly one week, during which a patient typically experiences a few seizures. An epileptologist studies the observed activity and labels electrodes that exhibit abnormal ictal and interictal neural spiking, as shown in Figure 2.1.

Typical presurgical planning involves acquisition of a post-implantation CT volume, which is used to identify the actual location of the implanted electrodes. Modern electrodes produce a CT intensity higher than brain, bone and most connector wires, so can be localized with relative ease. As the number of electrodes is typically less than one hundred, manual localization is a practical approach. A semi-automated method for electrode localization has been previously demonstrated [48], but it still requires manual identification of a few electrodes and also requires a post-implantation T1 MRI volume, which is not always available.

Our goal is to evaluate the accuracy of epileptogenic zone localization in the cortex of the brain. The cortex can be extracted from a pre-implantation T1 MRI volume using the Freesurfer software package [15]. Initially, the post-implantation electrode locations need to be transferred to the pre-implantation T1 volume. Despite the displacement of tissue and bone due to the craniotomy, and the presence of the electrode arrays, there is usually enough structural overlap for accurate rigid alignment between the

CT and MRI volumes. As the image intensity of bone is relatively high in CT, but relatively low in T1 MRI, a mutual information cost criterion is typically used to drive the alignment [46].

Unfortunately, the craniotomy and electrode implantations may cause significant deformation of the brain [24], which cannot be accounted for by rigid alignment of the CT and MRI volumes. Standard volume based non-rigid alignment methods are not a viable option due to the presence of the electrodes in the CT volume.

However, we know that the electrodes must lie on the cortical surface of the brain. As the electrodes are configured in grids and strips, they strictly lie on the hull that envelops the pial surface, rather than in sulcal folds. We refer to this hull as the *dural* surface and can approximate it by smoothing the extracted pial surface [38]. A simple way to correct for post-implantation brain deformation is to project each electrode location to the closest point on the dural surface. A more complex approach [10] also attempts to minimize the distortion of the electrode array configuration.

■ 2.2 Resting State fMRI

Clinical iEEG is the gold standard for identifying the onset and irritative zones, but only covers a limited area of the cortex and has a coarse spatial resolution. By contrast, fMRI has a spatial resolution of around 2mm and allows measurement, albeit indirect, of neural activity across the entire brain. This indirect measure is captured by the BOLD contrast, which measures local changes in blood oxygenation [36]. As brain regions are activated, energy requirements force oxygen metabolism to increase, thereby linking the BOLD contrast to neural activity [18].

Resting state fMRI probes brain activity in the absence of any particular stimulus. The measured BOLD time series capture fluctuations in functional activity that the subject happened to be undergoing at the time of acquisition, as illustrated in Figure 2.2. These measurements are of interest because they reveal the intrinsic functional connectivity between regions of the brain, unbiased by a specific task or stimu-

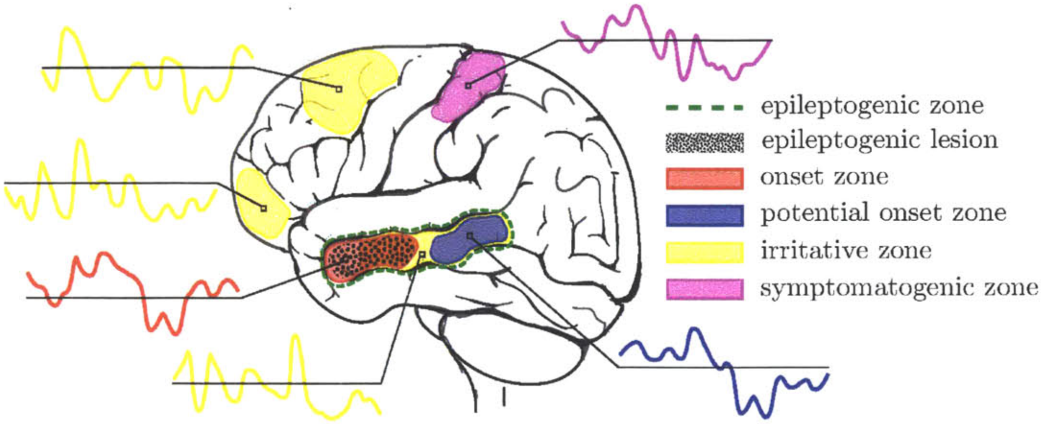

Figure 2.2: In resting state fMRI an interictal BOLD time series is observed at each voxel. The temporal resolution is too coarse to accurately detect abnormal spiking. Instead, correlations between the time series are used to infer the functional connectivity between voxels. Connections associated with epileptogenic regions are thought to exhibit a deviation from those of healthy controls.

lus [4]. A functional connection in this context does not necessarily correspond to an anatomical tract of white matter directly connecting two regions. Instead, it simply indicates that two regions exhibit functional synchrony through the correlation of their time series. However, multiple studies have found that functional and anatomical connections tend to co-occur [22, 40], suggesting that functional connections capture the underlying brain network. Epilepsy is thought to be strongly related to brain network disorganization [23, 25]. Resting state fMRI studies have found significant differences between the functional connectivity of epilepsy patients and those of a healthy population [30, 32, 37]. In some cases, these differences have been found to stem from regions in the vicinity of the epileptogenic zone [3, 41].

In order to find the functional connectivity differences between a patient and a group of healthy subjects, we must first pre-process the raw sequence of fMRI volumes so that confounds specific to the acquisition of each subject are removed. These confounds may be caused by differences in the shape and size of the subjects' brains, subject motion in the scanner and biological processes not directly linked to brain activity, such as the respiratory and cardiac cycles [29].

A detailed study on the effects of these confounds and robust ways to remove them can be found in [42]. For each acquisition, the first four volumes are discarded to allow for T1 equilibration effects. As fMRI scans typically employ an echo planar imaging protocol, each slice within a single volume is acquired at a different time. To correct for this, the slices are temporally aligned by shifting the time series accordingly. Furthermore, each volume in the sequence is rigidly aligned to the first volume to account for patient motion in the scanner [20]. Since functional correlations are most consistently produced by BOLD fluctuations in the frequency range of 0.01-0.08Hz [4], bandpass filtering is applied to the time series to remove any signal outside of this range. Finally, a few confounding signals are removed from the time series using linear regression. These consist of the mean time series across the whole brain, which likely accounts for the signal induced by breathing; that of the deep cerebral white matter, which carries no regionally specific functional signal; and that of the lateral ventricles, which carries no functional signal at all. Regressors are also formed from translation and rotation parameter values estimated from motion correction, as subject motion induces signal in the acquisition. Additionally, derivatives of all regressors are included to remove temporally shifted versions of the signals.

The pre-processing pipeline also includes spatial normalization. The anatomy of the brain is variable in size and shape, even across subjects in a healthy population. Spatial normalization translates, rotates and smoothly deforms each subject's anatomy into a common anatomical template, so that locations in the brain can be directly compared across subjects. This template may be a specific subject, an average brain [13] or a probabilistic atlas of healthy brains [31]. There are inevitable normalization errors because the brains of different subjects cannot be smoothly and exactly transformed to one another. This misalignment can be ameliorated by applying spatial smoothing, so that each voxel contains a weighted average of its neighbors' time series. While this step effectively reduces the spatial resolution of the data, it typically increases the statistical

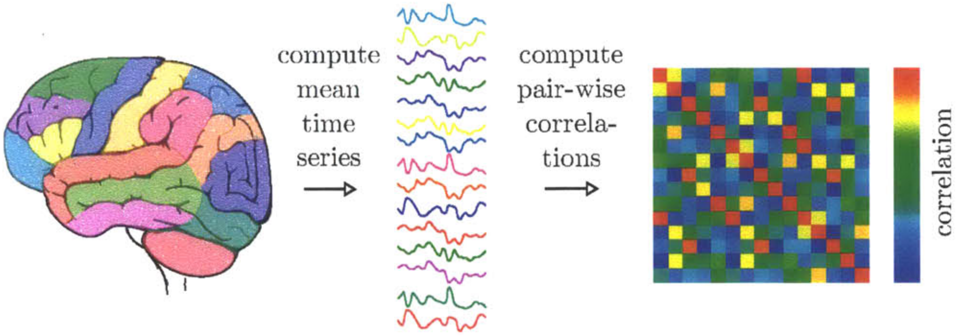

Figure 2.3: The brain can be parcelated into a relatively small number of regions of anatomical and functional interest. Correlations between their associated mean time series are used as manageable observations of functional connectivity in the brain.

power of comparison across a group of subjects.

The functional connectivity between two voxels can be measured by computing the Pearson correlation coefficient between their fMRI time series. Assuming the brain covers N voxels, the functional connectivity of the entire brain can be characterized by the $N \times N$ symmetric positive semi-definite matrix, where a large positive element suggests the presence of a functional connection. Large negative elements may indicate the presence of an antagonistic functional connection [34], but should be treated with caution as the neurophysiological basis of such negative correlations is not well understood [47].

Our goal is to automatically identify salient regions and networks in the brain from this matrix. However, the number of gray matter voxels in an isotropic 2mm brain is roughly $N = 2 \times 10^5$, which makes computation with this matrix impractical. One way to overcome this computational problem is to partition the whole brain into a smaller number of relatively large regions, and associate each with its mean time course. Correlations between these mean time courses can then be calculated to produce a smaller matrix, as shown in Figure 2.3, enabling efficient analysis. For example, the FreeSurfer software package [14] automatically partitions the brain into $N = 150$ cortical parcels of anatomical and functional interest [9]. Alternatively, smaller regions can be defined by uniformally subdividing the cortical surface, producing as many as $N = 1175$

regions to analyze [49].

2.2.1 Statistics of Functional Correlations

In resting state fMRI, each correlation is an observation of functional connectivity between two regions. We aim to discover whether such a correlation observed in an epilepsy patient deviates from the corresponding correlations observed in a healthy population of subjects, as illustrated in Figure 2.4. Alternatively, we can declare a null hypothesis that this correlation does not deviate from the healthy population. The associated Normal null distribution over this correlation is completely defined by its mean and variance, which can be estimated from a sample of H healthy subjects as

$$\mu_{nm} = \frac{1}{H}\sum_{h=1}^{H} b_{nmh}, \quad \sigma_{nm}^2 = \frac{1}{H-1}\sum_{h=1}^{H}(b_{nmh} - \mu_{nm})^2,$$

where b_{nmh} is the correlation between regions n and m of healthy subject h.

The correlation b_{nmu} in an epilepsy patient u can then associated with a z-score that takes on higher magnitudes as it deviates from the mean of the null distribution:

$$z_{nmu} = \frac{b_{nmu} - \mu_{nm}}{\sigma_{nm}}.$$

We aim to detect both abnormally low and high correlations, and therefore choose the absolute value of z_{nmu} as our connection statistic. This approach has been applied to studies of epilepsy [3, 30] and other neurological disorders [7], and typically highlights many connections associated with correlations that significantly deviate from normality.

By setting a threshold α, the abnormal connectivity associated with a region can be summarized by calculating the proportion of significantly different correlations:

$$p_{nu}(\alpha) = \frac{1}{N-1}\sum_{m \neq n} \mathbb{1}(|z_{nmu}^b| > \alpha). \tag{2.1}$$

where $\mathbb{1}(\cdot)$ is equal to one when its argument is true and zero otherwise.

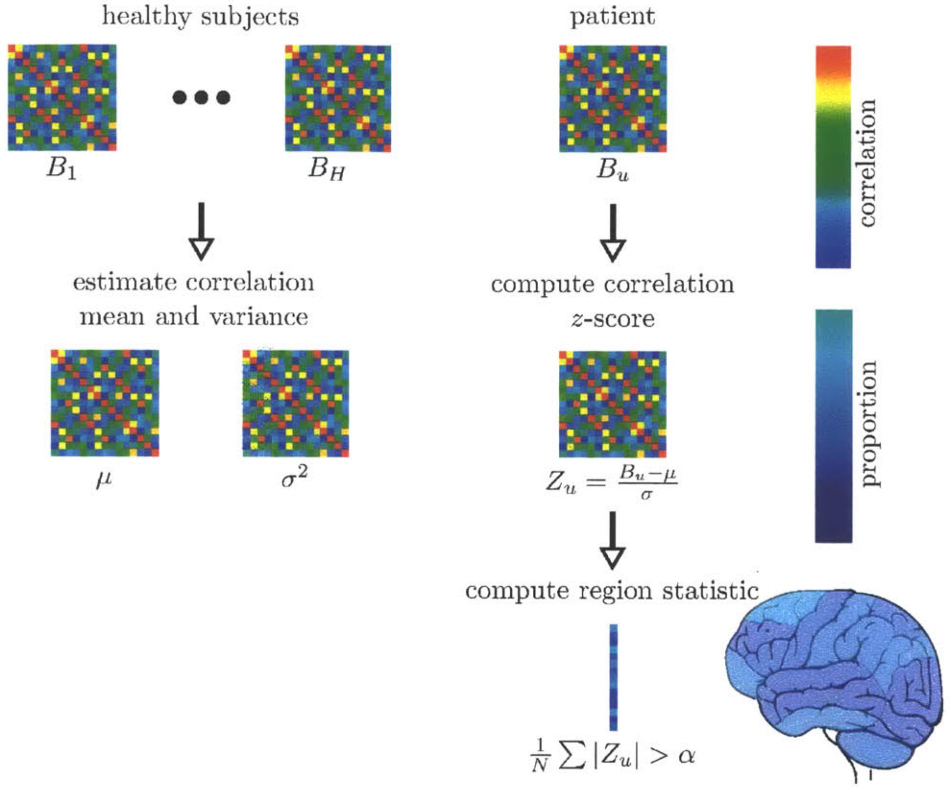

Figure 2.4: One approach to identifying anomalous regions is to compute a z-score for each element of the correlation matrix B_u of patient u. This represents the deviation from corresponding correlations of a sample of H healthy subjects. A statistic is computed for a region by counting the proportion of correlations over a threshold α.

■ 2.2.2 Statistics of Graph Theoretical Features

A general limitation of correlation statistics is that it is unclear how to summarize the resulting statistics for each region. Instead, we can treat the functional brain network as a graph, where nodes correspond to brain regions and the presence of edges is determined by whether the correlation between two regions is above a pre-selected threshold [6, 17]. Graph theory can be used to compute features of the nodes, which are scalar values that characterize some notion of importance or centrality in the graph.

For example, the degree centrality d_{nu} associated with region n of patient u is the proportion of other regions that region n is connected to:

$$d_{nu}(\beta) = \frac{1}{N-1} \sum_{m \neq n} \mathbb{1}(b_{nmu} > \beta),$$

where β is a threshold that determines whether a correlation corresponds to a functional connection. We aim to detect those regions that exhibit significantly different degrees in patient u compared to those in a healthy population, as illustrated in Figure 2.5. This property is captured by the absolute value of the degree z-score:

$$z_{nu}(\beta) = \left| \frac{d_{nu}(\beta) - \mu_n(\beta)}{\sigma_n(\beta)} \right|, \tag{2.2}$$

where $\mu_n(\beta)$ and $\sigma_n(\beta)$ denote the mean and variance of the healthy degree and are estimated as

$$\mu_n(\beta) = \frac{1}{H} \sum_{h=1}^{H} d_{nh}(\beta), \quad \sigma_n^2(\beta) = \frac{1}{H-1} \sum_{h=1}^{H} (d_{nh}(\beta) - \mu_n(\beta))^2.$$

In functional connectivity analysis, the degree centrality can be restricted to local or distal regions of the brain to produce different characterizations [39]. These showed promise for localization of epileptogenic regions in a recent study of pre-surgical epilepsy patients [41]. Another feature is betweenness centrality, which is the number of shortest paths between all pairs of nodes a particular node lies on. This measure represents how important a particular node is for efficient communication in a graph. Nodes with high betweenness centrality were found to be reproducible across healthy subjects, and were found to correlate with regions of the brain that exhibited high amyloid-β deposits in a study of Alzheimer's patients [5]. When the number of nodes in the graph is large, computing the betweenness centrality becomes impractical, but can be approximated in brain networks with the more easily computable eigenvector centrality [27].

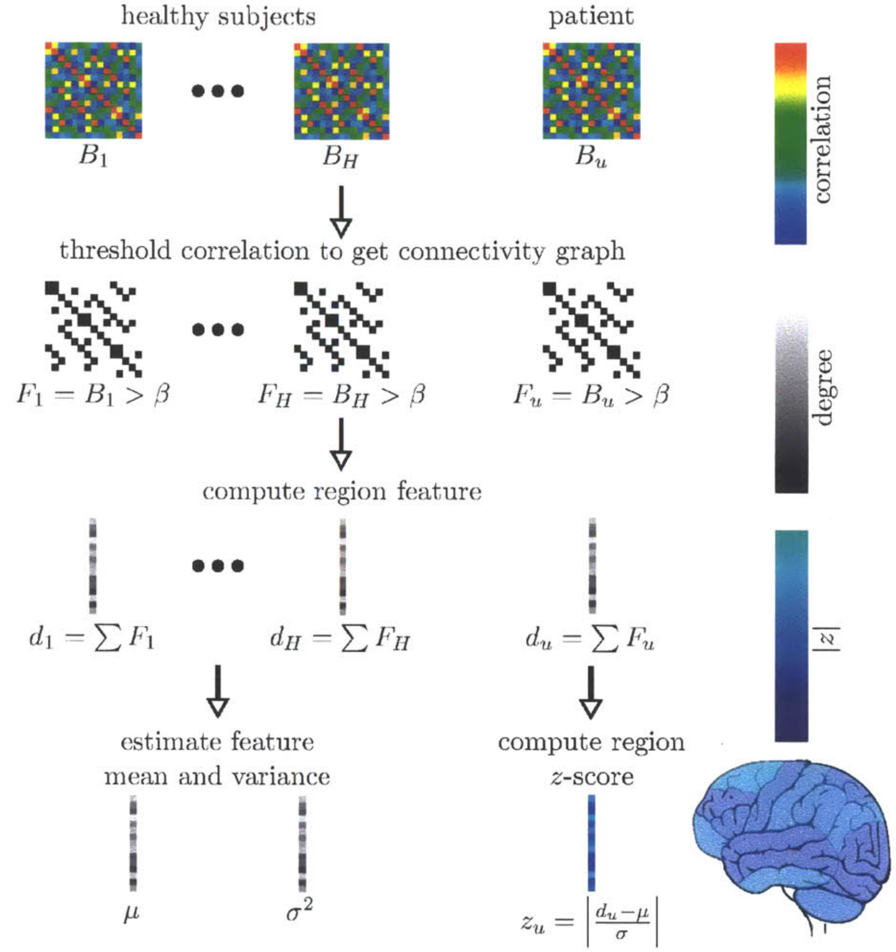

Figure 2.5: We can construct a binary graph of functional connectivity by thresholding the correlation matrix B_u of patient u with a threshold β. In this graph, nodes correspond to regions. Graph theory can be used to compute features of these nodes in the patient, such as the degree centrality, and compute the deviation of corresponding features in H healthy subjects.

■ 2.2.3 Generative Models of Functional Connectivity

Using discriminative statistics is by far the most common approach in connectivity studies of neurological disorders. Rather than searching for a sensitive statistic of the observed correlations, we can instead attempt to build a probabilistic generative model of them. This approach has found widespread use and success in computer vision due

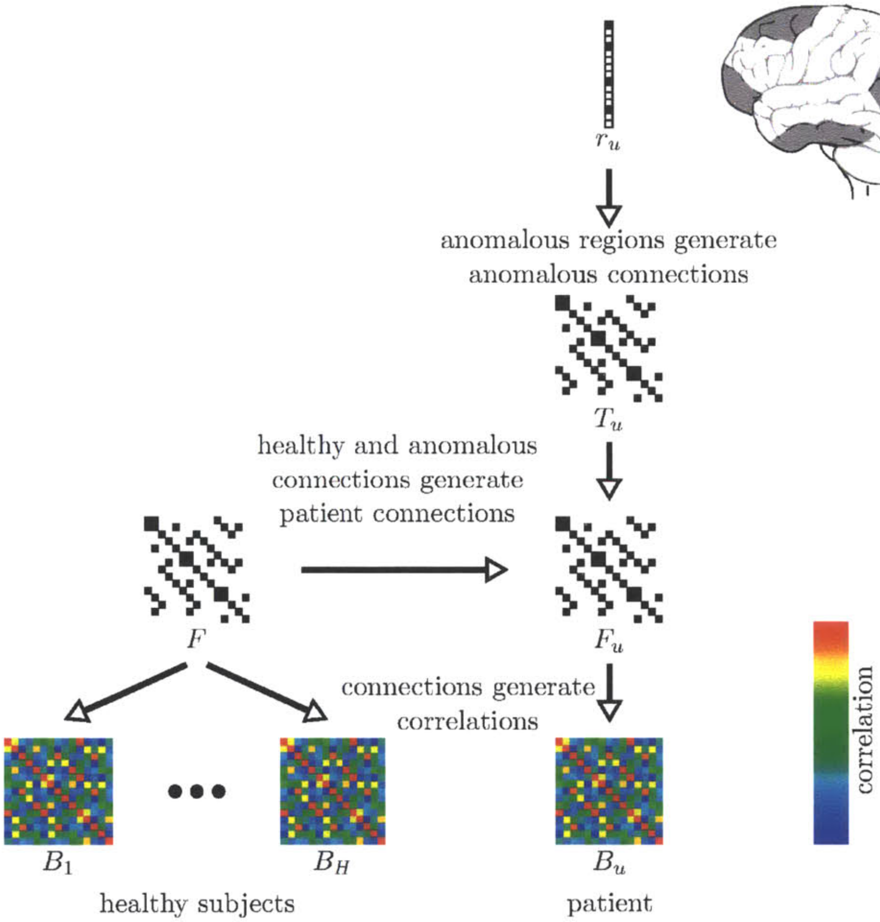

Figure 2.6: The generative model of anomalous regions proposed by [44]. Here, latent random variables represent anomalous regions r_u and connections T_u. The model captures the probabilistic relationship between these latent variables and the observed correlations. By designing a suitable algorithm, the latent anomalous regions can be recovered from the observed correlations without the need for parameter selection.

to its robustness and interpretability. However, generative models have seen far less use in resting state fMRI connectivity analysis.

A notable exception is a random effects model of symmetric positive definite matrices, which is used to define a Normal distribution of correlation matrices in healthy subjects [43]. This method uses a tangent space parameterization of symmetric positive definite matrices, which allows the parameterized elements to be treated as independent, while maintaining the dependency structure in the underlying correlation matrix. Non-parametric sampling of the healthy subject distribution is used to define a null distribution of healthy deviations away from the mean correlation matrix, which is then used to perform a statistical test on the deviations observed in stroke patients. The method is similar in spirit to the approach of performing statistical tests on correlations and also requires an adhoc summary of the correlation statistics for each region.

Another approach proposes a generative model of functional and anatomical brain connectivity in a healthy population and of the deviations from this in a schizophrenia patient population [45]. The model uses latent discrete random variables to represent different states of functional connectivity. Given a connectivity state, a correlation is generated from a corresponding Normal distribution that is shared across subjects and patients. In contrast to standard graph theoretic approaches, continuous correlations are softly assigned to connections without the need to choose a threshold. This model can also be extended to describe how anomalous regions give rise to anomalous connections shared across an unhealthy population [44]. The basic idea is illustrated in Figure 2.6.

Given the anomalous state of two regions, an anomalous connection is drawn from a Bernoulli distribution defined such that anomalous regions are associated with a large number of anomalous connections. In contrast to correlation statistic approaches, this number is automatically determined from the data. The goal is to invert the generative process and estimate the posterior probability of each region being anomalous, given the observed correlations of healthy subjects and patients. The current approach uses a combination of Gibbs sampling [16] and a variational approximation [21] to perform this estimation. In the application of the model, the number of regions $N = 70$ is small. However, for the large number of regions required in our application, Gibbs sampling is computationally impractical. In the next chapter, we adapt this method for anomaly detection in epilepsy patients.

Chapter 3

Methods

We start this chapter by describing our generative model of anomalous regions in Section 3.1. This construction builds on the generative model described in [44]. In contrast to the previously proposed method, the anomalous regions in our model are not shared across patients, which is appropriate for application to epilepsy patients. However, we do allow for parameters associated with anomalies to be shared across subjects. In other words, the characteristics of change are common across patients, but can be located in different regions. We present two mean field variational algorithms for posterior estimation that scale well as the number of regions grows. The first algorithm, described in Section 3.2, approximates posteriors over all purely latent variables. The second, described in Section 3.3, marginalizes some of the latent variables to improve the accuracy of the estimation.

■ 3.1 Generative Model of Anomalous Regions

Let R_{nu} be a Bernoulli random variable indicating that region n of patient u is anomalous. R_{nu} is drawn from the distribution

$$p(r_{nu}; \pi) = \pi^{r_{nu}}(1-\pi)^{1-r_{nu}}, \tag{3.1}$$

where $\pi \in (0,1)$ is the parameter of a Bernoulli distribution.

Let T_{nmu} be a Bernoulli random variable indicating that the connection between regions n and m of patient u is anomalous. T_{nmu} is dependent on the anomalous state of the regions at either end of the connection, and is drawn from the distribution

$$p(t_{nmu}|r_{nu}, r_{mu}; \eta) = \begin{cases} \delta(t_{nm}) & \text{if } r_{nu} = r_{mu} = 0, \\ \delta(1 - t_{nm}) & \text{if } r_{nu} = r_{mu} = 1, \\ \eta^{t_{nmu}}(1-\eta)^{1-t_{nmu}} & \text{if } r_{nu} \neq r_{mu}, \end{cases} \tag{3.2}$$

where δ is the Dirac delta function and $\eta \in (0, 1)$ is the parameter of a Bernoulli distribution. T_{nmu} is deterministic if the anomalous state of regions n and m in patient u is the same, and is a Bernoulli random variable with parameter η if they are different. This distribution encourages anomalous networks containing cliques of anomalous nodes, where larger values of η allow more edges outside of cliques to be affected.

Let F_{nm} be a multinomial random variable indicating the state of healthy connectivity between regions n and m. We use three states of connectivity: $f_{nm}[-1] = 1$ denotes a negative connection; $f_{nm}[0] = 1$ denotes no connection; and $f_{nm}[1] = 1$ denotes a positive connection. Exactly one component of f_{nm} must be equal to one. F_{nm} is drawn from the distribution

$$p(f_{nm}; \gamma) = \prod_{k=-1}^{1} \gamma_k^{f_{nm}[k]}, \tag{3.3}$$

where $\gamma = (\gamma_{-1}, \gamma_0, \gamma_1)$ is the parameter vector of a Multinomial distribution such that $\gamma_k \in (0, 1)$ and $\sum_{k=-1}^{1} \gamma_k = 1$.

Let $\tilde{F}_{nmu}$ be a multinomial random variable indicating the state of connectivity between regions n and m of patient u. $\tilde{F}_{nmu}$ is dependent on T_{nmu}, the anomalous state of the connection between regions n and m of patient u, and on F_{nm}, the healthy connectivity state between regions n and m. $\tilde{F}_{nmu}$ is drawn from the distribution

$$p(\tilde{f}_{nmu}|f_{nm}, t_{nmu}; \epsilon) = \begin{cases} (1-\epsilon)^{f_{nm}^{\top}\tilde{f}_{nmu}} \left(\frac{\epsilon}{2}\right)^{1-f_{nm}^{\top}\tilde{f}_{nmu}} & \text{if } t_{nmu} = 0, \\ \epsilon^{f_{nm}^{\top}\tilde{f}_{nmu}} \left(\frac{1-\epsilon}{2}\right)^{1-f_{nm}^{\top}\tilde{f}_{nmu}} & \text{if } t_{nmu} = 1, \end{cases} \tag{3.4}$$

where $\epsilon \in (0, 1)$ is the parameter of a Bernoulli distribution. If the connection between regions n and m of patient u is anomalous, the connectivity state is perturbed from the healthy template with high probability $1 - \epsilon$. Conversely, if the connection is normal, the connectivity state is perturbed with small probability ϵ.

Let B_{nmh} be the random correlation coefficient between the time series of regions n and m of healthy subject h. B_{nmh} is dependent on the healthy connectivity state, and is drawn from a mixture of Normal distributions:

$$p(b_{nmh}|f_{nm}; \mu, \sigma) = \prod_{k=-1}^{1} \mathcal{N}(b_{nmh}; \mu_k, \sigma_k^2)^{f_{nm}[k]}, \tag{3.5}$$

where $\mu = (\mu_{-1}, \mu_0, \mu_1)$, $\sigma = (\sigma_{-1}, \sigma_0, \sigma_1)$ and $\mathcal{N}(\cdot; \mu_k, \sigma_k^2)$ is a Normal distribution with mean μ_k and variance σ_k^2.

Similarly, let $\tilde{B}_{nmu}$ denote the random correlation coefficient between the time series of regions n and m of patient u. $\tilde{B}_{nmu}$ is dependent on the connectivity state of the patient, and is drawn from the same mixture of Normal distributions as the healthy correlations:

$$p(\tilde{b}_{nmu}|\tilde{f}_{nmu}; \mu, \sigma) = \prod_{k=-1}^{1} \mathcal{N}(\tilde{b}_{nmu}; \mu_k, \sigma_k^2)^{\tilde{f}_{nmu}[k]}. \tag{3.6}$$

We assume independence between all healthy subjects and patients, independence between healthy connections and independence between regions, and thus obtain the full joint distribution:

$$\begin{aligned} p(f, b, r, t, \tilde{f}, \tilde{b}; \theta) &= p(f; \gamma)p(b|f; \mu, \sigma)p(r; \pi)p(t|r; \eta)p(\tilde{f}|f, t; \epsilon)p(\tilde{b}|\tilde{f}; \mu, \sigma) \\ &= \left(\prod_{n=1}^{N} \prod_{m>n} p(f_{nm}; \gamma) \prod_{h=1}^{H} p(b_{nmh}|f_{nm}; \mu, \sigma) \right) \qquad (3.7) \\ &\quad \left(\prod_{n=1}^{N} \prod_{u=1}^{U} p(r_{nu}; \pi) \prod_{m>n} p(t_{nmu}|r_{nu}, r_{mu}; \eta) p(\tilde{f}_{nmu}|f_{nm}, t_{nmu}; \epsilon) p(\tilde{b}_{nmu}|\tilde{f}_{nmu}; \mu, \sigma) \right), \end{aligned}$$

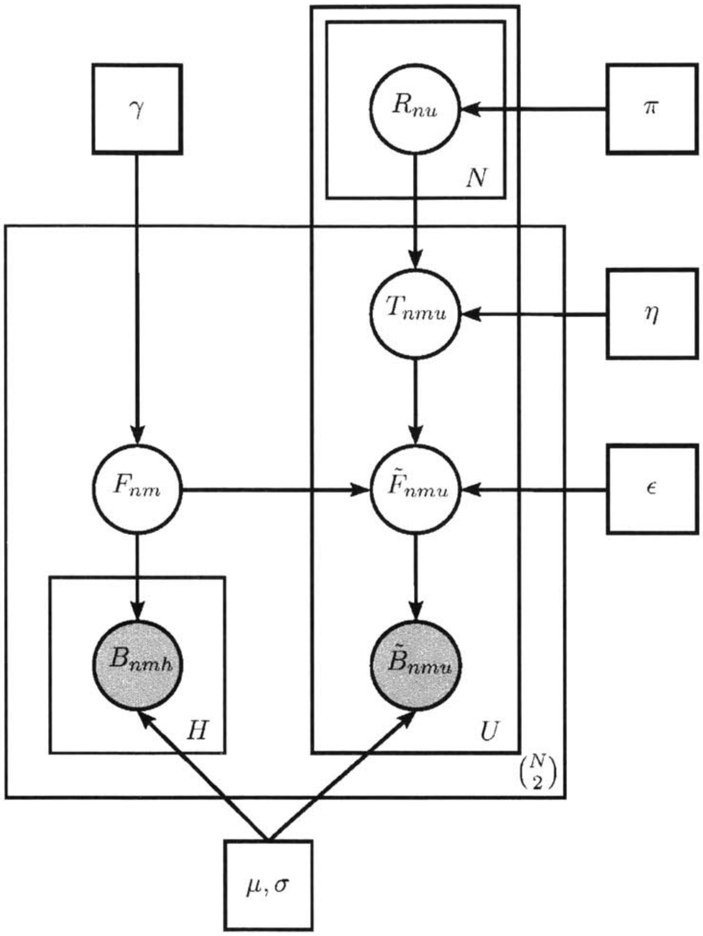

Figure 3.1: The directed graph that represents the interaction between the random variables and parameters in the probabilistic model we use for anomalous region detection. Circles correspond to random variables, and squares correspond to parameters. Shaded circles indicate observed values.

where $\theta = (\pi, \gamma, \eta, \epsilon, \mu, \sigma)$.

We summarize this generative model with the graphical model shown in Figure 3.1. The interaction between the region label R_{nu} and the edge T_{nmu} is detailed in the graphical model shown in Figure 3.2.

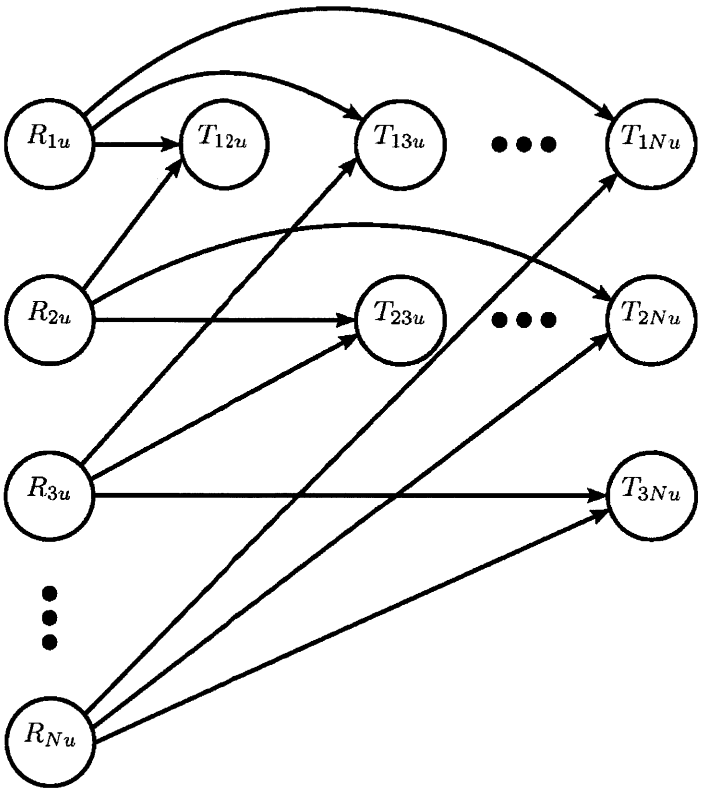

Figure 3.2: The directed graph that represents the interaction between anomalous regions $\{R_{nu} : n = 1 \dots N\}$ and connectivity states $\{T_{nmu} : n = 1 \dots N, m > n\}$ for patient u. Parameters are omitted for clarity.

3.2 Mean Field T-Algorithm

Our goal is to compute the posterior probability $p(r_{nu}|\tilde{b}, b; \theta)$ for all regions $n \in \{1, \dots, N\}$ in all patients $u \in \{1, \dots, U\}$. This requires marginalizing out all latent random variables in the model to compute the partition function $p(\tilde{b}, b; \theta)$.

We can easily sum over T to obtain

$$p(\tilde{f}_{nmu}|f_{nm}, r_{nu}, r_{mu}; \eta, \epsilon) = \begin{cases} (1-\epsilon)^{f_{nm}^{\mathsf{T}}\tilde{f}_{nmu}} \left(\frac{\epsilon}{2}\right)^{1-f_{nm}^{\mathsf{T}}\tilde{f}_{nmu}} & \text{if } r_{nu} = r_{mu} = 0, \\ \epsilon^{f_{nm}^{\mathsf{T}}\tilde{f}_{nmu}} \left(\frac{1-\epsilon}{2}\right)^{1-f_{nm}^{\mathsf{T}}\tilde{f}_{nmu}} & \text{if } r_{nu} = r_{mu} = 1, \\ \epsilon_1^{f_{nm}^{\mathsf{T}}\tilde{f}_{nmu}} \left(\frac{1-\epsilon_1}{2}\right)^{1-f_{nm}^{\mathsf{T}}\tilde{f}_{nmu}} & \text{if } r_{nu} \neq r_{mu}, \end{cases} \quad (3.8)$$

where $\epsilon_1 = \eta\epsilon + (1-\eta)(1-\epsilon)$.

Now, the full joint distribution is of the form

$$\begin{aligned} p(f,b,r,\tilde{f},\tilde{b};\theta) &= p(f;\gamma)p(b|f;\mu,\sigma)p(r;\pi)p(\tilde{f}|f,r;\eta,\epsilon)p(\tilde{b}|\tilde{f};\mu,\sigma) \\ &= \left(\prod_{n=1}^{N}\prod_{m>n} p(f_{nm};\gamma)\prod_{h=1}^{H} p(b_{nmh}|f_{nm};\mu,\sigma)\right) \qquad (3.9) \\ &\left(\prod_{n=1}^{N}\prod_{u=1}^{U} p(r_{nu};\pi)\prod_{m>n} p(\tilde{f}_{nmu}|f_{nm},r_{nu},r_{mu};\epsilon)p(\tilde{b}_{nmu}|\tilde{f}_{nmu};\mu,\sigma)\right). \end{aligned}$$

Due to the dependence of $\tilde{F}_{nmu}$ on R_{nu} and R_{mu}, marginalizing out R requires a summation over all possible binary vectors of length N representing possible values of the anomalous region vector r_u for each patient. There are $\mathcal{O}(2^N)$ such vectors, making this sum computationally intractible. We choose to approximate the sum by estimating a mean field variational factorization of the posterior probability distribution, which gives us a lower bound on the partition function [21].

Computing the posterior $p(r_{nu}|\tilde{b},b;\theta)$ also requires an estimate of the parameter vector θ. Here, we choose the variational EM algorithm for this estimation [2], which is described in Algorithm 1. This algorithm also requires the posteriors $p(f_{nm}|\tilde{b},b;\theta)$ and $p(\tilde{f}_{nmu}|\tilde{b},b;\theta)$, which the mean field factorization provides. The factorization takes the form

$$\begin{aligned} &p(f,r,\tilde{f}|\tilde{b},b;\theta) \approx q(f,r,\tilde{f}) = q_F(f)q_R(r)q_{\tilde{F}}(\tilde{f}) \qquad (3.10) \\ &= \left(\prod_{n=1}^{N}\prod_{m>n}(q_{F_{nm}[k]})^{f_{nm}[k]}\right)\left(\prod_{n=1}^{N}\prod_{u=1}^{U}\prod_{l=0}^{1}(q_{R_{nu}[l]})^{r_{nu}[l]}\prod_{m>n}\prod_{k=-1}^{1}(q_{\tilde{F}_{nmu}[k]})^{\tilde{f}_{nmu}[k]}\right). \end{aligned}$$

We update the factors $q = \{q_{F_{nm}}, q_{R_{nu}}, q_{\tilde{F}_{nmu}}\}$ by minimizing the variational free

```
Data: Correlation coefficients of healthy subjects $b$ and patients $\tilde{b}$.
Result: Posterior probability estimates $q$ and parameter estimates $\theta$.
initialize $q^{(0)}$ as uniform distributions
initialize $\theta^{(0)}$ randomly from uniform distributions
$s \leftarrow 0$
$\mathcal{E}^{(0)} \leftarrow \infty$
$\mathcal{E}^{(1)} \leftarrow \mathcal{E}(q^{(0)}, \theta^{(0)}; b, \tilde{b})$
while $(\mathcal{E}^{(s)} - \mathcal{E}^{(s+1)})/|\mathcal{E}^{(s)}| > 10^{-4}$) do
    $s \leftarrow s + 1$
    $t \leftarrow 0$
    $\mathcal{E}^{(s,0)} \leftarrow \infty$
    $\mathcal{E}^{(s,1)} \leftarrow \mathcal{E}^{(s)}$
    $q^{(s)} \leftarrow q^{(s-1)}$
    while $(\mathcal{E}^{(s,t)} - \mathcal{E}^{(s,t+1)})/|\mathcal{E}^{(s,t)}| > 10^{-4}$) do
        $t \leftarrow t + 1$
        $q^{(s)} \leftarrow \text{update}(q^{(s)}, \theta^{(s-1)}, b, \tilde{b})$
        $\mathcal{E}^{(s,t+1)} \leftarrow \mathcal{E}(q^{(t+1)}, \theta^{(s-1)}; b, \tilde{b})$
    end
    $\theta^{(s)} \leftarrow \text{update}(q^{(s)}, \theta^{(s-1)}, b, \tilde{b})$
    $\mathcal{E}^{(s+1)} \leftarrow \mathcal{E}(q^{(s)}, \theta^{(s)}; b, \tilde{b})$
end
return $q^{(s)}, \theta^{(s)}$
```

Algorithm 1: The high level structure of the variational EM inference algorithm. For each iterative update of the parameters θ in the outer loop, we must iteratively update the posterior probability estimates q in the inner loop.

energy $\mathcal{E}(q, \theta; b, \tilde{b})$ for fixed observations $(b, \tilde{b})$ and parameters θ:

$$\begin{aligned}
\mathcal{E}(q, \theta; \tilde{b}, b) &= -\mathbb{E}_q[\log p(F, b, R, \tilde{F}, \tilde{b}; \theta)] + \mathbb{E}_q[\log q(R, \tilde{F}, F)] \\
&= -\mathbb{E}_{q_F}[\log p(F; \gamma)] - \mathbb{E}_{q_F}[\log p(b|F; \mu, \sigma)] - \mathbb{E}_{q_R}[\log p(R; \pi)] \\
&\quad - \mathbb{E}_{q_F q_R q_{\tilde{F}}}[\log p(\tilde{F}|F, R; \epsilon, \eta)] - \mathbb{E}_{q_{\tilde{F}}}[\log p(\tilde{b}|\tilde{F}; \mu, \sigma)] \\
&\quad + \mathbb{E}_{q_R}[\log q_R(R)] + \mathbb{E}_{q_F}[\log q_F(F)] + \mathbb{E}_{q_{\tilde{F}}}[\log q_{\tilde{F}}(\tilde{F})]. \qquad (3.11)
\end{aligned}$$

The factors are initialized as uniform distributions and are iteratively updated to minimize the energy according to the following update equations.

Assuming the number of healthy subjects is substantially larger than the number of patients, the update for the healthy connectivity factor is mostly driven by the Normal likelihoods of the correlations in the healthy subjects:

$$\begin{aligned}\log q_{F_{nm}[k]} = {}& \log \gamma_k + \sum_{h=1}^{H} \log \mathcal{N}(b_{nmh}; \mu_k, \sigma_k^2) + \\ & \sum_{u=1}^{U} q_{\tilde{F}_{nmu}[k]} \Bigg(q_{R_{nu}[0]} q_{R_{mu}[0]} \log \frac{2(1-\epsilon)}{\epsilon} + q_{R_{nu}[1]} q_{R_{mu}[1]} \log \frac{2\epsilon}{1-\epsilon} + \\ & \big(q_{R_{nu}[1]} q_{R_{mu}[0]} + q_{R_{nu}[0]} q_{R_{mu}[1]} \big) \log \frac{2\epsilon_1}{1-\epsilon_1} \Bigg) + \text{const.}\end{aligned} \tag{3.12}$$

This estimate is only influenced by a patient when their corresponding connectivity posterior probability is high. In this case, a high probability of healthy regions at either end of the connection increases the estimate and a high probability of unhealthy regions decreases it.

Updates for the connectivity posterior of patient u are similar:

$$\begin{aligned}\log q_{\tilde{F}_{nmu}[k]} = {}& \log \mathcal{N}(\tilde{b}_{nmu}; \mu_k, \sigma_k^2) + \\ & q_{F_{nm}[k]} \Bigg(q_{R_{nu}[0]} q_{R_{mu}[0]} \log \frac{2(1-\epsilon)}{\epsilon} + q_{R_{nu}[1]} q_{R_{mu}[1]} \log \frac{2\epsilon}{1-\epsilon} + \\ & \big(q_{R_{nu}[1]} q_{R_{mu}[0]} + q_{R_{nu}[0]} q_{R_{mu}[1]} \big) \log \frac{2\epsilon_1}{1-\epsilon_1} \Bigg) + \text{const.}\end{aligned} \tag{3.13}$$

This update is affected by the Normal likelihood of the unhealthy correlation, and is also affected by the subject's region label posterior probability as for $q_{F_{nm}[k]}$ when the corresponding healthy connectivity posterior probability is high.

Updates for $q_{R_{nu}[1]}$ take on relatively high values compared to $q_{R_{nu}[0]}$ when $q_{F_{nm}}^{\top} q_{\tilde{F}_{nmu}}$

has a low value for most connections from region n:

$$\begin{aligned}\log q_{R_{nu}[0]} &= \log(1-\pi) + \sum_{m \neq n} q_{R_{mu}[0]} \left(q_{F_{nm}}^{\top} q_{\tilde{F}_{nmu}} \log(1-\epsilon) + (1 - q_{F_{nm}}^{\top} q_{\tilde{F}_{nmu}}) \log \frac{\epsilon}{2} \right) + \\ &\quad q_{R_{mu}[1]} \left(q_{F_{nm}}^{\top} q_{\tilde{F}_{nmu}} \log \epsilon_1 + (1 - q_{F_{nm}}^{\top} q_{\tilde{F}_{nmu}}) \log \frac{1-\epsilon_1}{2} \right) + \text{const.}, \qquad (3.14)\end{aligned}$$

$$\begin{aligned}\log q_{R_{nu}[1]} &= \log \pi + \sum_{m \neq n} q_{R_{mu}[1]} \left(q_{F_{nm}}^{\top} q_{\tilde{F}_{nmu}} \log \epsilon + (1 - q_{F_{nm}}^{\top} q_{\tilde{F}_{nmu}}) \log \frac{1-\epsilon}{2} \right) + \\ &\quad q_{R_{mu}[0]} \left(q_{F_{nm}}^{\top} q_{\tilde{F}_{nmu}} \log \epsilon_1 + (1 - q_{F_{nm}}^{\top} q_{\tilde{F}_{nmu}}) \log \frac{1-\epsilon_1}{2} \right) + \text{const.} \qquad (3.15)\end{aligned}$$

A low value of $q_{F_{nm}}^{\top} q_{\tilde{F}_{nmu}}$ occurs when the connectivity states associated with region n of patient u differs greatly from that of the healthy population. This is accentuated when the regions associated with differing connectivity also have a high chance of being anomalous.

We iteratively update these posterior probability estimates in the order provided above until there is little decrease in the variational energy. Specifically, we continue iterating while $(\mathcal{E}^{(s)} - \mathcal{E}^{(s+1)})/|\mathcal{E}^{(t)}| > 10^{-4}$.

After convergence, the posterior probability estimates are fixed and we estimate the parameters that minimize the variational energy. The parameters are randomly initialized from uniform distributions, such that $\pi \in [0.01, 0.1]$, $\gamma_{-1}, \gamma_1 \in [0.1, 0.3], \gamma_0 = 1 - \gamma_{-1} - \gamma 1$, $\mu_{-1} \in [-0.5, -0.2]$, $\mu_1 \in [0.2, 0.5]$, $\sigma_k^2 \in [0.01, 0.1] \forall k$, $\eta \in [0.1, 0.5]$ and $\epsilon \in [0.01, 0.05]$. We initialize $\mu_0 = 0$ and keep this value fixed throughout the algorithm. The other parameters are updated as described below.

The Bernoulli parameter π of an anomalous region is determined by the corresponding average posterior probability:

$$\pi = \frac{1}{UN} \sum_{u=1}^{U} \sum_{n=1}^{N} q_{R_{nu}[1]}. \qquad (3.16)$$

The multinomial parameter vector of the prior over connectivity states is determined in a similar way:

$$\gamma_k = \sum_{n=1}^{N} \sum_{m>n} q_{F_{nm}[k]} + \text{const.} \tag{3.17}$$

The Normal parameters are updated in a Maximum-Likelihood fashion and pooled over healthy subjects and patients:

$$\mu_k = \frac{\sum_{n=1}^{N} \sum_{m>n} \left(\sum_{h=1}^{H} q_{F_{nm}[k]} b_{nmh} + \sum_{u=1}^{U} q_{\tilde{F}_{nm}[k]} \tilde{b}_{nmu} \right)}{\sum_{n=1}^{N} \sum_{m>n} \left(\sum_{h=1}^{H} q_{F_{nm}[k]} + \sum_{u=1}^{U} q_{\tilde{F}_{nm}[k]} \right)}, \tag{3.18}$$

$$\sigma_k^2 = \frac{\sum_{n=1}^{N} \sum_{m>n} \left(\sum_{h=1}^{H} q_{F_{nm}[k]} (b_{nmh} - \mu_k)^2 + \sum_{u=1}^{U} q_{\tilde{F}_{nm}[k]} (\tilde{b}_{nmu} - \mu_k)^2 \right)}{\sum_{n=1}^{N} \sum_{m>n} \left(\sum_{h=1}^{H} q_{F_{nm}[k]} + \sum_{u=1}^{U} q_{\tilde{F}_{nm}[k]} \right)}, \tag{3.19}$$

where each correlation is weighted by the appropriate posterior probability.

The Bernoulli parameters η and ϵ are inherently coupled. That is, the derivatives of the variational energy with respect to η and ϵ contain both parameters. Therefore, we iteratively estimate optimal parameter values using a Gauss-Newton type descent method. Specifically, we use the trust-region reflective algorithm implemented in MATLAB and iteratively update ϵ and η until the relative decrease in the variational energy is less than 10^{-4}. The asymptotic algorithmic complexity of computing the update rules above and the required derivatives is $\mathcal{O}(N^2)$.

■ 3.3 Mean Field $T\tilde{F}$-Algorithm

In the existing model, we approximate the posterior distribution over R, F and $\tilde{F}$ because summing over all possible values of r is intractible. However, summing over all possible values of f and $\tilde{f}$ is much easier. Here, we marginalize out $\tilde{F}$, as the posterior over this variable must be approximated for each patient and is therefore likely to be more poorly estimated than F, which is estimated by pooling over observations from many subjects. Marginalizing out $\tilde{F}$ provides

$$\begin{aligned} p(\tilde{b}_{nmu}|f_{nm}, r_{nu}, r_{mu}; \theta) &= \sum_{\tilde{f}_{nmu}} p(\tilde{f}_{nmu}|f_{nm}, r_{nu}, r_{mu}; \eta, \epsilon) p(\tilde{b}_{nmu}|\tilde{f}_{nmu}; \mu, \sigma) \\ &= \prod_{k=-1}^{1} \mathcal{M}_k^0(\tilde{b}_{nmu}; \theta)^{f_{nm}[k] r_{nu} r_{mu}} \mathcal{M}_k^1(\tilde{b}_{nmu}; \theta)^{f_{nm}[k](1-r_{nu})(1-r_{mu})} \\ &\quad \mathcal{M}_k^{\neq}(\tilde{b}_{nmu}; \theta)^{f_{nm}[k](r_{nu}(1-r_{mu})+(1-r_{nu})r_{mu})}, \end{aligned} \tag{3.20}$$

where

$$\mathcal{M}_k^0(\tilde{b}_{nmu}; \theta) = (1-\epsilon)\mathcal{N}(\tilde{b}_{nmu}; \mu_k, \sigma_k^2) + \frac{\epsilon}{2} \sum_{l \neq k} \mathcal{N}(\tilde{b}_{nmu}; \mu_l, \sigma_l^2), \tag{3.21}$$

$$\mathcal{M}_k^1(\tilde{b}_{nmu}; \theta) = \epsilon \mathcal{N}(\tilde{b}_{nmu}; \mu_k, \sigma_k^2) + \frac{1-\epsilon}{2} \sum_{l \neq k} \mathcal{N}(\tilde{b}_{nmu}; \mu_l, \sigma_l^2), \tag{3.22}$$

$$\mathcal{M}_k^{\neq}(\tilde{b}_{nmu}; \theta) = \epsilon_1 \mathcal{N}(\tilde{b}_{nmu}; \mu_k, \sigma_k^2) + \frac{1-\epsilon_1}{2} \sum_{l \neq k} \mathcal{N}(\tilde{b}_{nmu}; \mu_l, \sigma_l^2). \tag{3.23}$$

As ϵ is assumed to be small, $\mathcal{M}_k^0$ is dominated by the likelihood of the correlation being drawn from the k^{th} Normal distribution, whereas $\mathcal{M}_k^1$ is dominated by the likelihoods of the correlation being drawn from the other Normal distributions. As $\epsilon_1 > \epsilon$, $\mathcal{M}_k^{\neq}$ is an interpolation between these two terms as the value of η changes.

Now we can construct the joint probability distribution over all remaining random variables as

$$\begin{aligned} p(f, b, r, \tilde{b}; \theta) &= p(f; \gamma) p(b|f; \mu, \sigma) p(r; \pi) p(\tilde{b}|f, r; \theta) \\ &= \left(\prod_{n=1}^{N} \prod_{m>n} p(f_{nm}; \gamma) \prod_{h=1}^{H} p(b_{nmh}|f_{nm}; \mu, \sigma) \right) \\ &\quad \left(\prod_{u=1}^{U} \prod_{n=1}^{N} p(r_{nu}; \pi) p(\tilde{b}_{nmu}|f_{nm}, r_{nu}, r_{mu}; \eta, \epsilon, \mu, \sigma) \right). \end{aligned} \tag{3.24}$$

As we have marginalized out $\tilde{F}$ exactly, the variational factorization no longer accounts for it and takes the form

$$\begin{aligned} p(r,f|\tilde{b},b;\theta) &\approx q(r,f) = q_R(r)q_F(f) \\ &= \left(\prod_{n=1}^{N}\prod_{m>n}\prod_{k=-1}^{1}(q_{F_{nm}[k]})^{f_{nm}[k]}\right)\left(\prod_{u=1}^{U}\prod_{n=1}^{N}\prod_{l=0}^{1}(q_{R_{nu}[l]})^{r_{nu}[l]}\right). \end{aligned} \tag{3.25}$$

The resulting variational free energy is

$$\begin{aligned} \mathcal{E}(q,\theta;\tilde{b},b) &= -\mathbb{E}_q[\log p(F,R,\tilde{b},b;\theta)] + \mathbb{E}_q[\log q(F,R)] \\ &= -\mathbb{E}_{q_F}[\log p(F;\gamma)] - \mathbb{E}_{q_F}[\log p(b|F;\mu,\sigma)] - \mathbb{E}_{q_R}[\log p(R;\pi)] \\ &\quad - \mathbb{E}_{q_F q_R}[\log p(\tilde{b}|F,R;\epsilon,\eta,\mu,\sigma)] + \mathbb{E}_{q_R}[\log q_R(R)] + \mathbb{E}_{q_F}[\log q_F(F)]. \end{aligned} \tag{3.26}$$

As in Section 3.2, we update the posterior probability estimates to minimize this variational energy, but according to the following equations.

The healthy connectivity state posterior probability is updated in a similar fashion as in the T-algorithm:

$$\begin{aligned} \log q_{F_{nm}[k]} &= \log\gamma_k + \sum_{h=1}^{H}\log\mathcal{N}(b_{nmh};\mu_k,\sigma_k) \\ &+ \sum_{u=1}^{U} q_{R_{nu}[0]}q_{R_{mu}[0]}\log\mathcal{M}_k^0(\tilde{b}_{nmu};\theta) + q_{R_{nu}[1]}q_{R_{mu}[1]}\log\mathcal{M}_k^1(\tilde{b}_{nmu};\theta) \\ &+ \left(q_{R_{nu}[1]}q_{R_{mu}[0]} + q_{R_{nu}[0]}q_{R_{mu}[1]}\right)\log\mathcal{M}_k^{\neq}(\tilde{b}_{nmu};\theta) + \text{const.} \end{aligned} \tag{3.27}$$

The anomalous region state posterior probability update is also similar to the T-

algorithm:

$$\begin{aligned}\log q_{R_{nu}[0]} &= \log(1-\pi) && (3.28)\\ &+ \sum_{m\neq n}\sum_{k=-1}^{1} q_{F_{nm}[k]}\Big(q_{R_{mu}[0]}\log\mathcal{M}_k^0(\tilde{b}_{nmu};\theta) + q_{R_{mu}[1]}\log\mathcal{M}_k^{\neq}(\tilde{b}_{nmu};\theta)\Big) + \text{const},\\ \log q_{R_{nu}[1]} &= \log(\pi) && (3.29)\\ &+ \sum_{m\neq n}\sum_{k=-1}^{1} q_{F_{nm}[k]}\Big(q_{R_{mu}[1]}\log\mathcal{M}_k^1(\tilde{b}_{nmu};\theta) + q_{R_{mu}[0]}\log\mathcal{M}_k^{\neq}(\tilde{b}_{nmu};\theta)\Big) + \text{const}.\end{aligned}$$

By fixing the posterior probability estimates, the updates for the parameters π and γ are unchanged from Equations 3.16 and 3.17, as the energy terms dependent on these parameters do not change. However, by marginalizing over $\tilde{F}$, we introduce a new dependence between (ϵ, η) and $(\mu,\ \sigma)$. Therefore, we use the derivatives of the variational energy with respect to all these parameters to perform minimization of the energy using the same trust-region reflective optimization method as in the T-algorithm. Computing these derivatives is more computationally demanding than those in the T-algorithm, and convergence is slower due to increased dimensionality of the domain. Therefore, this optimization becomes the bottleneck of the $T\tilde{F}$-algorithm. The asymptotic algorithmic complexity of computing the updates and required derivatives is still $\mathcal{O}(N^2)$.

We can ease the computational demand of these updates by ignoring this dependence. We can simply update μ and σ as in Equations 3.18 and 3.19, but ignore the contributions of correlations from patients, as we no longer have estimates of the corresponding connectivity posterior probabilities. This relaxation assumes that we can accurately estimate the Normal distribution parameters accurately from the correlations of healthy subjects alone, and may not hold in real data where correlations of patients may be distributed differently. We then update only (ϵ, η) using the trust-region reflective optimization method. While these are updates not guaranteed to reduce the overall energy, they do minimize the terms with the largest contribution to that energy, and we find that minimization is stable in practice. We refer to this variant of the algorithm as the $T\tilde{F}H$-algorithm.

■ 3.4 Summary of the Algorithms

In summary, we propose three different algorithms for posterior probability and parameter estimation with respect to our model of anomalous region detection. These algorithms have the same high level structure common to the variational EM algorithm, which is described in Algorithm 1.

The T-algorithm marginalizes out the latent variable T that represents a binary graph of connection anomalies and approximates posterior probabilites of all other latent variables using a mean field factorization. All posterior probability and parameter updates have closed form solutions, except for ϵ and η, which are jointly optimized using an iterative descent method. The $T\tilde{F}$-algorithm additionally marginalizes out the latent variable $\tilde{F}$, which may be poorly approximated in the T-algorithm. This marginalization introduces a dependence between the parameters (ϵ, η) and the parameters (μ, σ), all of which must be jointly optimized using the iterative descent method. The $T\tilde{F}H$-algorithm relaxes this dependence by estimating (μ, σ) from healthy subjects only, and by optimizing (ϵ, η) separately using the iterative descent method.

These three variants of our inference algorithms are likely to have performance and efficiency tradeoffs, which we explore in the following chapter by applying them to synthetic data, generated from the model. We then apply the most appropriate algorithms to real data from a clinical study of epilepsy patients.

Chapter 4

Experiments

In this chapter, we evaluate the algorithms derived in the previous chapter on synthetic and real data from fMRI scans of epilepsy patients. In Section 4.1, we sample synthetic data from our generative model described in the previous chapter and evaluate the detection performance of the corresponding inference algorithms. In Section 4.2, we compare the detection performance of the most appropriate algorithms with some of the existing methods reviewed in Chapter 2, when applied to a group of epilepsy patients.

■ 4.1 Synthetic Experiments

We create synthetic data by generating samples of the latent and observed random variables from the model introduced in the previous chapter. Estimating model parameters from such data enables us to evaluate the proposed inference algorithms, and to identify how and when they fail. We also measure and compare the empirical computation time of the algorithms. For reference, we implement the algorithms in MATLAB and run them on a machine with 6 × 2.8GHz Intel Xeon X5660 processor cores.

■ 4.1.1 Data

Our main interest is investigating the quality of inference for different values of the parameters ϵ and η. In clinical populations, these parameters may vary for different disorders and even for different patients with the same disorder. As ϵ increases, the number of noisy abnormal connections between healthy regions increases. As η

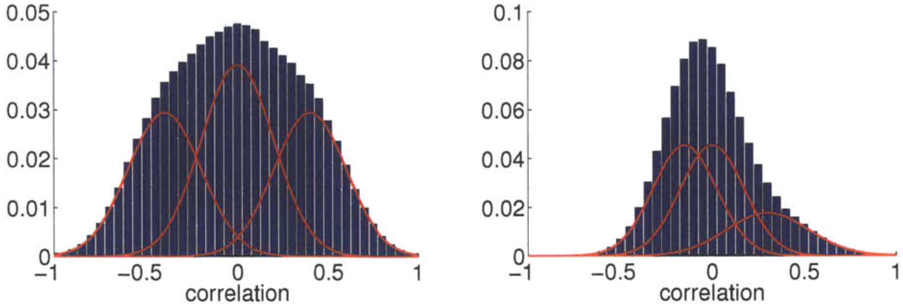

Figure 4.1: The two mixtures of Normal distributions (red) used for synthetic experiments with sample observations binned into histograms (blue). The *ideal* configuration (left) separates negative and positive connections from noise. The *real* configuration (right) uses parameter values estimated from epilepsy data, where connections are harder to distinguish.

decreases, the number of abnormal connections from unhealthy regions to healthy regions decreases, making unhealthy regions easier to miss. Ideally, we require an inference algorithm that performs well for relatively high values of ϵ and relatively low values of η. Here we generate data for pairs of these parameter values $(\epsilon, \eta) \in \{0.01, 0.02, \ldots, 0.05\} \times \{0.1, 0.2, \ldots, 0.5\}$.

The other parameters roughly describe how well we distinguish between different types of connections and unhealthy regions. Here we consider two configurations of $(\pi, \gamma, \mu, \sigma)$. The *ideal* configuration corresponds to values of these parameters where the Gaussian components are clearly separated, in that $\mu = (-0.4, 0, 0.4)$ and $\sigma = (0.04, 0.04, 0.04)$. Connections and anomalous regions are fairly frequent, in that $\gamma = (0.3, 0.4, 0.3)$ and $\pi = 0.1$. The *real* configuration corresponds to values of these parameters estimated from real epilepsy data. Here, the Gaussian components overlap considerably, in that $\mu = (-0.15, 0, 0.35)$ and $\sigma = (0.025, 0.035, 0.05)$. Connections and anomalous regions are less common, in that $\gamma = (0.4, 0.4, 0.2)$ and $\pi = 0.05$. We show these mixtures of Normal distributions with example observations in Figure 4.1. For each parameter configuration, we generate 10 samples of latent and observed random variables. We set $N = 148$, $H = 38$ and $U = 6$, which corresponds to the number of regions, healthy subjects and patients in the real epilepsy dataset.

■ 4.1.2 Evaluation

For each sample, each inference algorithm produces a set of estimated parameters and posterior probabilities. We perform inference using the T-algorithm, $T\tilde{F}$-algorithm and $T\tilde{F}H$-algorithm, first, by treating each patient individually, and second, by pooling patients together for parameter estimation. For each algorithm, we perform estimation with five different random parameter value initializations and choose the solution that corresponds to the lowest final variational energy. For evaluation purposes, we compare the MAP prediction of the estimated region anomaly posterior probability, $q_{R_{nu}}[1] > 0.5$, with the sample r_{nu} for all regions $n \in \{1, 2, \ldots, N\}$ and patients $u \in \{1, 2, \ldots, U\}$. We count the number of false positives and negatives to assess detection performance for different values of the parameters ϵ and η, and average these across patients.

■ 4.1.3 Results

Detection performance for the ideal configuration is summarized in Figure 4.2. The T-algorithm performs poorly. As $\pi = 0.1$, we expect around 15 anomalous regions. Yet, for a low value of $\eta = 0.1$ this algorithm misses almost half of these. Grouping patients for parameter estimation has little effect on and may even degrade performance. By comparison, the $T\tilde{F}$-algorithm is almost faultless, making fewer than 0.2 region detection errors, of either type, on average for all values of (ϵ, η). Grouping patients for parameter estimation decreases the average number of false negatives for small values of $\epsilon = 0.01$, but the effect is small. The $T\tilde{F}H$-algorithm performs almost identically, with results omitted for brevity.

Detection performance on the *real* configuration is summarized in Figure 4.3. Here, the T-algorithm makes a large number of false positive errors, especially when grouping unhealthy subjects for parameter estimation. As $\pi = 0.05$, the expected number of anomalous regions is around 7. Thus, the number of false negative errors is also high, especially for low values of η, but decreases when grouping patients. The $T\tilde{F}$-algorithm

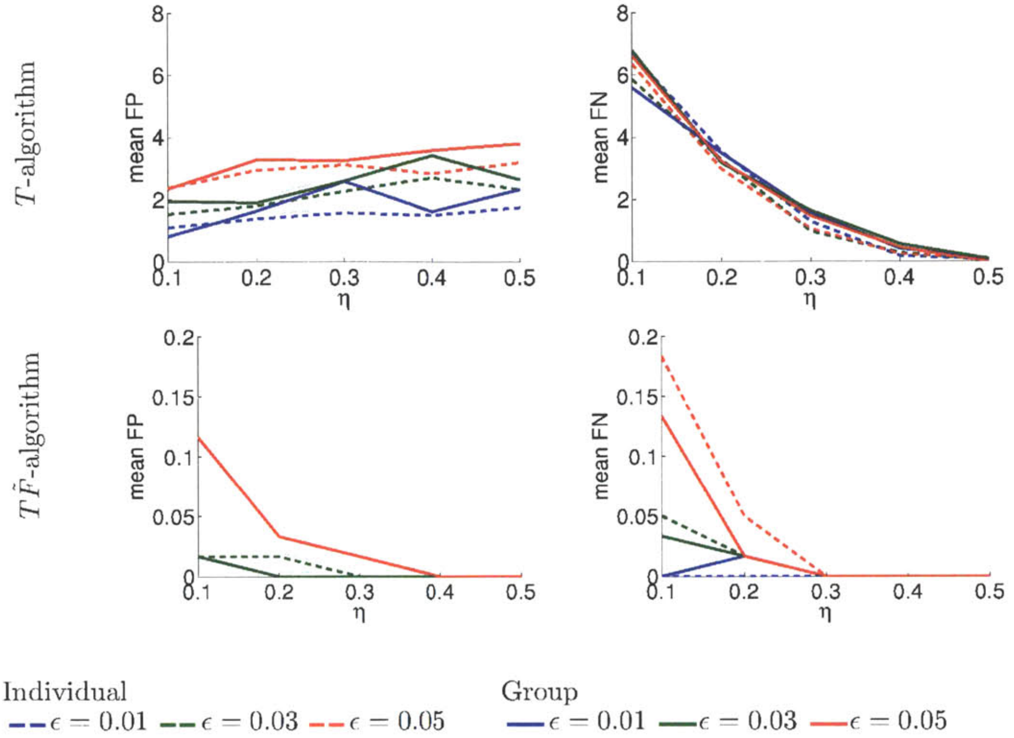

Figure 4.2: Detection performance of inference algorithms on *ideal* synthetic data. For combinations of ϵ and η, we show the mean number of false positives (left) and false negatives (right) per subject, for the T-algorithm (top), the $T\tilde{F}$-algorithm (bottom).

generally makes few false positive errors across all values of (ϵ, η), and makes few false negative errors except for low values of $\eta \leq 0.2$. Higher values of ϵ increase the number of false negatives and positives for these low values of η. The $T\tilde{F}H$-algorithm performs similarly, making slightly more false positive errors and slightly fewer false negative errors.

In summary, the $T\tilde{F}$-algorithm clearly outperforms the T-algorithm for both the ideal configuration of synthetic data and, more importantly, the real configuration. This performance comes at the cost of empirical run time, as the T-algorithm takes around 4s per initialization, whereas the $T\tilde{F}$-algorithm takes around 60s per initialization. The $T\tilde{F}H$-algorithm effectively matches the performance of the $T\tilde{F}$-algorithm, but is far

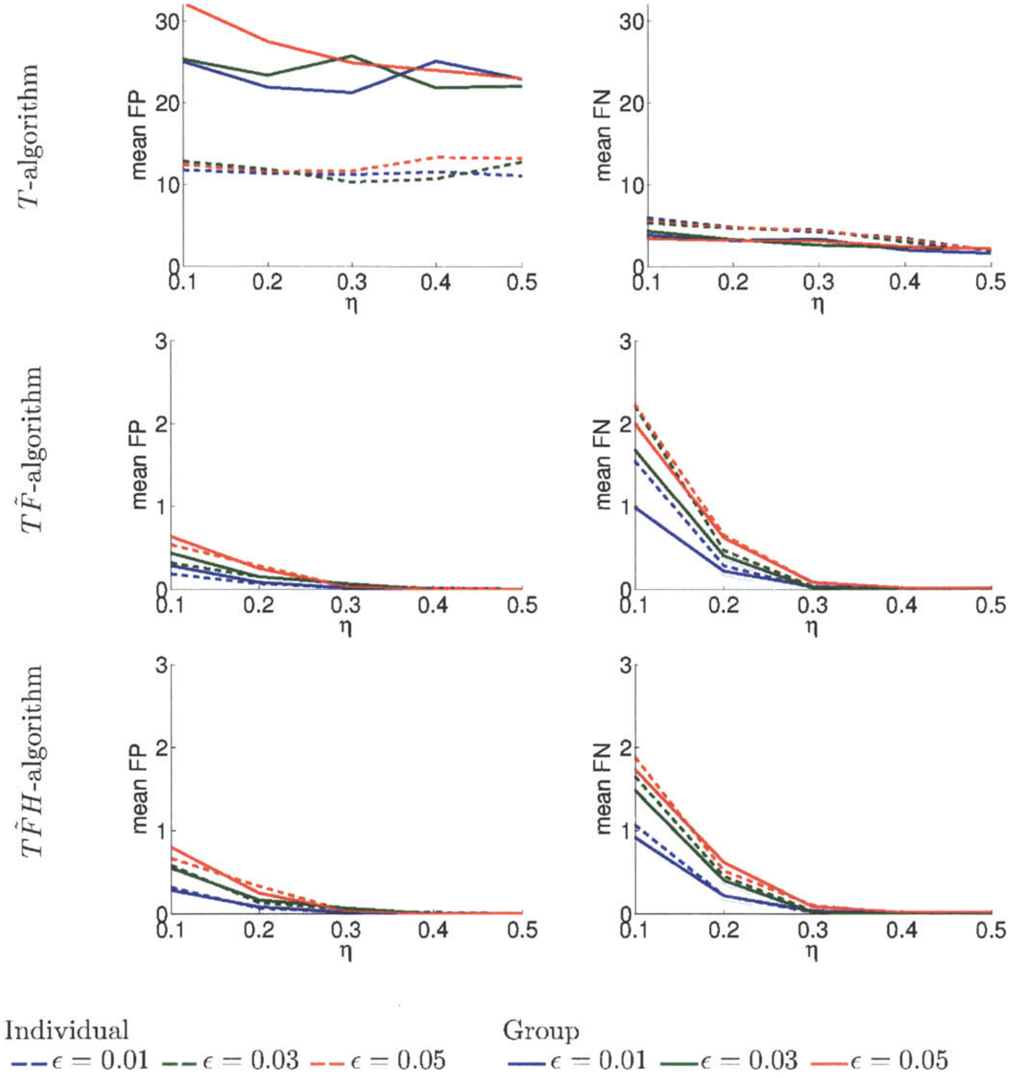

Figure 4.3: Detection performance of inference algorithms on *real* synthetic data. For combinations of ϵ and η, we show the mean number of false positives (left) and false negatives (right) per subject, for the T-algorithm (top), the $T\tilde{F}$-algorithm (middle) and the $T\tilde{F}H$-algorithm (top).

more computationally efficient, taking around 2.5s per initialization. Assuming the model holds in the epilepsy data, performance on the real configuration lets us know that the $T\tilde{F}$-algorithm and $T\tilde{F}H$-algorithms are more likely to miss anomalous regions than raise false alarms for healthy regions. In general, grouping the patients improves performance a little, but only for low values of η.

4.2 Epilepsy Experiments

4.2.1 Data & Pre-processing

We thank Steve Stufflebeam, Naoro Tanaka and Hesheng Liu for providing data of six epilepsy patients undergoing pre-surgical planning. For each patient, we obtain a pre-implantation 1mm isotropic MPRAGE T1 MRI volume, a single post-implantation CT volume with an isotropic in-plane resolution of 0.5mm and a slice thickness of 2.5mm, electrode spiking labels, and two to six runs of 2mm isotropic resting state BOLD fMRI volumes, each with 76 frames sampled every 5 seconds. For patient 5, the post-implantation CT volume has a slice thickness of 5mm instead of 2.5mm. We thank Randy Buckner for providing data of 38 subjects from a healthy population using the same MRI acquisition protocols. For each subject, we are provided with a pre-implantation 1mm isotropic MPRAGE T1 MRI volume, and three to four runs of 2mm isotropic resting state BOLD fMRI volumes each with 76 frames sampled every 5 seconds.

For each fMRI run, we perform the steps of the pre-processing pipeline discussed in Section 2.2, spatially normalize each fMRI volume to the MNI152 2mm template, and perform Gaussian spatial smoothing with a full-width half-maximum of 6mm. As the multiple fMRI runs of a single subject are aligned to the same template, we concatenate the time series at each voxel to produce a single sequence for each subject.

We use FreeSurfer [14] to produce a *coarse* parcelation of the cortex into 150 anatomical regions [9], from which we select $N = 148$ common regions of functional interest across subjects. We also subdivide the Freesurfer cortical surface template uniformly into $N = 1153$ cortical regions. The parcelations are illustrated in Figure 4.4. We align both parcelations to the MNI152 2mm template, so that they can be used to extract mean time series from the aligned fMRI sequence of each subject. Pearson correla-

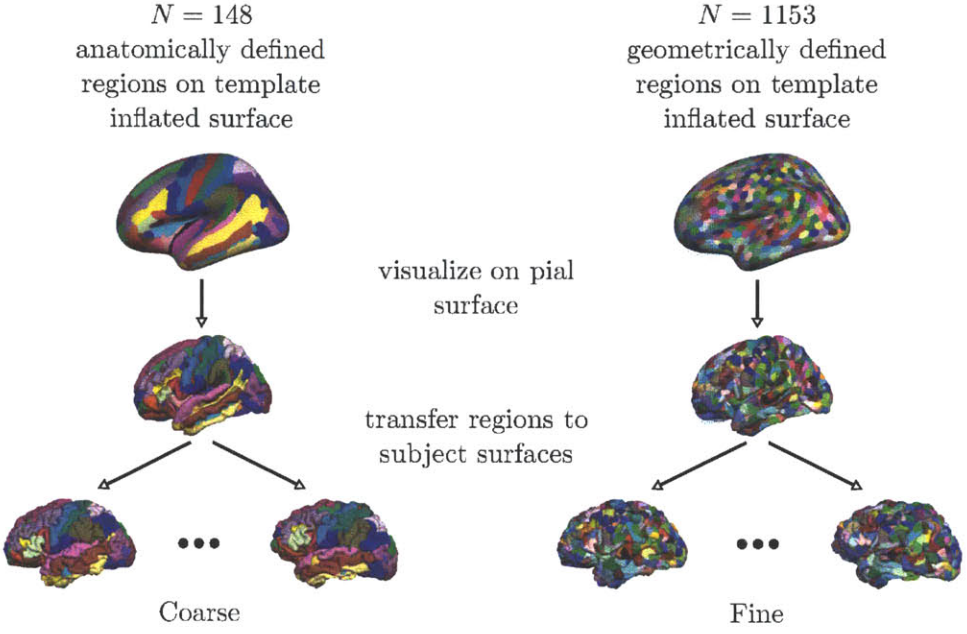

Figure 4.4: The *coarse* parcelation of 148 anatomically defined regions (left) and the *fine* parcelation into 1153 geometrically defined regions. Regions are defined on the Freesurfer template surface and then transferred to each subject's surface. Colors are randomly assigned to denote regions' identities.

tion coefficients are computed between the mean time series of the regions to provide observed correlations for each subject.

The post-implantation CT volume is rigidly registered to the MPRAGE T1 MRI volume using FSL's FLIRT [20] with a normalized mutual information similarity measure. We find that registration is generally accurate, as exemplified by the results of a single subject shown in Figure 4.5. We manually identify the electrode centers on each grid and strip, and correct for brain shift by simply projecting each electrode center to the closest point on the dural surface extracted using the method described in [38].

■ 4.2.2 Evaluation

For each electrode of each patient, we obtain a spiking label that represents one of the four possible observations: no spiking; interictal spiking only; ictal spiking only;

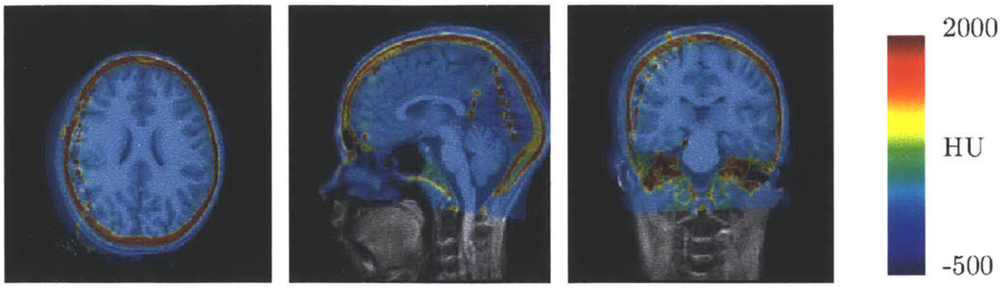

Figure 4.5: An axial, sagittal and coronal slice of the pre-implantation MPRAGE T1 MRI volume of a patient overlaid with the rigidly registered post-implantation CT volume. The CT overlay colors correspond to Hounsfield unit (HU) values. Registration is generally accurate, with the skull in both modalities closely aligned. However, due to brain shift deformation, the electrodes sometimes appear inside the cortical surface of the MRI volume.

and both interictal and ictal spiking. For this work, the last two labels are treated identically as indicators of the onset zone. For visualization purposes, electrode labels are transferred onto the dural surface by propagating them to all vertices within a 5mm radius of the corresponding corrected center. Detected anomalous regions in the cortical ribbon are also projected onto each vertex of the dural surface, by computing the maximum probability of anomaly along the vertex's normal from the dural surface to 20mm inside the brain, as illustrated in Figure 4.6. A maximum is used because spiking activity measured on the dural surface could be caused by a small epileptogenic region in the cortical ribbon.

■ 4.2.3 Coarse Parcelation Results

First, we use the coarse parcelation to generate a 148×148 correlation matrix for each subject. We run the $T\tilde{F}$-algorithm and $T\tilde{F}H$-algorithm for each patient individually and for the group of patients. For each variant of the algorithm, we repeat the estimation procedure with 20 different random parameter initializations and choose the solution with the lowest final variational energy. We find little difference in the resulting posterior probabilities, as illustrated for a single patient in Figure 4.7. Parameter estimates lie in

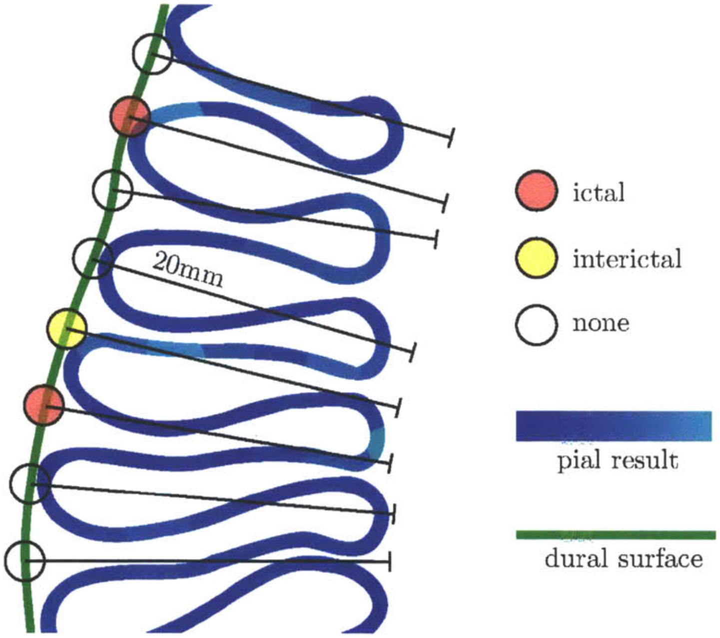

Figure 4.6: Our representation of the data means that results are associated with regions in the cortical ribbon. However, we wish to evaluate these results with electrode spiking labels on the dural surface that envelops the cortex. We associated each vertex on the dural surface with the maximum value along its normal vector 20mm inside the cortex. This allows us to visualize anomalous regions buried deep inside sulcal folds.

the following ranges across all patients: $\pi \in [0.04, 0.16]$, $\eta \in [0.2, 0.4]$, $\gamma \approx [0.4, 0.4, 0.2]$, $\epsilon \in [0.001, 0.1]$, $\mu \approx [-0.15, 0, 0.35]$, $\sigma^2 \approx [0.025, 0.035, 0.05]$. With the exception of patient 1, where $\epsilon \approx 0.1$, parameter values fall in the ranges where we expect the algorithms to perform well.

For baseline comparison, we also apply the two statistical approaches described in Sections 2.2.1 and 2.2.2. As a reminder, the region correlation statistic $p_{nu}(\alpha)$, defined in Equation 2.1, is the proportion of correlation z-scores whose absolute value is above a threshold α. The region degree statistic $z_{nu}(\beta)$, defined in Equation 2.2, is the absolute z-score of the proportion of correlations above a threshold β. We compute each statistic with a range of thresholds, $\alpha \in \{1, 1.5, \ldots, 5\}$ and $\beta \in \{0.2, 0.3, 0.4, 0.5\}$, and find that both statistics are sensitive to the choice of threshold, as illustrated for a single patient

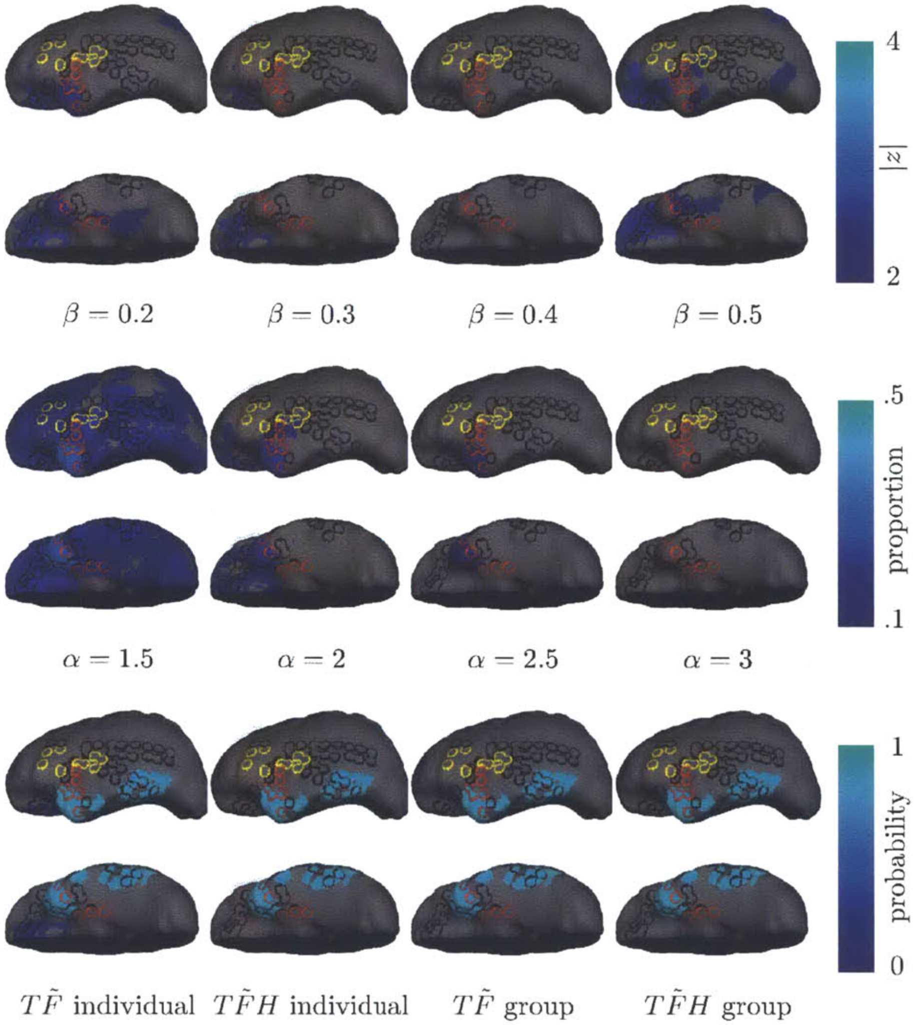

Figure 4.7: Coarse parcelation results for Patient 2 when sweeping the threshold β on the degree statistic, the threshold α on the correlation statistic (middle) and the variants of our algorithm (bottom).

in Figure 4.7.

Across all patients, we find that the optimal correlation statistic $p_{nu}(\alpha = 2)$ generally produces better results than the optimal degree statistic $z_{nu}(\beta = 0.5)$ when visually evaluated with respect to the electrode labels. We compare this optimal correlation statistic with the posterior probabilities of the $T\tilde{F}$-algorithm for individual patients in Figure 4.8. In general, we observe that the correlation statistic detects regions that overlap with almost all electrodes that exhibit ictal spiking, except in patient 5. However, detections also overlap with many electrodes that exhibit no spiking and performance is far from ideal. For example, almost all regions are detected in Patient 1. Detections occasionally overlap with interictal spiking, but the relationship is unclear. The posterior probabilities of the $T\tilde{F}$-algorithm are typically binary, and effectively select a subset of the regions with high correlation statistics. However, we can see that this selection is poor with respect to the electrode labels. For example, ictal spiking regions detected by the correlation statistic in patients 3 and 6 are not detected by the algorithm.

Poor performance of the algorithm could be due a number of reasons. The actual value of correlation may be informative for detection of the epileptogenic zone. The correlation statistic can distinguish between a correlation of high value, say 0.5, and very high value, say 0.7, whereas the model assumes both these correlations are almost certainly observations of the same underlying connected state. Additionally, the model assumes that anomalous regions tend to form cliques in the abnormal graph T, which may not be a valid assumption for epileptogenic networks in the coarse parcelation. We explore this latter possibility in the following section by using a finer parcelation.

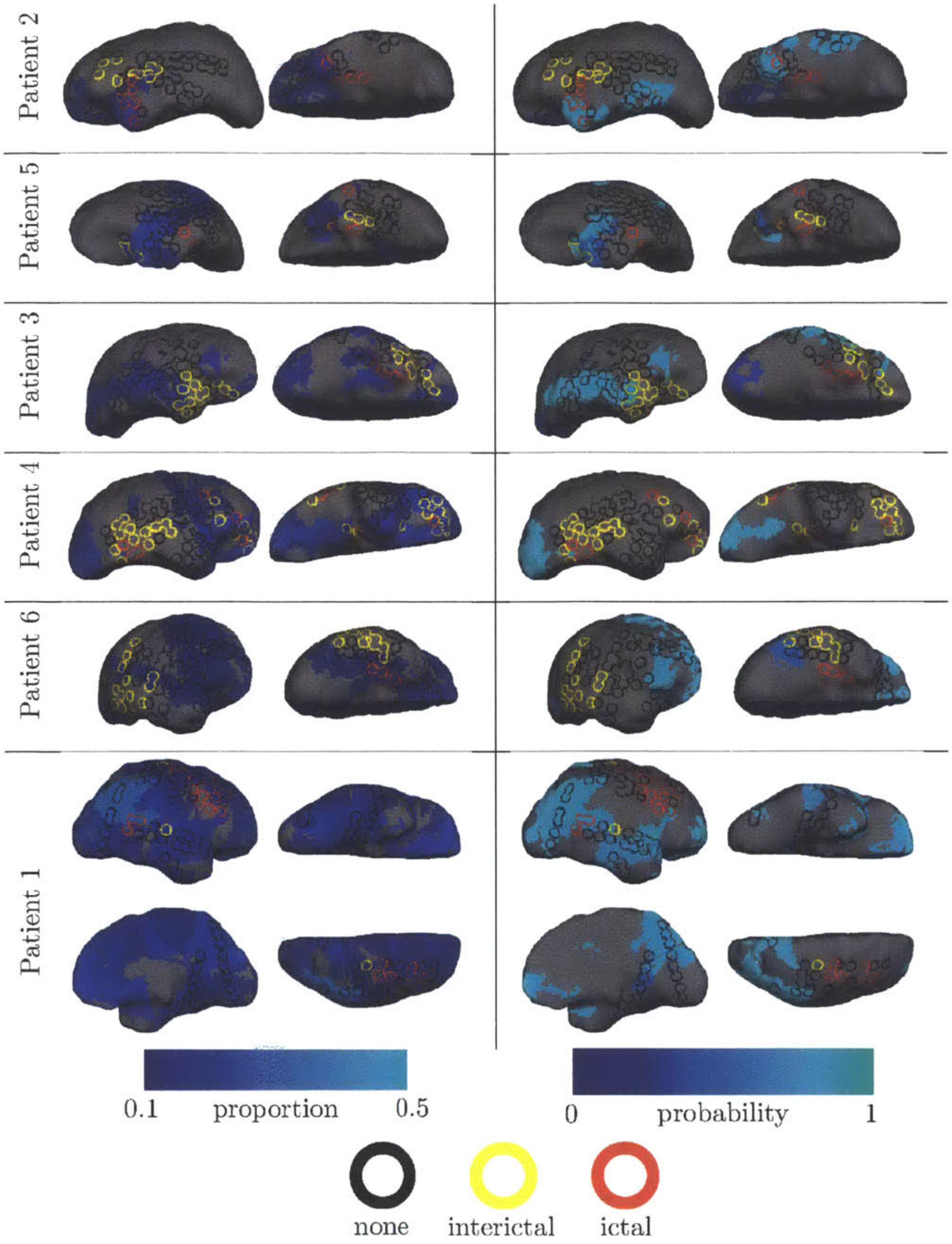

Figure 4.8: The coarse regional correlation statistic (left) and the posterior probabilities of the $T\tilde{F}$-algorithm (right), projected to the dural surface of each patient. Electrode spiking labels are also shown as 5mm radius circles on the surface. Only views with electrode coverage are shown.

4.2.4 Fine Parcelation Results

In order to explore whether the resolution of the coarse parcelation is limiting the performance of localization, we use the fine parcelation to generate a 1153×1153 correlation matrix for each subject. We run the same algorithms as described in Section 4.2.3. We find that resulting posterior probability estimates are similar across algorithms with some small differences, as illustrated for a single patient in Figure 4.9. We also generate the same statistics as described in Section 4.2.3, and find that they are again sensitive to the choice of threshold, as illustrated for a single patient in Figure 4.9.

Across all patients, we find that the optimal correlation statistic $p_{nu}(\alpha = 2)$ generally produces better results than the optimal degree statistic $z_{nu}(\beta = 0.5)$ when visually evaluated with respect to the electrode labels. We also find that the $T\tilde{F}$-algorithm applied to individual patients produces the best results of our algorithm variants. Grouping the patients together for parameter estimation reduces the number of regions detected, and misses one or two spiking regions as a result. Parameter estimates for the individual $T\tilde{F}$-algorithm lie in the following ranges: $\pi \in [0.04, 0.22]$, $\gamma \approx [0.3, 0.5, 0.2]$, $\eta \in [0.18, 0.52]$, $\epsilon \in [0.001, 0.05]$, $\mu_{-1} \in [-0.17, -0.1]$, $\mu_1 \in [0.22, 0.29]$ $\sigma^2 \approx [0.025, 0.025, 0.05]$. Although the values for π, η and ϵ fall in the range where we expect the algorithms to perform well, the values of μ and γ are untested in synthetic experiments and may make posterior recovery more difficult.

We compare these two results in Figure 4.10. Here, we see that the fine parcelation allows the correlation statistic to detect some epileptogenic regions that were undetected from the coarse parcelation. For example, two of the ictal spiking electrodes in patient 5 are now detected. However, this small improvement in sensitivity to ictal spiking regions comes at the expense of a lack of specificity, as even more regions are found in the vicinity of electrodes that exhibit no spiking. Again, we see that posterior probabilities estimated by our algorithm are typically binary and mostly select a subset of the regions with high correlation statistics. If we concentrate on areas of the surface with electrode coverage, we see that the model removes many of the false positive detections detected

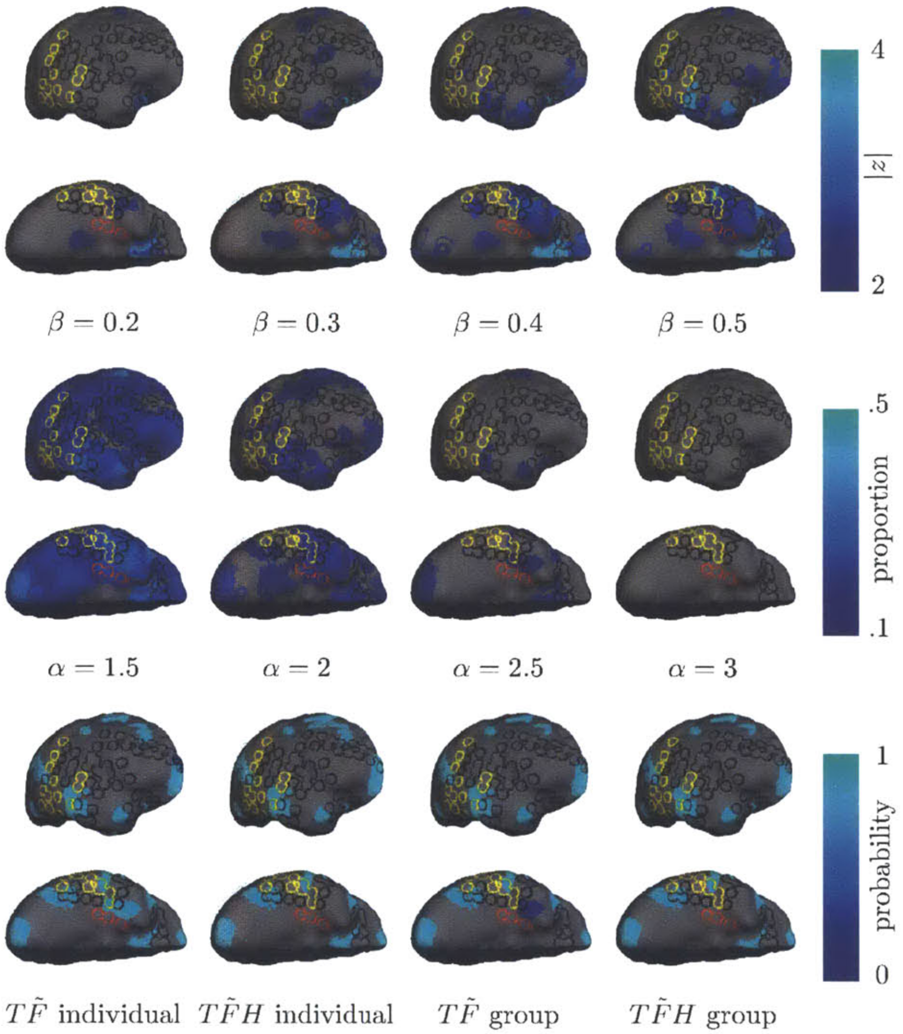

Figure 4.9: Fine parcelation results for Patient 2 when sweeping the threshold β on the degree statistic, the threshold α on the correlation statistic (middle) and the variants of our algorithm (bottom).

by the correlation statistic. However, it also makes some false negative errors, missing detections around ictal spiking electrodes. For example, in patient 6, the algorithm fails to select a spiking region that is detected by the correlation statistic.

Both approaches also detect regions where there is no electrode coverage. Although we can only speculate about whether these regions are epileptogenic, they may be suitable candidate positions for electrodes. In this case, one potential advantage of the algorithm is that it detects a relatively small number of coherent regions, whereas the correlation statistic detects more regions that are scattered across the cortex.

■ 4.3 Summary

By using our generate model to produce synthetic data, we find that the $T\tilde{F}$-algorithm clearly outperforms the T-algorithm, but requires significantly more computation time. The $T\tilde{F}H$-algorithm effectively matches the performance of the $T\tilde{F}$-algorithm, but is far more computationally efficient. When applied to data from epilepsy patients, we observe that using the geometrically defined fine parcelation instead of the anatomically defined coarse parcelation improves the detection of ictal spiking regions, both for the baseline statistical approaches and for our algorithms. The improved performance is characterized by higher sensitivity to these regions, but comes at the cost of reduced specificity, especially in the case of correlation statistics. We observe that our algorithm produces close to binary posterior probabilities and effectively selects a relatively small subset of the regions with high correlation statistics. In the next chapter, we clarify the contributions of the thesis with respect to the results and the original goals of the work. We also propose further work to address some of the limitations illustrated by the results.

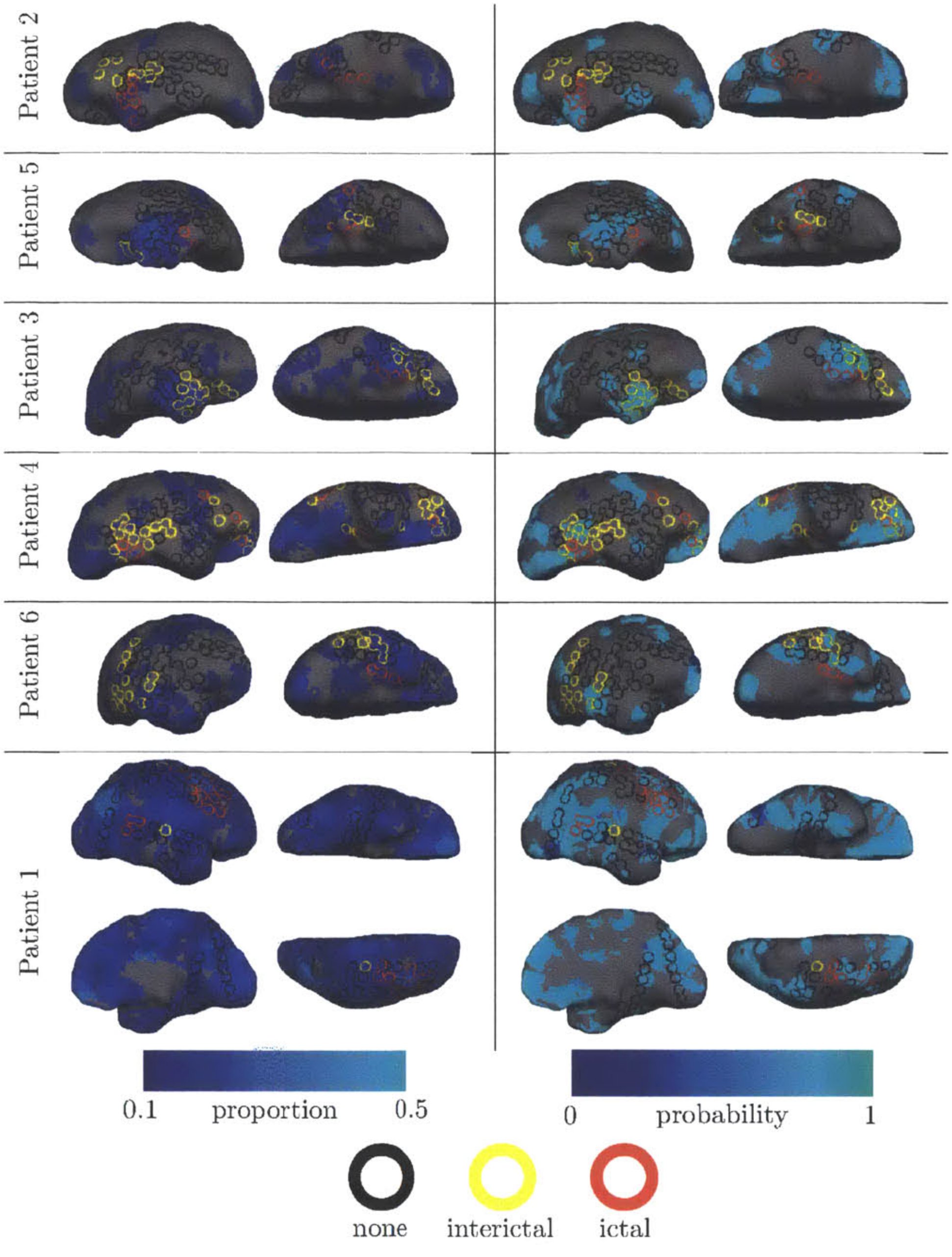

Figure 4.10: The fine regional correlation statistics (left) and the posterior probabilities of the $T\tilde{F}$-algorithm (right), projected to the dural surface of each patient. Electrode spiking labels are also shown as 5mm radius circles on the surface. Only views with electrode coverage are shown.

Chapter 5

Discussion

In this chapter, we summarize the contributions of this thesis and connect these back to the original goals of the work. Based on the limitations of these contributions and insights from this project, we propose avenues of investigation for future work.

■ 5.1 Contributions

The main contribution of this thesis is the adaptation of the method described in [44] so that it can be efficiently run when the number of regions of interest is large. This is important because we find our method produces much better results on epilepsy data for the fine parcelation than when applied to the coarse parcelation. For the fine parcelation, we find that the method frequently detects onset zone regions where iEEG electrodes exhibit ictal spiking. While baseline statistical approaches at least match the performance of the method, they require manual selection of threshold parameters, which would not be optimal without the electrode labels. By contrast, our method determines all its parameters automatically. However, performance is far from ideal and clearly the method could not be used to replace iEEG for onset zone localization in clinical practice. Yet, the proposed algorithm shows some promise as a coarse localizer that could be used to improve the placement of iEEG electrodes. Not only does it detect regions in the vicinity of almost all ictal spiking electrodes, but it also identifies a relatively small number of other regions that are not covered by electrodes and may potentially contain other epileptogenic areas.

■ 5.2 Limitations & Future Work

Even though the electrode spiking labels are effectively used as a ground truth, we limit our evaluation of results to qualitative visual comparison rather than compute any quantitative measure of overlap. This is largely because iEEG electrodes are located on the dural surface of the brain, whereas the epileptogenic source regions that cause abnormal spiking observed in the electrodes' time series can be located anywhere in the cortex. Localizing the cortical source of abnormal neural spiking from the dural iEEG measurements is a difficult challenge in itself and is worthy of further work.

Even so, our qualitative evaluation of the results suggests that proposed method has some limitations. In its existing form, our method estimates posterior probabilities and parameter values using a number of different random initializations of the parameter values, and chooses the solution with the lowest final variational energy. The $T\tilde{F}$-algorithm, which exhibits the best performance in our evaluation, produces solutions with slightly different estimates of the posterior probabilities, all of which tend to have similar variational energies. Therefore, taking the solution with the lowest energy may be somewhat arbitrary. Instead, we could produce a consensus result of the posterior probabilities across all solutions. For example, simply computing the arithmetic mean posterior probabilities seems to improve the performance of the algorithm in some cases, as illustrated in Figure 4.10. Here, we identify the ictal spiking region in Patient 6, which is missed the current approach.

This result suggests that our model does not entirely account for the data we observe in epilepsy patients. One limitation of the model is that the template of healthy connectivity between two regions is assumed to be entirely captured by the multinomial random variable F. This means that we cannot account for variability of a connection across the healthy population. The correlation statistic does account for this variability, which may explain why it exhibits better performance in some patients. Another way to address this limitation is to perform permutation tests with healthy subjects to create a null distribution of region posterior probabilities. This construction would allow us to capture the probability of detecting each region by chance, and use it to evaluate the significance of each patient's anomalous region posterior probability value.

We can also consider alternative pre-processing strategies that may improve epileptogenic zone localization. In this work, we only consider cortical regions. However, subcortical structures, such as the hippocampus, thalamus and basal ganglia, also contain functional signals that may better capture the network abnormalities associated with epileptogenic regions. Note that due to their location, iEEG spiking labels could not be used to evaluate detections in these subcortical regions. For the coarse parcelation, we have experimented with including these subcortical regions, but found worse performance. For the fine parcelation, it is unclear how to segment these subcortical regions into parcels of a similar size.

In this work, we find that grouping epilepsy patients has little effect on the performance for synthetic or epilepsy data. Instead, if we consider an individual patient, we may be able to improve localization by biasing pre-processing to that patient. For example, instead of aligning the patient and all healthy subjects to the MNI152 template, we could align the healthy subjects to the patient. Furthermore, we could define a parcelation based the resting state fMRI signals of the patient and propagate the parcelation to the aligned healthy subjects. Each region would then correspond to a discrete functional unit of the patient of interest, rather than a shared anatomically or geometrically defined area.

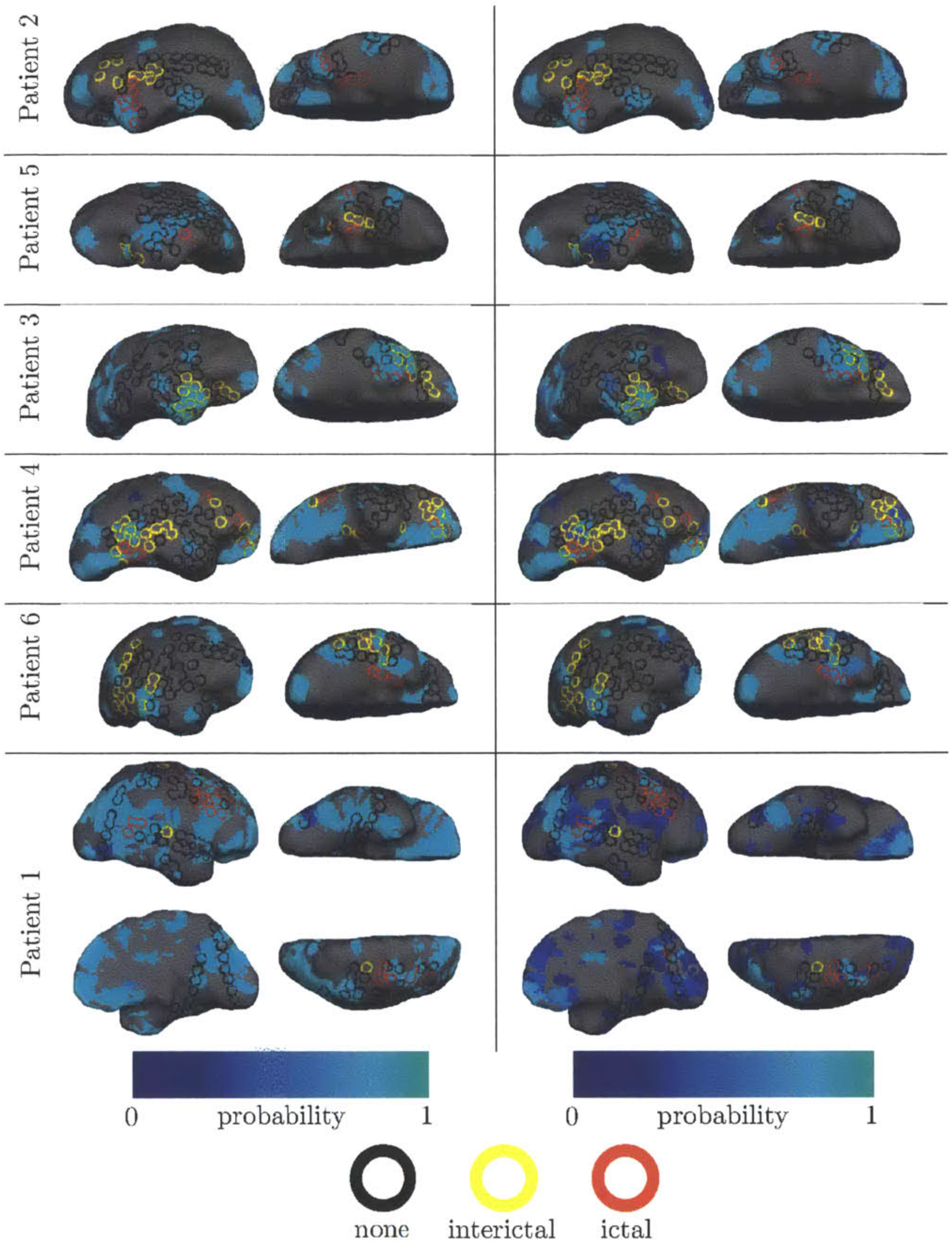

Figure 5.1: The posterior probabilities of the $T\tilde{F}$-algorithm using the solution with the lowest variational energy (left) and using the average solution (right), projected to the dural surface of each patient. Electrode spiking labels are also shown as 5mm radius circles on the surface. Only views with electrode coverage are shown.

Bibliography

[1] K. Abraham and C. A. Marsan. Patterns of Cortical Discharges and their Relation to Routine Scalp Electroencephalography. *Electroencephalography and Clinical Neurophysiology*, 10(3):447–461, 1958.

[2] M. J. Beal and Z. Ghahramani. The Variational Bayesian EM Algorithm for Incomplete Data: with Application to Scoring Graphical Model Structures. *Bayesian Statistics*, 7:1–10, 2003.

[3] G. Bettus, E. Guedj, F. Joyeux, S. Confort-Gouny, E. Soulier, V. Laguitton, P. J. Cozzone, P. Chauvel, J.-P. Ranjeva, F. Bartolomei, and M. Guye. Decreased Basal fMRI Functional Connectivity in Epileptogenic Networks and Contralateral Compensatory Mechanisms. *Human Brain Mapping*, 30(5):1580–1591, 2009.

[4] B. Biswal, F. Z. Yetkin, V. M. Haughton, and J. S. Hyde. Functional Connectivity in the Motor Cortex of Resting Human Brain using Echo-planar MRI. *Magnetic Resonance in Medicine*, 34(4):537–541, 1995.

[5] R. L. Buckner, J. Sepulcre, T. Talukdar, F. M. Krienen, H. Liu, T. Hedden, J. R. Andrews-Hanna, R. A. Sperling, and K. A. Johnson. Cortical Hubs Revealed by Intrinsic Functional Connectivity: Mapping, Assessment of Stability, and Relation to Alzheimer's Disease. *The Journal of Neuroscience*, 29(6):1860–1873, 2009.

[6] E. Bullmore and O. Sporns. Complex Brain Networks: Graph Theoretical Analysis of Structural and Functional Systems. *Nature Reviews: Neuroscience*, 10(3):186–198, 2009.

[7] X. J. Chai, S. Whitfield-Gabrieli, A. K. Shinn, J. D. E. Gabrieli, A. Nieto Castañón, J. M. McCarthy, B. M. Cohen, and D. Ongür. Abnormal Medial Prefrontal Cortex Resting-state Connectivity in Bipolar Disorder and Schizophrenia. *Neuropsychopharmacology*, 36(10):2009–2017, 2011.

[8] R. Cooper, A. Winter, H. Crow, and W. Walter. Comparison of Subcortical, Cortical and Scalp Activity using Chronically Indwelling Electrodes in Man. *Electroencephalography and Clinical Neurophysiology*, 18(3):217–228, 1965.

[9] R. S. Desikan, F. Ségonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, M. S. Albert, and R. J. Killiany. An Automated Labeling System for Subdividing the Human Cerebral Cortex on MRI Scans into Gyral Based Regions of Interest. *NeuroImage*, 31(3):968–80, 2006.

[10] A. R. Dykstra, A. M. Chan, B. T. Quinn, R. Zepeda, C. J. Keller, J. Cormier, J. R. Madsen, E. N. Eskandar, and S. S. Cash. Individualized Localization and Cortical Surface-based Registration of Intracranial Electrodes. *NeuroImage*, 59(4):3563–3570, 2012.

[11] J. Engel. Surgery for Seizures. *New England Journal of Medicine*, 334(10):647–652, 1996.

[12] J. Engel. Finally, a Randomized, Controlled Trial of Epilepsy Surgery. *New England Journal of Medicine*, 345(5):365–367, 2001.

[13] A. Evans and D. Collins. 3D Statistical Neuroanatomical Models from 305 MRI Volumes. In *IEEE Nuclear Science Symposium and Medical Imaging Conference*, pages 1813–1817, 1993.

[14] B. Fischl. FreeSurfer. *NeuroImage*, 62(2):774–781, 2012.

[15] B. Fischl, M. I. Sereno, and A. M. Dale. Cortical Surface-based Analysis II: Inflation, Flattening, and a Surface-based Coordinate System. *NeuroImage*, 9(2):195–207, 1999.

[16] S. Geman and D. Geman. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–41, 1984.

[17] M. Guye, G. Bettus, F. Bartolomei, and P. J. Cozzone. Graph Theoretical Analysis of Structural and Functional Connectivity MRI in Normal and Pathological Brain Networks. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 23(5-6):409–421, 2010.

[18] D. J. Heeger and D. Ress. What does fMRI tell us about Neuronal Activity? *Nature Reviews: Neuroscience*, 3(2):142–151, 2002.

[19] G. Jackson, R. Kuzniecky, and S. Berkovic. Introduction to Epilepsy. In G. Jackson and R. Kuzniecky, editors, *Magnetic Resonance in Epilepsy: Neuroimaging Techniques*, chapter 1, pages 1–16. Elsevier Academic Press, Burlington, second edition, 2005.

[20] M. Jenkinson and S. Smith. A Global Optimisation Method for Robust Affine Registration of Brain Images. *Medical Image Analysis*, 5(2):143–156, 2001.

[21] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul. An Introduction to Variational Methods for Graphical Models. *Machine Learning*, 37:183–233, 1999.

[22] M. A. Koch, D. G. Norris, and M. Hund-Georgiadis. An Investigation of Functional and Anatomical Connectivity using Magnetic Resonance Imaging. *NeuroImage*, 16(1):241–250, 2002.

[23] M. A. Kramer and S. S. Cash. Epilepsy as a Disorder of Cortical Network Organization. *The Neuroscientist*, 18(4):360–72, 2012.

[24] P. S. LaViolette, S. D. Rand, B. M. Ellingson, M. Raghavan, S. M. Lew, K. M. Schmainda, and W. Mueller. 3D Visualization of Subdural Electrode Shift as Measured at Craniotomy Reopening. *Epilepsy Research*, 94(1-2):102–109, 2011.

[25] L. Lemieux, J. Daunizeau, and M. C. Walker. Concepts of Connectivity and Human Epileptic Activity. *Frontiers in Systems Neuroscience*, 5:12, 2011.

[26] T. Loddenkemper, C. Kellinghaus, E. Wyllie, I. M. Najm, A. Gupta, F. Rosenow, and H. O. Lüders. A Proposal for a Five-dimensional Patient-oriented Epilepsy Classification. *Epileptic Disorders*, 7(4):308–316, 2005.

[27] G. Lohmann, D. S. Margulies, A. Horstmann, B. Pleger, J. Lepsien, D. Goldhahn, H. Schloegl, M. Stumvoll, A. Villringer, and R. Turner. Eigenvector Centrality Mapping for Analyzing Connectivity Patterns in fMRI Data of the Human Brain. *PloS one*, 5(4):e10232, 2010.

[28] H. O. Lüders, I. Najm, D. Nair, P. Widdess-Walsh, and W. Bingman. The Epileptogenic Zone: General Principles. *Epileptic Disorders*, 8(Suppl 2):S1–9, 2006.

[29] T. E. Lund, K. H. Madsen, K. Sidaros, W.-L. Luo, and T. E. Nichols. Non-white Noise in fMRI: does Modelling Have an Impact? *NeuroImage*, 29(1):54–66, 2006.

[30] C. Luo, C. Qiu, Z. Guo, J. Fang, Q. Li, X. Lei, Y. Xia, Y. Lai, Q. Gong, D. Zhou, and D. Yao. Disrupted Functional Brain Connectivity in Partial Epilepsy: a Resting-state fMRI Study. *PloS one*, 7(1):e28196, 2011.

[31] J. C. Mazziotta, A. W. Toga, A. Evans, P. Fox, and J. Lancaster. A Probabilistic Atlas of the Human Brain: Theory and Rationale for its Development. *NeuroImage*, 2(2):89–101, 1995.

[32] M. L. McGill, O. Devinsky, C. Kelly, M. F. Milham, X. Castellanos, B. T. Quinna, J. DuBois, J. R. Young, C. Carlson, J. French, R. Kuzniecky, E. Halgren, and T. Thesen. Default Mode Network Abnormalities in Idiopathic Generalized Epilepsy. *Epilepsy & Behavior*, 23(3):353–359, 2012.

[33] V. L. Morgan, R. R. Price, A. Arain, P. Modur, and B. Abou-Khalil. Resting functional MRI with Temporal Clustering Analysis for Localization of Epileptic Activity without EEG. *NeuroImage*, 21(1):473–481, 2004.

[34] K. Murphy, R. M. Birn, D. a. Handwerker, T. B. Jones, and P. a. Bandettini. The Impact of Global Signal Regression on Resting State Correlations: are Anti-correlated Networks Introduced? *NeuroImage*, 44(3):893–905, 2009.

[35] E. Niedermeyer and F.L. da Silva. *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields.* Lippincott, Williams & Wilkins, fifth edition, 2004.

[36] S. Ogawa, T. M. Lee, A. R. Kay, and D. W. Tank. Brain Magnetic Resonance Imaging with Contrast Dependent on Blood Oxygenation. *Proceedings of the National Academy of Sciences of the United States of America*, 87(24):9868–9872, 1990.

[37] F. R. S. Pereira, A. Alessio, M. S. Sercheli, T. Pedro, E. Bilevicius, J. M. Rondina, H. F. B. Ozelo, G. Castellano, R. J. M. Covolan, B. P. Damasceno, and F. Cendes. Asymmetrical Hippocampal Connectivity in Mesial Temporal Lobe Epilepsy: Evidence from Resting State fMRI. *BMC Neuroscience*, 11:66, 2010.

[38] M. Schaer, M. B. Cuadra, L. Tamarit, F. Lazeyras, S. Eliez, and J.-P. Thiran. A Surface-based Approach to Quantify Local Cortical Gyrification. *IEEE Transactions on Medical Imaging*, 27(2):161–170, 2008.

[39] J. Sepulcre, H. Liu, T. Talukdar, I. n. Martincorena, B. T. T. Yeo, and R. L. Buckner. The Organization of Local and Distant Functional Connectivity in the Human Brain. *PLoS Computational Biology*, 6(6):e1000808, 2010.

[40] O. Sporns, G. Tononi, and G. M. Edelman. Theoretical Neuroanatomy: Relating Anatomical and Functional Connectivity in Graphs and Cortical Connection Matrices. *Cerebral Cortex*, 10(2):127–141, 2000.

[41] S. M. Stufflebeam, H. Liu, J. Sepulcre, N. Tanaka, R. L. Buckner, and J. R. Madsen. Localization of Focal Epileptic Discharges using Functional Connectivity Magnetic Resonance Imaging. *Journal of Neurosurgery*, 114(6):1693–7, 2011.

[42] K. R. Van Dijk, T. Hedden, A. Venkataraman, K. C. Evans, S. W. Lazar, and R. L. Buckner. Intrinsic Functional Connectivity as a Tool for Human Connectomics: Theory, Properties, and Optimization. *Journal of Neurophysiology*, 103(1):297–321, 2010.

[43] G. Varoquaux, F. Baronnet, A. Kleinschmidt, P. Fillard, and B. Thirion. Detection of Brain Functional-connectivity Difference in Post-stroke Patients using Group-level Covariance Modeling. In *MICCAI: International Conference on Medical Image Computing and Computer Assisted Intervention*, pages 200–208, 2010.

[44] A. Venkataraman, M. Kubicki, and P. Golland. From Brain Connectivity Models to Identifying Foci of a Neurological Disorder. In *MICCAI: International Conference*

on Medical Image Computing and Computer Assisted Intervention, pages 715–722, 2012.

[45] A. Venkataraman, Y. Rathi, M. Kubicki, C.-F. Westin, and P. Golland. Joint Generative Model for fMRI/DWI and its Application to Population Studies. In *MICCAI: International Conference on Medical Image Computing and Computer Assisted Intervention*, pages 191–199, 2010.

[46] P. Viola and W. W. III. Alignment by Maximization of Mutual Information. *International Journal of Computer Vision*, 24(2):1–29, 1997.

[47] A. Weissenbacher, C. Kasess, F. Gerstl, R. Lanzenberger, E. Moser, and C. Windischberger. Correlations and Anticorrelations in Resting-state Functional Connectivity MRI: a Quantitative Comparison of Preprocessing Strategies. *NeuroImage*, 47(4):1408–1416, 2009.

[48] A. I. Yang, X. Wang, W. K. Doyle, E. Halgren, C. Carlson, T. L. Belcher, S. S. Cash, O. Devinsky, and T. Thesen. Localization of Dense Intracranial Electrode Arrays using Magnetic Resonance Imaging. *NeuroImage*, 63(1):157–165, 2012.

[49] B. T. T. Yeo, F. M. Krienen, J. Sepulcre, M. R. Sabuncu, M. Hollinshead, J. L. Roffman, J. W. Smoller, J. R. Polimeni, B. Fischl, H. Liu, R. L. Buckner, J. Lu, M. Zhang, D. Wang, Y. Cao, Q. Ma, X.-N. Zuo, R. Ehmke, M. Mennes, D. Imperati, F. X. Castellanos, M. P. Milham, D. Lashkari, and L. Zöllei. The Organization of the Human Cerebral Cortex Estimated by Intrinsic Functional Connectivity. *Journal of Neurophysiology*, 106:1125–1165, 2011.