2015

MODELING BRAIN DESYNCHRONIZATION BY EEG SENSOR VARIANCE IN EPILEPTIC PATIENTS

Craig Michael Krebsbach
University of Rhode Island, craigkrebsbach@gmail.com

Recommended Citation

Krebsbach, Craig Michael, "MODELING BRAIN DESYNCHRONIZATION BY EEG SENSOR VARIANCE IN EPILEPTIC PATIENTS" (2015). *Open Access Master's Theses.* Paper 545.
https://digitalcommons.uri.edu/theses/545

MODELING BRAIN DESYNCHRONIZATION BY EEG SENSOR VARIANCE IN EPILEPTIC PATIENTS

BY

CRAIG MICHAEL KREBSBACH

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
STATISTICS

UNIVERSITY OF RHODE ISLAND
2015

MASTER OF SCIENCE THESIS

OF

CRAIG MICHAEL KREBSBACH

APPROVED:

Thesis Committee:

Major Professor Gavino Puggioni

Natallia Katenka

Kunal Mankodiya

Nasser Zawia

DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND

2015

ABSTRACT

The use of electroencephalogram (EEG) for predictive purposes of seizures in epileptic patients has grown steadily with the access to greater computing power. Methods of seizure analysis to date have focused on modeling and computer aided machine learning to help increase sensitivity and specificity of seizure detection. Brain synchronization between various areas of the brain at the onset of seizures tends to be a common feature of seizures, followed by a resynchronization of the various brain areas at the end of seizures. While previous methods have looked at the cross-correlation or lag-correlation of only two areas the brain, most EEG data these days has a vast array of sensors that can easily exceed 15-20 areas of the brain. The goal of this research is to take a relatively new approach to statistical modeling of multivariate EEG data, by use of the variance of multiple sensors as an extended measure of brain desynchronization in a time series format. Use of the Children's Hospital Boston and Massachusetts Institute of Technology (CHB-MIT) scalp database from PhysioNet is used to demonstrate the potential effectiveness of modeling multivariate EEG data by assessing the overall variance between sensors for three young patients with intractable seizures by use of a DCC-GARCH model, Bayesian regime switching mixture model, and a Bayesian change-point model.

ACKNOWLEDGMENTS

I would like to thank my wonderful advisor Dr. Gavino Puggioni for all his time and patience over the past years. Without his insights this research wouldn't have been possible. Thanks also to my committee members Dr. Natallia Katenka and Dr. Kunal Mankodiya for their resourcefulness. Special thanks to Dr. Walter Besio for his EEG knowledge and continual assistance. Thanks to my Behavioral Science advisor Dr. Lisa Harlow for her years of support and for being my Statistics defense chair. Last, but certainly not least, thank you to Dr. Liliana Gonzalez for accepting me into the Statistics department and for her invaluable input and support.

TABLE OF CONTENTS

CHAPTER

APPENDIX

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

1.1 Seizure Background

Epilepsy is a prevalent disease, with the World Health Organization (WHO) estimating its toll to be around 50 million individuals [1]. A seizure itself is classified as overabundance of neuronal electrical activity in the brain resulting in moderate to severe shaking along with the potential loss of consciousness and excrement control [1]. The obvious impact that a seizure can have on an individual is not only one of physical concern, but also that of mental state and well-being. While individuals can have independent seizures at one point and never have them again, patients with epilepsy are those that suffer more than just one seizure in their lifetime, and in most cases more than one in a day itself when not properly treated [1].

Intractable seizures can occur in patients with epilepsy, meaning they do not respond to medication based treatment. It is estimated that about 70% of patients with epilepsy respond to medication, however this also means approximately 3 in 10 patients do not respond [1]. Surgical options are then considered, with implanted sensors in the brain to determine the actual area of seizure activity in the brain to be resected [2].

Commonly, scalp electroencephalogram (EEG) is used to measure the electrical brain activity at various areas of the brain simultaneously. A common EEG recording of many sensors simultaneously is displayed in Figure 1.1 [3, 4]. Recent work has extended the use of EEG recordings to ambulatory monitoring, which allows patients to go about their daily routines without being wired to a machine [2, 4, 5, 6]. Real time monitoring of EEG allows for a much larger amount of available data as patients can wear the ambulatory headsets throughout the day doing their normal routines. Post-hoc time series analysis of the ambulatory EEG in recent years has led to many alternative approaches to the analysis of seizures and their onset [7].

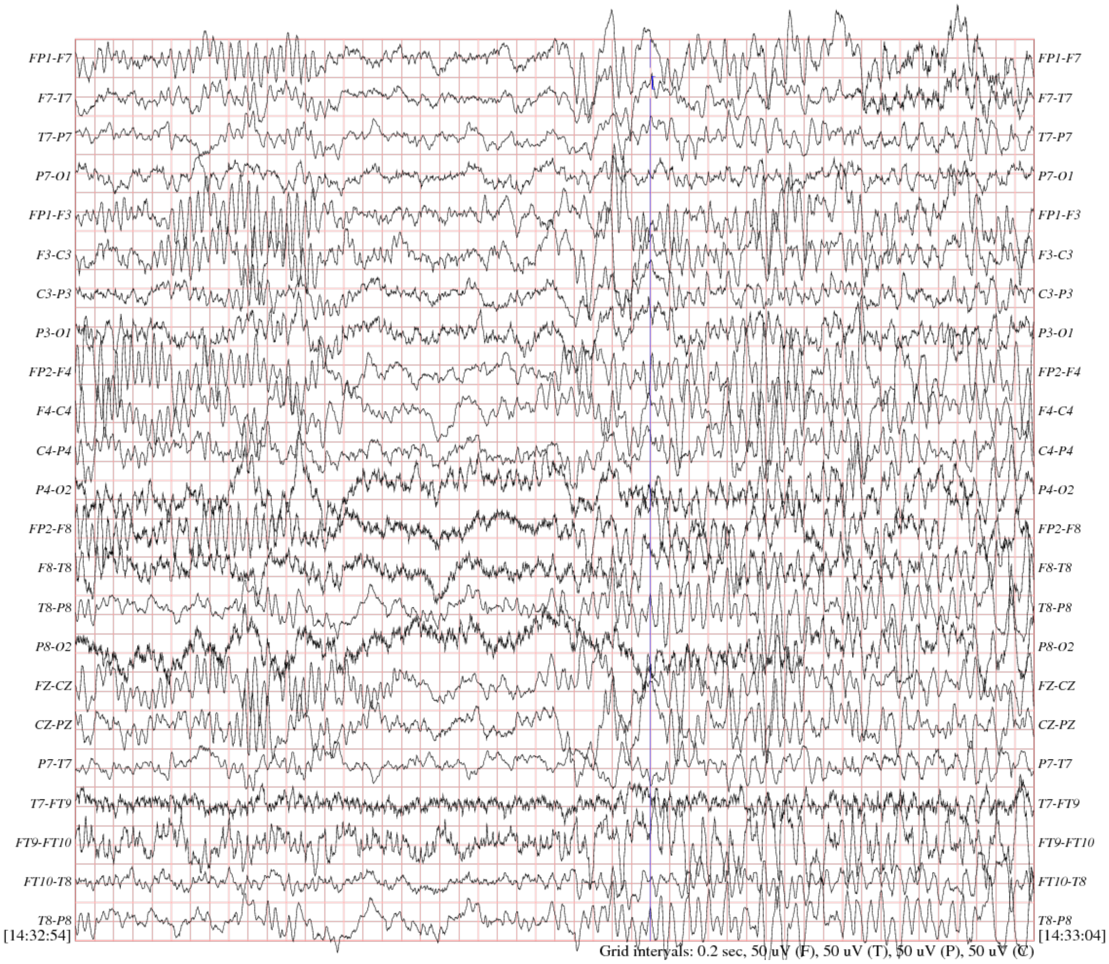

Figure 1.1. A sample EEG of multiple sensors recorded consecutively over the same time interval.

The use of battery powered ambulatory devices for monitoring EEG as it occurs or in a post-hoc fashion allows for a much richer and more complex amount of data [6]. One common study for example typically uses 22 sensors recording simultaneously with a resolution of $256\,Hz$ for one hour, a total of $921,600$ data-points for each of the 22 sensors is available for analysis, or $20,275,200$ data-points for one hour [3, 4]. Supervised multivariate machine learning algorithms have dominated current research in seizure analysis, however the use of power required for a real-time analysis needs to also be considered [5, 6, 7]. Bayesian modeling approaches have also been applied to EEG recordings with focus on modeling the EEG behavior versus computer extracted decisions based on maximum likelihood estimation [8, 9].

For consistency, monitoring of EEG recordings are referenced by scalp locations as defined by the international 10-20 system of electrode placement [10, 11, 12]. Figure 1.2 highlights the areas located within the 10-10 system. It is worth noting that not all areas are necessarily included, as EEG electrode set-ups vary in number and positioning. The 10-20 system is a necessity for consistency in noting electrode placement. For example, the CHB-MIT scalp database utilizes almost two dozen electrodes [3, 4] while other studies have used varying sizes of electrodes [7] and configurations [13].

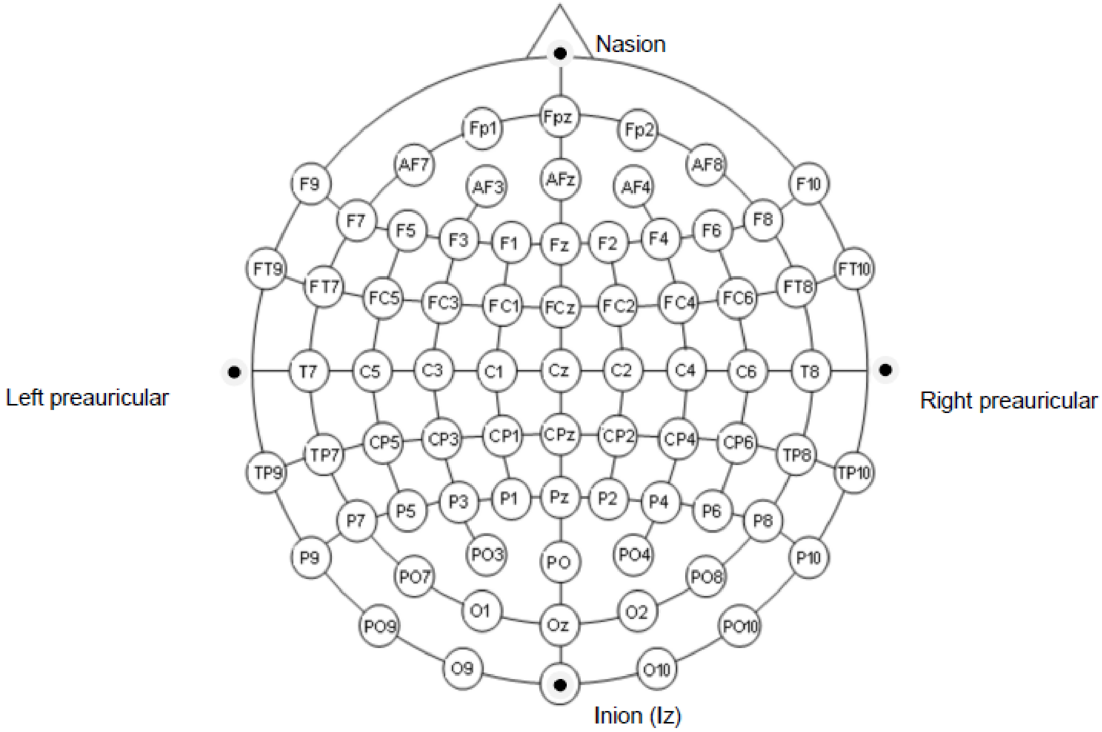

Figure 1.2. The standard EEG sensor areas as defined by the international 10-10 system.

The use of EEG recording allows for not only time series analysis of multivariate data, but also allows for the potential use of spatio-temporal monitoring [8, 13]. Location of seizures is another issue in itself, one of importance, however the actual modeling of seizures is still a developing field. The benefit of the 10-20 system does allow for not only the ability of analysis across time, but the inclusion of spatial location can also be added.

1.2 Brain Synchronization

A common feature of seizures is the desynchronization and resynchronization of the synaptic activity in the brain [14, 15, 16, 17]. Typically, the electrical activity of the brain will show high levels of desynchronization at the onset of a seizure followed by strong resynchronization of the signals to end the seizure [17]. The scalp electrodes of the EEG monitor the synaptic activity in the neurons of the brain with a minimum of 108 neurons and 6 cm^2 necessary to present an EEG sensor response [15]. Some research has suggested a strong synchronization between various areas of the brain potentially hours leading into a seizure, however this seems to only be extrapolated when knowledge of a seizure occurring is also present [16]. Measuring levels of neuronal synchronization can either be measured in regards to a set threshold or across time, which can thus be interpreted as increasing or decreasing trends [17]. Because of the cortical changes in synchronization that are many times present in EEG recordings, it offers a strong potential for seizure location and detection [14].

Typical approaches to addressing brain synchronization have focused on a correlational aspect of two sensors. A common measure is called *mean phase coherence* R , or similarly called *first Fourier mode* and *phase locking value* [14]. Another method is to utilize the correlation matrix of the normalized sensor data at each time point and analyze the eigenvalues of the matrix across time [16]. Other autoregressive measures and transformations have been utilized, but generally are only beneficial when the knowledge of an ictal state occurring is included in a post-hoc analysis [16].

1.3 A Look Forward

Surely, investigation of EEG recordings to help monitor and potentially prevent seizures in epileptic patients has a tremendous potential impact on the health of such patients. Pre-ictal seizure states have been a relative mystery, but current increases in data and computing ability have allowed for a much more thorough focus on seizure prediction. Modeling EEG data offers many benefits, challenges, and possibilities, particularly because of the massive amount of data combined in every patients record. Various multivariate and univariate analytic methods have shown predictive potential, as have frequentist and Bayesian approaches. Melding these ideas into a solid modeling approach that has the potential of low-power usage ambulatory monitoring and care becomes a main goal for any seizure based research.

Use of the Children's Hospital Boston and Massachusetts Institute of Technology (CHB-MIT) scalp database from Physionet [3, 4] is accessed to demonstrate the potential effectiveness of modeling multivariate EEG data by assessing the overall variance between sensors for young patients with intractable seizures. Recent articles using the same data set have addressed the question of real life practical use of EEG analysis and power reduction approaches to make acquisition and possible prevention of seizures more plausible [5, 6]. Several other articles recently have addressed the CHB-MIT dataset through machine learning, with good predictive success but at the cost of lower potential ambulatory usage [2, 18]. What lacks in the research of this dataset is a thorough statistical modeling approach to the CHB-MIT database, particularly with the goal in mind of being less computationally expensive and more practicality based.

By focusing on the idea of brain desynchronization through modeling EEG sensor variance, it is anticipated that predictions will display both a strong sensitivity and specificity to detecting seizures before they actually occur. Time series modeling will be utilized to delineate the processes preceding the seizure onset. Early detection of seizures is useful in the case of patients such as those in the CHB-MIT database, who have not responded to medication and could benefit from the use of implanted devices to counteract seizures before they occur. To do this, natural patterns of the levels of synchrony in the brain will be analyzed by assessing the variance within 22 scalp EEG sensors, followed with appropriate statistical modeling.

This thesis follows with three more chapters. Chapter 2 addresses modeling approaches to EEG data through multivariate dynamic conditional correlation generalized autoregressive conditional heteroskedasticity (DCC-GARCH) models and Bayesian analysis (Markov regime switching mixture models and change point models). Chapter 3 utilizes a DCC-GARCH model on three CHB-MIT patients and then addresses a new approach to modeling brain synchronization through the use of the variance between sensors at every time-point, followed by the application of a Markov switching model with a mixture of normal densities and a change point model for all three patients. Chapter 4 serves as a discussion of the three various models utilized in this thesis in terms of the three subjects analyzed for each. Future goals and directions based on this research are also presented in Chapter 4.

List of References

[1] World Health Organization, January 2015. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs999/en/

[2] Y. U. Khan, O. Farooq, and P. Sharma, "Automatic detection of seizure onset in pediatric EEG," *International Joural of Embeded Systems and Applications*, vol. 2, no. 3, pp. 81–89, 2012.

[3] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[4] A. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment," Ph.D. dissertation, Massachusetts Institute of Technology, September 2009.

[5] K. H. Lee and N. Verma, “A low-power processor with configurable embedded machine-learning accelerators for high-order and adaptive analysis of medical-sensor signals,” *Solid-State Circuits, IEEE Journal of*, vol. 48, no. 7, pp. 1625–1637, 2013.

[6] J. Chiang and R. K. Ward, “Energy-efficient data reduction techniques for wireless seizure detection systems,” *Sensors*, vol. 14, no. 2, pp. 2036–2051, 2014.

[7] L. Orosco, A. G. Correa, and E. Laciar, “Review: A survey of performance and techniques for automatic epilepsy detection,” *Journal of Medical and Biological Engineering*, vol. 33, no. 6, pp. 526–537, 2013.

[8] R. Prado and M. West, *Time series: modeling, computation, and inference.* CRC Press, 2010.

[9] D. F. Wulsin, “Bayesian nonparametric modeling of epileptic events,” Ph.D. dissertation, University of Pennsylvania, 2013.

[10] F. Sharbrough, G. Chatrian, R. Lesser, H. Lüders, M. Nuwer, and T. Picton, “American electroencephalographic society guidelines for standard electrode position nomenclature,” *J. Clin. Neurophysiol*, vol. 8, no. 2, pp. 200–202, 1991.

[11] R. T. Pivik, R. J. Broughton, R. Coppola, R. J. Davidson, N. Fox, and M. R. Nuwer, “Guidelines for the recording and quantitative analysis of electroencephalographic activity in research contexts,” *Psychophysiology*, vol. 30, no. 6, pp. 547–558, 1993.

[12] R. L. Gilmore, “American-electroencephalographic-society guidelines in electroencephalography, evoked-potentials, and polysomnography,” *Journal of Clinical Neurophysiology*, vol. 11, no. 1, pp. 1–142, 1994.

[13] W. G. Besio, K. Koka, R. Aakula, and W. Dai, “Tri-polar concentric ring electrode development for laplacian electroencephalography,” *Biomedical Engineering, IEEE Transactions on*, vol. 53, no. 5, pp. 926–933, 2006.

[14] F. Mormann, T. Kreuz, R. G. Andrzejak, P. David, K. Lehnertz, and C. E. Elger, “Epileptic seizures are preceded by a decrease in synchronization,” *Epilepsy research*, vol. 53, no. 3, pp. 173–185, 2003.

[15] P. Olejniczak, “Neurophysiologic basis of EEG,” *Journal of clinical neurophysiology*, vol. 23, no. 3, pp. 186–189, 2006.

[16] P. R. Carney, S. Myers, and J. D. Geyer, “Seizure prediction: methods,” *Epilepsy & Behavior*, vol. 22, pp. S94–S101, 2011.

[17] P. Jiruska, M. de Curtis, J. G. Jefferys, C. A. Schevon, S. J. Schiff, and K. Schindler, “Synchronization and desynchronization in epilepsy: controversies and hypotheses,” *The Journal of physiology*, vol. 591, no. 4, pp. 787–797, 2013.

[18] H. Khammari and A. Anwar, “A spectral based forecasting tool of epileptic seizures,” *IJCSI International Journal of Computer*, 2012.

CHAPTER 2

Statistical Modeling Approaches

2.1 DCC-GARCH Modeling

A very practical approach to modeling multivariate time-series data with volatility can be achieved by use of a dynamic conditional correlation generalized autoregressive conditional heteroskedasticity (DCC-GARCH) model. When the number of time points gets rather large such as a typical EEG dataset, the DCC-GARCH models can be rather efficient at multivariate modeling of the various sensors. While typically used in financial data applications, the GARCH models have been applied to EEG wavelets in several studies but not in a multivariate DCC-GARCH approach [1, 2, 3, 4].

2.1.1 ARCH & GARCH Models

The use of multivariate measures with a large amount of data points can be tedious, however the dynamic conditional correlation generalized autoregressive conditional heteroscedasticity (DCC-GARCH) model allows for a stream-lined approach to the task. Initially, the autoregressive conditional heteroskedasticity (ARCH) model was introduced to deal with volatile financial data [5]. The ARCH(p) model relates a model with mean zero to it's autoregressive volatility when the assumption of normality is met s.t.:

$$y_t \,|\, \phi_{(t-1)} \sim \mathrm{N}\left(0,\, h_t\right)$$

$$h_t = \alpha_0 + \alpha_i y^2_{(t-1)}$$

where ϕ_t is the information available up until time t and h_t represents the variance function of the data, which can also be stated in terms of p autoregressive parameters estimated as α:

$$h_t = h\left(y_{(t-1)}, y_{(t-2)}, \ldots, y_{(t-p)}, \alpha\right)$$

with p ARCH model parameters are estimated using maximum likelihood, which results in the values for α that best optimize the model.

The ARCH model was further extended to the generalized autoregressive conditional heteroskedasticity (GARCH) a few years later, which allowed for not only autoregressive parameters p, but also moving average parameters q in an ARMA modeling approach for the error variance [6]. The stochastic GARCH model also assumes normality and is specified as:

$$
\begin{aligned}
y_t \,|\, \phi_{(t-1)} &\sim \mathrm{N}\,(0,\, h_t) \\
h_t &= \alpha_0 + \sum_{i=1}^{q} \alpha_i y^2_{(t-1)} + \sum_{(i=1)}^{p} \beta_i h_{(t-i)} \\
&= \alpha_0 + A(L) y_t^2 + B(L) h_t
\end{aligned}
$$

where:

$$
\begin{aligned}
&p \geq 0, \qquad q > 0 \\
&\alpha_0 > 0, \qquad \alpha_i \geq 0, \qquad i = 1, \ldots, q \\
&\beta_i \geq 0, \qquad i = 1, \ldots, p.
\end{aligned}
$$

The GARCH(p, q) model can also be expressed in another variation as follows [6]:

$$
\begin{aligned}
y_t^2 &= \alpha_0 + \sum_{i=1}^{q} \alpha_i y^2_{(t-1)} + \sum_{(j=1)}^{p} \beta_j y^2_{(t-j)} - \sum_{j=1}^{p} \beta_j \nu_{(t-j)} + \nu_t \\
\nu_t &= y_t^2 - h_t \\
&= \left(\eta_t^2 - 1\right) h_t
\end{aligned}
$$

where $\eta_t \sim \mathrm{N}(0,1)$ and ν_t is uncorrelated across time with a mean of zero. The latter GARCH(p,q) expressions allow for a time series ARMA model for y_t^2 with orders of $m = \max(p,q)$ and p.

2.2 Multivariate DCC-GARCH Models

A multivariate version of the GARCH model relies on the conditional correlation of the matrix of values correspondent with time. The DCC-GARCH model [7] focuses on the time-varying covariance matrix H_t such that:

$$H_t \;=\; D_t R D_t$$

where $D_t = \mathrm{diag}\left(\sqrt{h_{i,t}}\right)$ and h represents the univariate GARCH models. R is the conditional correlation matrix:

$$R \;=\; E_{t-1}(\epsilon_t \epsilon_t') = D_t^{-1} H_t D_t^{-1}$$

where $\epsilon_t = D_t^{-1} r_t$ and ϵ represents standard normal disturbances. In the DCC model, the correlation matrix R is allowed to vary with time and conditional variances must summate to unity [7]. The correlation matrix can be represented as a GARCH(1,1) model [7, 8] as:

$$
\begin{aligned}
q_{i,j,t} &= \bar{\rho}_{i,j} + \alpha(\epsilon_{i,(t-1)}\epsilon_{j,(t-1)} - \bar{\rho}_{i,j}) + \beta(q_{i,j,(t-1)} - \hat{\rho}_{i,j}) \\
&= \bar{\rho}_{i,j}\left(\frac{1-\alpha-\beta}{1-\beta}\right) + \alpha\sum_{s=1}^{\infty}\beta^{s}\epsilon_{i,(t-s)}\epsilon_{j,(t-s)}
\end{aligned}
$$

where the expected cross-product is $\bar{\rho}_{i,j}$ and the variances summate to 1. The correlation can be estimated as:

$$
\rho_{i,j,t} = \frac{q_{i,j,t}}{\sqrt{q_{i,i,t}q_{j,j,t}}}
$$

and the expectation is positive definite since the covariance matrix Q_t is a weighted average of positive definite and semidefinite matrices. A requirement of $\alpha + \beta < 1$ for the model to not become explosive and reman stationary must be met [7, 8]. The DCC estimation is computed through a maximum likelihood procedure.

The extension of the DCC-GARCH by Engle [7] helps to attain a more parsimonious (yet still very high) number of parameters. The multivariate GARCH is an extensive of the univariate GARCH proposed by Bollerslev [6]. Use of autoregressive (AR) and moving average (MA) parameters can also be included to the univariate measures to help increase multivariate model fit, however adding even more parameters to the model.

2.3 Basics of Bayesian Inference

Often in Bayesian modeling we need to simulate from the posterior distributions. In contrast to frequentist statistics, Bayesian statistics focuses on probabilities with given information and prior beliefs that can either be strong when evidence supports it, or weak when there isn't as much convincing evidence available. Use of Bayes' rule can be applied to updating probabilities of interest with the addition of new information [9]. Iterative simulation methods are used to approximate the joint posterior distribution by drawing samples from the full conditional distributions via processes such as a Gibbs sampler [9].

2.3.1 Inference of the Conditional Posterior Mean

Bayesian analysis can be utilized to determine posterior probability densities by use of Monte Carlo simulations which sufficiently approximate the actual posterior density when enough consecutive iterations (i) of sampling provide convergence. For example, estimation of a parameter θ in a sample of $y_1, \ldots, y_n$ from the distribution $\pi\left(y_1, \ldots, y_n \,|\, \theta\right)$ can be achieved via Markov Chain Monte Carlo (MCMC) methods [9]. For independent iterations i of sampling, the posterior distribution for θ is simulated s.t.:

$$\theta^{(1)}, \ldots, \theta^{(i)} \sim \text{i.i.d. } \pi\left(\theta \,|\, y_1, \ldots, y_n\right)$$

where increased values of i give a better approximation of the true posterior distribution [9]. Posterior joint inference of the mean when conditioned on the variance can be estimated for a normally distributed model with mean θ and variance σ^2 as follows when $(y_1, \ldots, y_n \,|\, \theta, \sigma^2) \sim$ i.i.d. $\mathrm{N}(\theta, \sigma^2)$:

$$
\begin{aligned}
\pi\left(y_1,\ldots,y_n \,|\, \theta,\, \sigma^2\right) &= \prod_{i=1}^{n} \pi\left(y_i \,|\, \theta,\, \sigma^2\right) \\
&= \prod_{i=1}^{n} \frac{\exp\left[-\frac{1}{2}\left(\frac{y_i-\theta}{\sigma}\right)^2\right]}{\sqrt{2\pi\sigma^2}} \\
&= \left(2\pi\sigma^2\right)^{(-n/2)} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{y_i-\theta}{\sigma}\right)^2\right] \\
&= \left(2\pi\sigma^2\right)^{(-n/2)} \exp\left[\frac{1}{\sigma^2}\sum_{i=1}^{n} y_i^2 - \frac{2\theta}{\sigma^2}\sum_{i=1}^{n} y_i + \frac{n\theta^2}{\sigma^2}\right].
\end{aligned}
$$

From the above equation derives a sufficient statistic based on $\left(\sum y_i^2, \sum y_i\right)$ which can also be transposed to $\left(\bar{y}, s^2\right)$ when we know $\bar{y} = \sum y_i/n$ and $s^2 = \sum \left(y_i - \bar{y}\right)^2 / \left(n-1\right)$ [9]. Posterior inference for θ conditioned on the known value of σ^2 with a conjugate prior for θ will result in:

$$
\begin{aligned}
\pi\left(\theta \,|\, y_1,\ldots,y_n,\, \sigma^2\right) &\propto \pi\left(\theta \,|\, \sigma^2\right) \times \exp\left[-\frac{1}{2\sigma^2}\sum\left(y_i-\theta\right)^2\right] \\
&\propto \pi\left(\theta \,|\, \sigma^2\right) \times \exp\left[c_1\left(\theta - c_2\right)^2\right]
\end{aligned}
$$

when the terms of $\exp\left[c_1\left(\theta-c_2\right)^2\right]$ are normally distributed we can then surmise that $\pi\left(\theta \,|\, \sigma^2\right)$ is an appropriate conjugate prior [9]. If $\theta \sim \mathrm{N}\left(\mu_0,\, \tau_0^2\right)$ then:

$$\begin{aligned}
\pi\left(\theta \,|\, y_1,\ldots,y_n,\, \sigma^2\right) &= \pi\left(\theta \,|\, \sigma^2\right)\left(\frac{\pi\left(y_1,\ldots,y_n \,|\, \theta,\, \sigma^2\right)}{\pi\left(y_1,\ldots,y_n \,|\, \sigma^2\right)}\right) \\
&\propto \pi\left(\theta \,|\, \sigma^2\right) \times \pi\left(y_1,\ldots,y_n \,|\, \theta,\, \sigma^2\right) \\
&\propto \exp\left[-\frac{1}{2\tau^0}{}^{2}\left(\theta-\mu_0\right)^2\right] \times \exp\left[-\frac{1}{2\sigma^2}\sum\left(y_i-\theta\right)^2\right] \\
&\propto \exp\left[\frac{1}{\tau_0^2}\left(\theta^2-2\theta\mu_0+\mu_0^2\right)+\frac{1}{2\sigma^2}\left(\sum y_i^2-2\theta\sum y_i+n\theta^2\right)\right] \\
&= a\theta^2-2b\theta+c
\end{aligned}$$

when the following are substituted:

$$\begin{aligned}
a &= \frac{1}{\tau_0^2}+\frac{n}{\sigma^2} \\
b &= \frac{\mu_0}{\tau_0^2}+\frac{\sum y_i}{\sigma^2} \\
c &= c\left(\mu_0,\, \tau_0^2,\, \sigma^2,\, y_1,\ldots,y_n\right).
\end{aligned}$$

From this a normal density can be extrapolated for $\pi\left(\theta \,|\, \sigma^2, y_1, \ldots, y_n\right)$ s.t.:

$$\begin{aligned}
\pi\left(\theta \,|\, \sigma^2, y_1, \ldots, y_n\right) &\propto \exp\left[-\frac{1}{2}\left(a\theta^2 - 2b\theta\right)\right] \\
&= \exp\left[-\frac{a}{2}\left(\theta^2 - \frac{2b\theta}{a} + \frac{b^2}{a^2}\right) + \frac{b^2}{2a}\right] \\
&\propto \exp\left[-\frac{a}{2}\left(\theta - \frac{b}{a}\right)^2\right] \\
&= \exp\left[-\frac{1}{2}\left(\frac{\theta - \frac{b}{a}}{\frac{1}{\sqrt{a}}}\right)^2\right]
\end{aligned}$$

where $1/\sqrt{a}$ represents the standard deviation and b/a the mean in terms of a normal distribution, and thus $\pi\left(\theta \,|\, \sigma^2,\, y_1, \ldots, y_n\right)$ is normally distributed where:

$$\begin{aligned}
\tau_n^2 &= \frac{1}{a} \\
&= \frac{1}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}} \\
\mu_n &= \frac{b}{a} \\
&= \frac{\frac{\mu_0}{\tau_0^2} + \frac{n\bar{y}}{\sigma^2}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}.
\end{aligned}$$

The data and prior information are combined to form the joint posterior precision s.t.:

$$\frac{1}{\tau_n^2} = \frac{1}{\tau_0^2} + \frac{n}{\sigma^2}$$

Similarly the posterior mean can be determined from the amalgamation of the data and the prior:

$$\mu_n = \frac{\frac{1}{\tau_0^2}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}\mu_0 + \frac{\frac{n}{\sigma^2}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}\bar{y}.$$

If there are κ_0 prior observations, the prior precision could be adjusted to $\tau_n^2 = \sigma^2/\kappa_0$. Based on this, the posterior mean would then be:

$$\mu_n = \frac{\kappa_0\mu_0}{\kappa_0 + n} + \frac{n\bar{y}}{\kappa_0 + n}.$$

Prediction of a new point $\tilde{Y}$ from previous observations $y_1, \ldots, y_n$ based on the normal posterior distribution thus follows:

$$
\begin{aligned}
\mathrm{E}\left[\tilde{Y}\,|\,y_1,\ldots,y_n,\,\sigma^2\right] &= \mathrm{E}\left[\theta+\tilde{\epsilon}\,|\,y_1,\ldots,y_n,\,\sigma^2\right] \\
&= \mathrm{E}\left[\theta\,|\,y_1,\ldots,y_n,\,\sigma^2\right] + \mathrm{E}\left[\tilde{\epsilon}\,|\,y_1,\ldots,y_n,\,\sigma^2\right] \\
&= \mu_n + 0 \\
&= \mu_n \\
\mathrm{Var}\left[\tilde{Y}\,|\,y_1,\ldots,y_n,\,\sigma^2\right] &= \mathrm{Var}\left[\theta+\tilde{\epsilon}\,|\,y_1,\ldots,y_n,\,\sigma^2\right] \\
&= \mathrm{Var}\left[\theta\,|\,y_1,\ldots,y_n,\,\sigma^2\right] + \mathrm{Var}\left[\tilde{\epsilon}\,|\,y_1,\ldots\ y_n,\,\sigma^2\right] \\
&= \tau_n^2 + \sigma^2.
\end{aligned}
$$

Since θ and $\tilde{\epsilon}$ are conditional on the normally distributed data in $y_1,\ldots,y_n$, it follows then that:

$$
\left(\tilde{Y}\,|\,\sigma^2,\,y_1,\ldots,y_n\right) \sim \mathrm{N}\left(\mu_n,\,\tau_n^2+\sigma^2\right)
$$

2.3.2 Joint Conditional Inference of Mean and Variance

In its simplest form, the Bayes rule for the posterior of the joint distribution of θ and σ^2 states that:

$$
\pi\left(\theta,\,\sigma^2\,|\,y_1,\ldots,y_n\right) = \frac{\pi\left(y_1,\ldots,y_n\,|\,\theta,\,\sigma^2\right)\pi\left(\theta,\,\sigma^2\right)}{\pi\left(y_1,\ldots,y_n\right)}.
$$

The axioms of probability for a joint distribution also state:

$$\pi\left(\theta,\, \sigma^2\right) \quad = \quad \pi\left(\theta \,|\, \sigma^2\right) \pi\left(\sigma^2\right)$$

when σ^2 is known and $\tau_0^2 = \sigma^2/\kappa_0$ where μ_0 and κ_0 are the mean and sample size from the prior observations, we can then state:

$$\pi\left(\theta \,|\, \sigma^2\right) \pi\left(\sigma^2\right) \quad = \quad \mathrm{dnorm}\left(\theta,\, \mu_0,\, \tau_0 = \sigma/\sqrt{\kappa_0}\right) \times \pi\left(\sigma^2\right).$$

Use of the gamma distribution is common for the prior of the precision, or $1/\sigma^2$. When expressed in terms of the prior precision certainty ν_0:

$$\frac{1}{\sigma^2} \quad \sim \quad \mathrm{G}\left(\frac{\nu_0}{2},\, \frac{\nu_0 \sigma_0^2}{2}\right)$$

The joint posterior distribution can also be represented by the axioms of probability s.t.:

$$\pi\left(\theta,\, \sigma^2 \,|\, y_1, \ldots, y_n\right) \quad = \quad \pi\left(\theta \,|\, \sigma^2,\, y_1, \ldots, y_n\right) \pi\left(\sigma^2 \,|\, y_1, \ldots, y_n\right).$$

When supplied with data $y_1, \ldots, y_n$ and σ^2, the conditional distribution of θ becomes:

$$\left(\theta \,|\, y_1, \ldots, y_n,\, \sigma^2\right) \quad \sim \quad \mathrm{N}\left(\mu_n,\, \frac{\sigma^2}{\kappa_n}\right)$$

when:

$$\begin{aligned}
\kappa_n &= \kappa_0 + n \\
\mu_n &= \frac{\dfrac{\kappa_0\mu_0}{\sigma^2} + \dfrac{n\bar{y}}{\sigma^2}}{\dfrac{\kappa_0}{\sigma^2} + \dfrac{n}{\sigma^2}} \\
&= \frac{\kappa_0\mu_0 + n\bar{y}}{\kappa_n}.
\end{aligned}$$

The posterior of σ^2 can be derived mathematically as:

$$\begin{aligned}
\pi\left(\sigma^2 \,|\, y_1, \ldots, y_n\right) &\propto \pi\left(y_1, \ldots, y_n \,|\, \sigma^2\right) \\
&= \pi\left(\sigma^2\right) \int \pi\left(y_1, \ldots, y_n \,|\, \theta,\, \sigma^2\right) \pi\left(\theta \,|\, \sigma^2\right)\, d\theta.
\end{aligned}$$

If thus follows that:

$$\left(\frac{1}{\sigma^2} \,|\, y_1, \ldots, y_n\right) \sim \mathrm{G}\left(\frac{\nu_n}{2}, \frac{\nu_n\sigma_n^2}{2}\right)$$

where:

$$\begin{aligned}
\nu_n &= \nu_0 + n \\
\sigma_n^2 &= \frac{\nu_0\sigma_0^2 + (n-1)\, s^2 + \dfrac{\kappa_0 n \left(\bar{y} - \mu_0\right)^2}{\kappa_n}}{\nu_n}
\end{aligned}$$

2.3.3 Gibbs Sampling for Bayesian Posterior Inference

Iterative Monte Carlo methods can be used to estimate posterior distributions, with convergence occurring after I total iterations. Gibbs sampling is a commonly used Monte Carlo approach to estimating the full conditional mean θ and precision $1/\sigma^2$ from data $y_1, \ldots, y_n$ and any supplied prior information [9, 10, 11, 12]. Assuming we have parameters ϕ^i at a current iteration i of the Gibbs sampler, we can infer θ conditional on $1/\sigma^2$ and $1/\sigma^2$ conditional on θ as follows:

1. Sample $\theta^{(i+1)} \sim \pi\left(\theta \,|\, \frac{1}{\sigma^{2(i)}},\, y_1, \ldots, y_n\right)$
2. Sample $\frac{1}{\sigma^{2(i+1)}} \sim \pi\left(\frac{1}{\sigma^2} \,|\, \theta^{(i+1)},\, y_1, \ldots, y_n\right)$
3. Set $\phi^{(i+1)} = \left(\theta^{(i+1)},\, \frac{1}{\sigma^{2(i+1)}}\right)$.

The flow of a Gibbs sampler follows a Markov Chain Monte Carlo (MCMC) series of steps where p estimated parameters from ϕ at time i are related only to the parameters at $i-1$ iteratively through:

1. Sample $\phi_1^i \sim \pi\left(\phi_1 \,|\, \phi_2^{i-1},\, \phi_3^{(i-1)}, \ldots, \phi_p^{(i-1)}\right)$
2. Sample $\phi_2^i \sim \pi\left(\phi_2 \,|\, \phi_1^{i-1},\, \phi_3^{(i-1)}, \ldots, \phi_p^{(i-1)}\right)$

   $\vdots$

p. Sample $\phi_p^i \sim \pi\left(\phi_p \,|\, \phi_1^{i-1},\, \phi_2^{(i-1)}, \ldots, \phi_{(p-1)}^{(i-1)}\right)$.

The Gibbs sampling steps produce sequentially dependent vectors which are independent for non-consecutive vectors based on the MCMC properties. The vectors are:

$$\begin{aligned}
\phi^{(1)} &= \left(\phi_1^{(1)}, \ldots, \phi_p^{(1)}\right) \\
\phi^{(2)} &= \left(\phi_1^{(2)}, \ldots, \phi_p^{(2)}\right) \\
&\vdots \\
\phi^{(i)} &= \left(\phi_1^{(i)}, \ldots, \phi_p^{(i)}\right).
\end{aligned}$$

The sampling distribution of $\phi^{(i)}$ becomes closer to the target joint distribution when i gets larger. This can be stated when $i \to \infty$ as:

$$\pi\left(\phi^{(i)} \in A\right) \to \int_A \pi\left(\phi\right) d\phi$$

and can be expanded to a function g as $I \to \infty$:

$$\frac{1}{I}\sum_{i=1}^{I} g\left(\phi^{(i)}\right) \to \mathrm{E}\left[g\left(\phi\right)\right] = \int g\left(\phi\right)\pi\left(\phi\right) d\phi.$$

Thus the mean of the complete set of iterative samples is a good approximation of $\mathrm{E}\left[g\left(\phi\right)\right]$. It is important to note that Gibbs sampling is an approximate description of the posterior distribution, not a form of model or the exact integral [9]. In other words, Gibbs sampling is a MCMC approximate of $\pi\left(\phi \,|\, y_1, \ldots, y_n\right)$ s.t.:

$$\begin{aligned}\frac{1}{I}\sum \phi^{(i)} &\approx \int \phi \, \pi\left(\phi \,|\, y_1, \ldots, y_n\right) \, d\phi \\ \frac{1}{I}\sum \left(\phi^{(i)} \leq c\right) &\approx \pi\left(\phi \leq \left(c \,|\, y_1, \ldots, y_n\right)\right) \\ &= \int_{-\infty}^{c} \pi\left(\phi \,|\, y_1, \ldots, y_n\right) \, d\phi.\end{aligned}$$

2.4 Markov Regime Switching Models

Estimation of the state variables S for all the time points in the series T can be found by using a Markov regime switching model. The state indicators in S are represented by $1, \ldots, K$ different states, the number of which can be determined by previous knowledge and verified by a Random Permutation Sampling (RPS) scheme as highlighted soon.

For the observations $y_1, \ldots, y_n$, a probabilistic observation can be made for the latent state variables in S which depend only on the state specific parameters of θ and σ^2 in $\phi_1^S, \ldots, \phi_K^S$ [13]. Use of the Bayesian Gibbs sampler can iteratively approximate the distributions for ϕ and thus we will also need a prior transition probability ξ. The priors for ϕ must meet the following:

1. Apriori independence of $(\theta_1, \ldots, \theta_K, \sigma_1^2, \ldots, \sigma_K^2)$ and ξ s.t.: $\pi(\phi) = \pi(\theta_1, \ldots, \theta_K, \sigma_1^2, \ldots, \sigma_K^2)\, \pi(\xi)$.
2. $\pi(\theta_1, \ldots, \theta_K, \sigma_1^2, \ldots, \sigma_K^2)$ is unaffected by the use of random indicator permutations s.t.: $\pi(\theta_1, \ldots, \theta_K, \sigma_1^2, \ldots, \sigma_K^2) = \pi\left(\theta_{\rho(1)}, \ldots, \theta_{\rho(K)}, \sigma_{\rho(1)}^2, \ldots, \sigma_{\rho(K)}^2\right)$ for permutations of $(\rho(1), \ldots, \rho(K))$ in $(1, \ldots, K)$.

3. A Markovian prior requires $\pi\left(\xi\right)$ and S to be invariant to random permutations for all K states. Thus, $\pi\left(\xi_{(1,1)},\ldots,\xi_{(1,K)},\ldots,\xi_{(K,1)},\ldots,\xi_{(K,K)}\right) = \pi\left(\xi_{(\rho(1),\rho(1))},\ldots,\xi_{(\rho(1),\rho(K))},\ldots,\xi_{(\rho(K),\rho(1))},\ldots,\xi_{(\rho(K),\rho(K))}\right)$ for permutations of $(\rho(1),\ldots,\rho(K))$ in $(1,\ldots,K)$.

4. The transition probabilities $\xi_1,\ldots,\xi_K$ should be independent a priori s.t:

$$\pi\left(\xi\right) \;=\; \prod_{i=1}^{K}\pi\left(\xi_i\right)$$

5. Finally, if $\forall_{(i,j)} \in (1,\ldots,K,\ i \neq j)$ than:

$$\pi\left(S_{1-s}=i\right) \;=\; \pi\left(S_{1-s}=j\right).$$

A Dirichlet ($\mathcal{D}$) Markovian prior is suggested based on the previous steps [13]. The prior of $\mathcal{D}\left(e_{(0,i1)},\ldots,e_{(0,iK)}\right)$ represents the transition likelihoods between the states S. The following restrictions must be met to ensure a proper Markovian prior:

1. $e_{(0,ii)} \equiv e_p$ where $i=1,\ldots,K$.

2. $e_{(0,ij)} \equiv e_q$, $\forall_{(i,j)} \in \left(1,\ldots,K\right),\, i \neq j$.

3. Prior transition beliefs can allow $e_p > e_q$ if deemed appropriate.

Bayesian posterior analysis for a Markov switching model based on $y_1,\ldots,y_n$ from the unconstrained space $\mathcal{A}$ results in a posterior of $P_{\mathcal{A}}\left(\bullet \,|\, y_1,\ldots,y_n\right)$.

2.4.1 Dirichlet Prior

Markov switching models have been defined to estimate the hidden states $S = \{S_{1:T}\}$ with transition parameters ξ and known state parameters θ [14, 15]. The algorithm estimates ξ and θ conditioned on S and then estimates S conditioned on ξ and θ. Bayesian Gibbs sampling is used to estimate the parameters by repetition of the following steps until the algorithm has sufficiently achieved convergence based on the individual iterations $i = 1 : I$.

1. Sample the value of $\xi^{(i)}$ from $p\left(\xi \,|\, S^{(i-1)}, \mathcal{D}_T\right)$ where $\mathcal{D}_T$ represents the Dirichlet distribution s.t.:

   $$\xi_j \sim \mathcal{D}(\alpha_j)$$

   where $\alpha_j = \left(\alpha_{(j,1)}, \ldots, \alpha_{(j,K)}\right)'$ and $j = 1 : K$. Each i samples ξ from $j = 1 : K$ and $N_{(j,K)}(S)$ represents the number of state transitions from S_j to S_k in S s.t.:

   $$\xi_j \sim \mathcal{D}\left(\alpha_{(j,1)} + N_{(j,1)}(S), \ldots, \alpha_{(j,K)} + N_{(j,K)}(S)\right)$$

2. Sample the value of $\theta^{(i)}$ from $p\left(\theta_{(1:K)} \,|\, S^{(i-1)}, \mathcal{D}_T\right)$.

3. Sample the value of $S^{(i-1)}$ from $\left(S \,|\, \xi^{(i)}, \theta^{(i)}, \mathcal{D}_T\right)$ which is accomplished by a forward filtering backward sampling routine (FFBS) [14, 15]. The algorithm is as follows:

   (a) Evaluate $Pr(S_t = k \,|\, \xi^{(i)}, \theta^{(i)}, \mathcal{D}_t)$ in $t = 1 : T$ where

$$\begin{aligned} Pr(S_t = k \mid \xi^{(i)}, \theta^{(i)}, \mathcal{D}_t) \;\propto\; & p(y_t \mid S_t = k, \mathcal{D}_{(t-1)}, \xi^{(i)}, \theta^{(i)}) \\ & \times Pr(S_t = k \mid \mathcal{D}_{(t-1)}, \xi^{(i)}, \theta^{(i)}), \end{aligned}$$

and sample $S_T^{(i)}$ from $Pr(S_T = k \mid \xi^{(i)}, \theta^{(i)}, \mathcal{D}_T)$;

(b) Sample latent state at time t for values of $t = (T-1) : 1$ from $Pr(S_t = j \mid S_{(t+1)}^{(i)} = k, \xi^{(i)}, \theta^{(i)}, \mathcal{D}_T)$ s.t.:

$$Pr(S_t = j \mid S_{(t+1)} = k, \xi, \theta, \mathcal{D}_T) \;\propto\; Pr(S_t = j \mid S_{(t+1)} = k, \xi, \theta, \mathcal{D}_T)$$

and

$$\begin{aligned} Pr(S_t = j \mid S_{(t+1)} = k, \xi, \theta, \mathcal{D}_t) \;\propto\; & Pr(S_{(t+1)} = k \mid S_t = j, \xi, \theta, \mathcal{D}_t) \\ & \times Pr(S_t = j \mid \xi, \theta, \mathcal{D}_t). \end{aligned}$$

(c) Set $S^{(i)} = \{S_{(1:T)}^{(i)}\}$.

2.4.2 Label Switching Problem

Because the state variable S is dummy coded, label switching poses a potential issue in the Gibbs sampling for Markov switching models. Use of a random permutation sampling (RPS) scheme has been advised to explore the full sampling space and determine if there is adequate separation for K [13, 14]. At each step of the Gibbs sampler, a RPS step can be added as such:

1. Randomly select a random permutation $\rho(1) \ldots \rho(K)$ of S at the current iteration i. Apply the random permutation s.t. $\left(\theta_1^{(i)}, \ldots, \theta_K^{(i)}\right) := \left(\theta_{\rho(1)}^{(i)}, \ldots, \theta_{\rho(K)}^{(i)}\right)$ and that $i^{(N)} = (i_0, i_1, \ldots, i_N) := \left(\rho(\tilde{i}_0), \rho(\tilde{i}_1), \ldots, \rho(\tilde{i}_N)\right)$.
2. Assuming a Markovian prior for ξ on i^N, relabelling can be done s.t. $\xi_i = \left(\xi_{(i1)} \ldots \xi_{(iK)}\right) := \left(\xi_{\rho(i)\rho(1)}, \ldots, \xi_{\rho(i)\rho(K)}\right)$ where $i = 1, \ldots, K$.

2.4.3 Random Permutation Sampling

Following the use of an unconstrained RPS sampling scheme with proper separation of K, a constrained sampling step can be included at each iteration i of the Gibbs sampler. A common approach is to relabel the indicators S with an identifiability constraint s.t. $(\theta_1 < \theta_2 < \ldots < \theta_K)$ [13, 14]. Gibbs sampling with a constraint to prevent label switching can be done through the standard Gibbs sampling procedure including the following at each i:

1. If the identifiability constraint on θ is violated, reorder the values of $(\theta_1, \ldots, \theta_K)$ as ordered pairs $(1, \theta_1), \ldots, (K, \theta_K)$ where the first component of the pairs represents the new labelling of the classes so that the identifiability constraint is now met.

2.5 Bayesian Change Point Modeling

Another Bayesian adapted approach is to that of change point analysis. Simply stated, change-points are those in which the distribution suddenly changes at a given time or multiple times. An example would be the change in state from pre-seizure to seizure and back to post-seizure. Gibbs sampling can again be used to estimate the posterior distributions of the parameters of the model, which now includes the change points [16]. Assume we have a partition $\rho = (i_0, i_1, \ldots, i_b)$ s.t.:

$$0 = i_0 < i_1 < i_2 < \ldots < i_b = n$$

and

$$\theta_i = \theta_{i_r}, \quad i_{r-1} < i \leq i_r$$

for $r = 1, 2, \ldots, b$ with change points value at $i_1 + 1, i_2 + 1, \ldots, i_{b-1} + 1$. Notation of time $i+1, \ldots, j$ can be notated as ij and the observations from the series as Y_{ij} [16]. It is assumed that the points in Y_{ij} are independent and follow a distribution s.t.:

$$Y \sim N\left(\mu_0, \frac{\sigma_0^2}{j-1}\right)$$

and is used to iteratively simulate the density of the partitions.

With a partition of $\rho = (U_1, U_2, \ldots, U_n)$ when the values of U represent the change points of the model, the conditional distributions are computed for every value of i in each partition. A transition probability p can thus iteratively be calculated for every value of i in the data Y s.t.:

$$\begin{aligned} \frac{p_i}{1-p_i} &= \frac{\pi(U_i = 1 \mid Y, U_j, j \neq i)}{\pi(U_i = 0 \mid Y, U_j, j \neq i)} \\ &= \frac{\int_0^\gamma p^b (1-p)^{(n-b-1)}\, dp \left(\int_0^\lambda \frac{w^{b/2}}{(W_1 + B_1 w)^{(n-1)}/2}\, dw \right)}{\int_0^\gamma p^{b-1} (1-p)^{(n-b)}\, dp \left(\int_0^\lambda \frac{w^{(b-1)/2}}{(W_0 + B_0 w)^{(n-1)}/2}\, dw \right)} \end{aligned}$$

where W_0 and W_1 are within-block sums of squares (SS) while B_0 and B_1 represent the between-block SS for blocks $U_i = 0$ and $U_i = 1$ from ρ. The values of γ and λ represent tuning parameters bounded by 0 and 1 to assist in model convergence when necessary [16, 17]. Because of potential divergence within a long series, the probability of a change point at i can conversely be calculated as:

$$\frac{p_i}{1-p_i} = \frac{\pi(U_i = 1 \,|\, Y, U_j, j \neq i)}{\pi(U_i = 0 \,|\, Y, U_j, j \neq i)} = \left(\frac{W_0}{W_1}\right)^{\frac{n-b-2}{2}} \times \left(\frac{B_0}{B_1}\right)^{\frac{b+1}{2}} \times \sqrt{\frac{W_1}{B_1}}$$

$$\times \frac{\int_0^{\frac{B_1\lambda/W_1}{1+B_1\lambda/W_1}} p^{(b+2)/2}(1-p)^{(n-b-3)/2}\,dp}{\int_0^{\frac{B_0\lambda/W_0}{1+B_0\lambda/W_0}} p^{(b+1)/2}(1-p)^{(n-b-2)/2}\,dp} \times \frac{\int_0^{\gamma} p^{b}(1-p)^{(n-b-1)}\,dp}{\int_0^{\gamma} p^{b-1}(1-p)^{(n-b)}\,dp}$$

which are incomplete beta integrals and will be stabile for longer series of observations in Y [16].

The Gibbs sampling routine can be achieved for change point models with finite variables U and V which have a joint density of $f(u,v)$ and conditionals of $f(u \,|\, v)$ and $f(v \,|\, u)$. With starting values of U^0 and V^0 we can simulate new states for U^i and V^i derived from U^{i-1} and V^{i-1} as follows:

$$f(u^i, v^i \,|\, u^{i-1}, v^{i-1}) = f(u^i \,|\, v^{i-1}) f(v^i \,|\, u^i).$$

Assuming the states have a stationary distribution with density $f(u,v)$ it can then be supposed that V^0 has the proper marginal density and U^1 has the proper conditional density given V^0, thus U^1 and V^0 have a proper joint density. It follows that U^1 and V^1 have a proper joint density which mimics U^0 and V^0 and therefore $f(u,v)$ is a stationary distribution and is unique if there is a finite chain of positive probabilities between any two change points. In reference to (u,v) and (u',v'), there must be a positive probability sequence where $u^1 = u$, $v^1 = v$, u^2, $v^2, \ldots, u^n = u'$, and $v^n = v'$ where $f(u^i, v^{i-1}) > 0$ and $f(u^i, v^i) > 0$ [16].

With a unique stationary distribution, the sampled points U^i and V^i only need the density $f(u,v)$. In a model with n variables in $U_1, \ldots, U_n$ where the transition probabilities between the change points are positive, a sample is drawn from $f(u_1, u_2, \ldots, u_n)$ with initial values of $u_1^0, \ldots, u_n^0$ and following transitions one at a time conditioned on the other variables. In general, U_k^i is sampled from the density $f(u_k \,|\, u_1^i, \ldots, u_{k+1}^{i-1}, \ldots, u_n^{i-1})$ [16].

In terms of partitions between change points with n observations, we can set $U_i = 1$ when $\theta_i = \theta_{i+1}$. It thus follows that $U_i = 0$ when $\theta_i \neq \theta_{i+1}$ for $1 \leq i < n$. We are left with a transition U_i of 0 or 1 based on the probabilistic nature of the partitions $f(U_1, \ldots, U_{i-1}, 1, U_{i+1}, \ldots, U_{n-1})$ and $f(U_1, \ldots, U_{i-1}, 0, U_{i+1}, \ldots, U_{n-1})$. [16].

The Bayes estimate is used to assess the change point probabilities in a Bayesian framework. A sweep through the data in Y iteratively draws a value from the distribution U_i for every i in the sweep given the data and the values of U_j, $j \neq i$. With b blocks and $i \geq 2$ and $U_i = 0$, it follows that:

$$\frac{P(U_i = 1 \,|\, Y, U_j, j \neq i)}{P(U_i = 0 \,|\, Y, U_j, j \neq i)} = \frac{\int_0^{p_0} p^b (1-p)^{n-b-1}\, dp \left(\int_0^{w_0} \frac{w^{b/2}}{(W_1 + B_1 w)^{(n-1)/2}}\, dw \right)}{\int_0^{p_0} p^{b-1} (1-p)^{n-b}\, dp \left(\int_0^{w_0} \frac{w^{(b-1)/2}}{(W_0 + B_0 w)^{(n-1)/2}}\, dw \right)}$$

where W_0 and B_0 are the within and between block SS respectively when $U_i = 0$. In a similar manner, W_1 and B_1 represent the corresponding SS when $U_i = 1$. We can then calculate $f(U_i = 1 \,|\, Y, U_j, j \neq i)$ and sample U_i [16].

For each sweep of the observations the posterior mean (μ_r) can be determined from the partition ρ and Y when $r \in ij \in \rho$ s.t.:

$$\hat{\mu}_r = (1 - w)\bar{Y}_{ij} + w\mu_0$$

where

$$w = \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2}.$$

For M sweeps through the data, the mean of M estimates is a valid approximation of μ_i when given the data in Y. The similar idea is applied to calculate the posterior mean of σ^2 for M passes through the data. It has been suggested that $50 \leq M \leq 500$ gives a fair approximation of the posterior of μ and σ^2 [16].

Bayesian change point analysis has been utilized in biostatistics to model zero-mean heteroskedastic data such as that in Electromyography (EMG) data [18]. The use of change point analysis in EMG seems to logically suggest its application in EEG data might also be warranted, especially with the change of seizure and non-seizure states.

List of References

[1] A. Galka, O. Yamashita, and T. Ozaki, "Garch modelling of covariance in dynamical estimation of inverse solutions," *Physics Letters A*, vol. 333, no. 3, pp. 261–268, 2004.

[2] K. F. K. Wong, A. Galka, O. Yamashita, and T. Ozaki, "Modelling non-stationary variance in EEG time series by state space GARCH model," *Computers in biology and medicine*, vol. 36, no. 12, pp. 1327–1335, 2006.

[3] A. Galka, K. Wong, and T. Ozaki, "Generalized state-space models for modeling nonstationary EEG time-series," in *Modeling Phase Transitions in the Brain*. Springer, 2010, pp. 27–52.

[4] S. Mihandoost, M. C. Amirani, and B. Z. Varghahan, "A new approach for feature extraction of EEG signal using GARCH variance series," in *Application of Information and Communication Technologies (AICT), 2011 5th International Conference on*. IEEE, 2011, pp. 1–5.

[5] R. F. Engle, "Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation," *Econometrica: Journal of the Econometric Society*, pp. 987–1007, 1982.

[6] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *Journal of econometrics*, vol. 31, no. 3, pp. 307–327, 1986.

[7] R. Engle, "Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models," *Journal of Business & Economic Statistics*, vol. 20, no. 3, pp. 339–350, 2002.

[8] M. Billio, M. Caporin, and M. Gobbo, "Flexible dynamic conditional correlation multivariate GARCH models for asset allocation," *Applied Financial Economics Letters*, vol. 2, no. 02, pp. 123–130, 2006.

[9] P. D. Hoff, *A first course in Bayesian statistical methods.* Springer, 2009.

[10] G. Casella and E. I. George, "Explaining the gibbs sampler," *The American Statistician*, vol. 46, no. 3, pp. 167–174, 1992.

[11] J. Diebolt and C. P. Robert, "Estimation of finite mixture distributions through bayesian sampling," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 363–375, 1994.

[12] S. Chib, "Calculating posterior distributions and modal estimates in markov mixture models," *Journal of Econometrics*, vol. 75, no. 1, pp. 79–97, 1996.

[13] S. Frühwirth-Schnatter, "Markov chain monte carlo estimation of classical and dynamic switching and mixture models," *Journal of the American Statistical Association*, vol. 96, no. 453, pp. 194–209, 2001.

[14] S. Frühwirth-Schnatter, *Finite Mixture and Markov Switching Models: Modeling and Applications to Random Processes.* Springer, 2006.

[15] R. Prado and M. West, *Time series: modeling, computation, and inference.* CRC Press, 2010.

[16] D. Barry and J. A. Hartigan, "A bayesian analysis for change point problems," *Journal of the American Statistical Association*, vol. 88, no. 421, pp. 309–319, 1993.

[17] C. Erdman, J. W. Emerson, *et al.*, "bcp: an r package for performing a bayesian analysis of change point problems," *Journal of Statistical Software*, vol. 23, no. 3, pp. 1–13, 2007.

[18] T. D. Johnson, R. M. Elashoff, and S. J. Harkema, "A bayesian change-point analysis of electromyographic data: detecting muscle activation patterns and associated applications," *Biostatistics*, vol. 4, no. 1, pp. 143–164, 2003.

CHAPTER 3

Modeling EEG Data

3.1 CHB-MIT Scalp Database

To demonstrate the use of a DCC-GARCH model and to be followed by modeling of the univariate measure of sensor variance, the CHB-MIT scalp database is accessed from PhysioNet [1, 2]. Recent articles using the same data set have addressed the question of real-life ambulatory use of EEG analysis and power reduction approaches to make acquisition and possible early detection of seizures more plausible [3, 4]. Several other articles recently have addressed the CHB-MIT dataset through machine learning, with good sensitivity and specificity but at the cost of lower potential ambulatory usage [5, 6]. What lacks in the research of this dataset is a thorough statistical modeling approach to the CHB-MIT database, particularly with the goal in mind of being less computationally expensive and more practicality based.

Three select patients and clips from the CHB-MIT scalp database are used for demonstration and analysis going forward (`chb05.13`, `chb08.02`, and `chb22.38`). All three of the clips are one hour in duration and follow the International 10-20 system for EEG sensor locations. A total of 23 sensors were used for all three patients presented, however only 22 of the channels are unique as one pair of sensors is duplicated (T8-P8) for referencing and thus analysis will utilize 22 sensors for each clip. The resolution for each clip is 256 *Hz*, or 256 samples per second. This results in a total of $921,600$ total samples for each sensor over the course of an hour, multiplied by 22 sensors to warrant $20,275,200$ total samples for just a single one hour clip. A tremendous computational burden exists with such large data, and thus the use of a univariate summary measure of the data certainly has its benefits. Because of the high resolution and autocorrelation of so many data points, one point per second will be utilized accordingly, spaced equally 256 points apart from start to end, leaving 3,600 total points in each clip to sequentially assess sensor variance as well as for use in DCC-GARCH modeling. Figure 3.1 displays the effect of decimating the full 921,600 points for one channel (FP1-F7) to 3,600 from clip `chb08.02`.

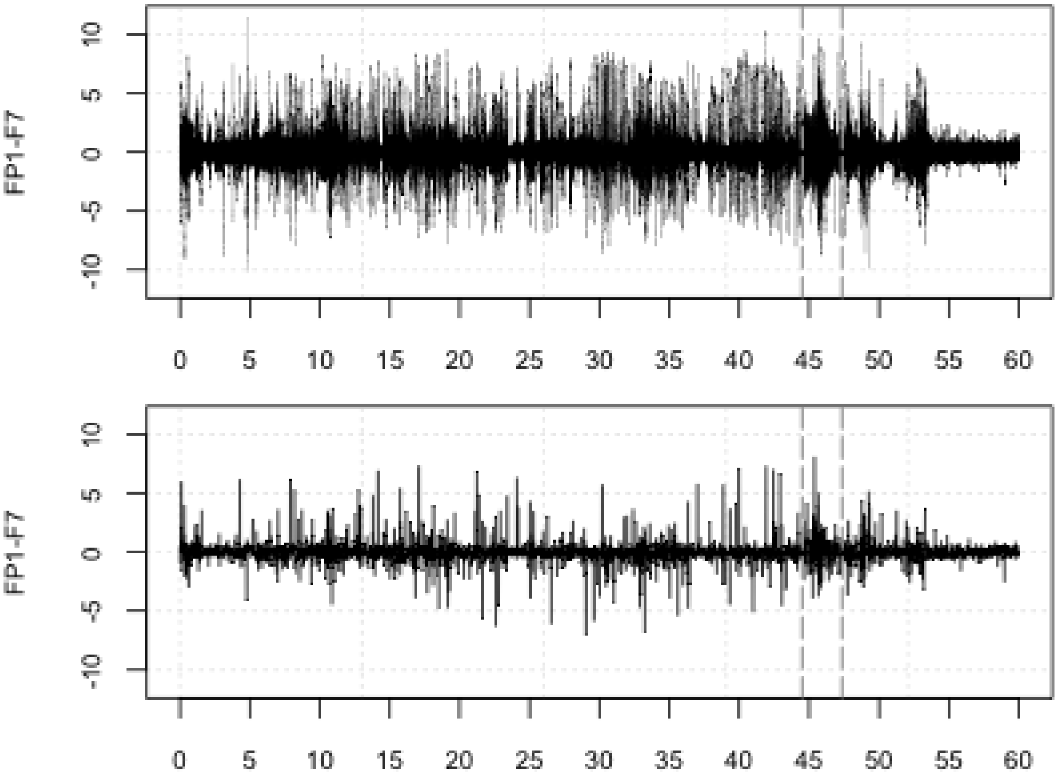

Figure 3.1. Plot of one individual sensor (FP1-F7) for clip `chb08.02`. The entire collection of 921,600 sequential points are plotted (top) as well as a decimated version with only one point per 256, equivalent to one time point per second (bottom).

Patients `chb05`, `chb08`, and `chb22`) were ages 7, 3.5, and 9 respectively at the time of data collection. Patients `chb05` and `chb22` are both female whereas `chb08` is male. The ages of these individuals seem beneficial in that they are much less likely to have seizures influenced by other sources such as alcohol, tobacco, or drugs as an older teenager or young adult might potentially have. The CHB-MIT database is very large and rather complex with hundreds more seizure and non-seizure clips available for future analysis. Some clips are longer than one hour and others contain multiple seizures within one clip. The three clips chosen for this analysis are all analogous in the sensors used, the fact that only one seizure occurs during a one hour clip, and the seizure locations aren't too close to the beginning or end of the clips.

Analysis of the individual sensors is commonplace in EEG research, as well as assessing differences between various spatial areas of the brain. Use of the international 10-20 EEG sensor placement in the CHB-MIT scalp dataset allows for varied sensors that can highlight changes in brain synchronization at the onset of a seizure. Thus use of multivariate measures of the sensors is beneficial, and will be examined more in depth with the use of DCC-GARCH soon. Figure 3.2 displays three individual sensors in the rostral area of the brain that go from left to right from clip `chb08.02`.

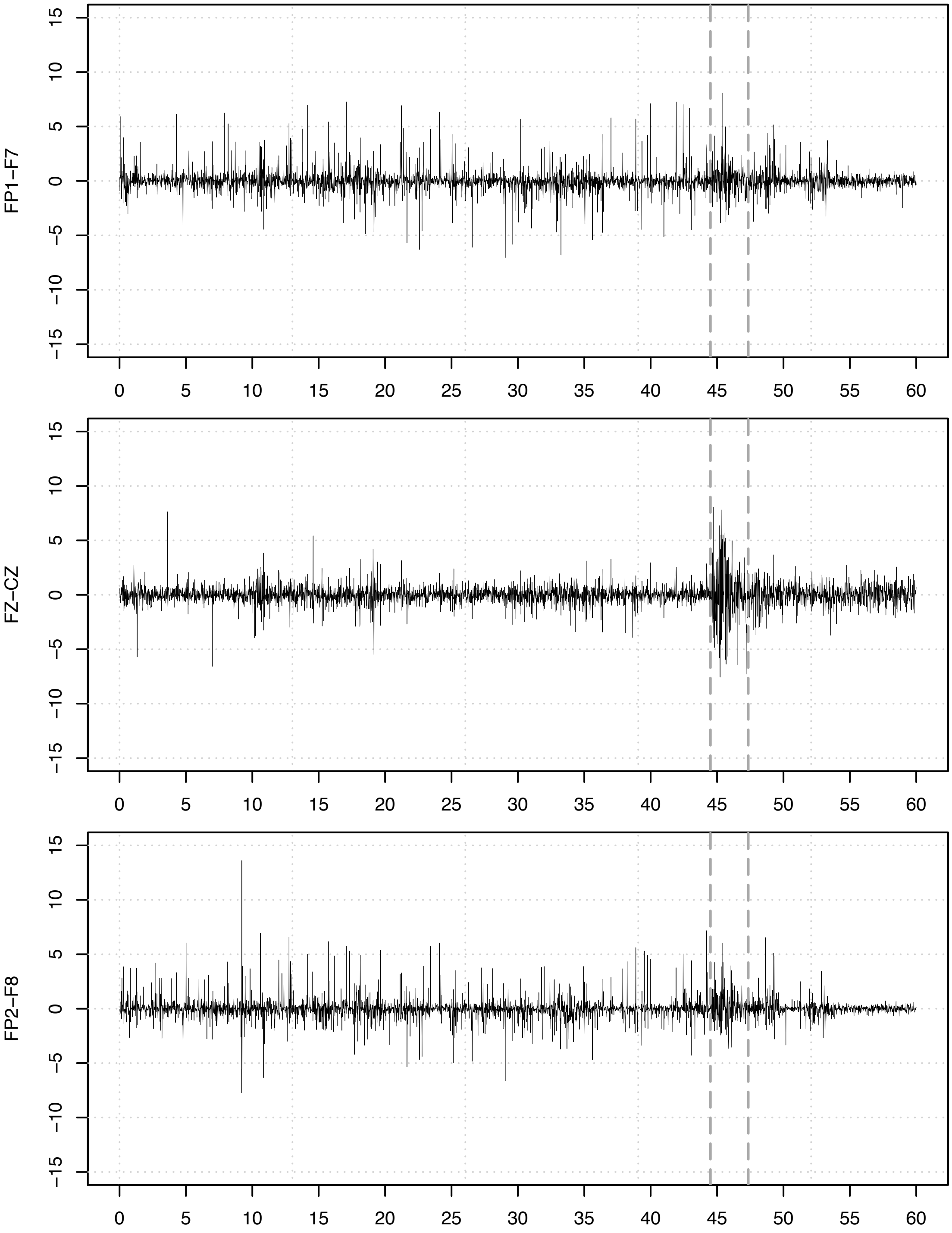

Figure 3.2. Three individual EEG sensor voltages in the rostral area of the brain for a 60 minute segment. In particular, the sensors displayed are located at FP1-F7 (top), FZ-CZ (middle), and FP2-F8 (bottom) based on the internal 10-20 system. The clips displayed from top to bottom are representative of going from the left hemisphere across the medial and to the right hemisphere respectively. Seizure locating is bounded by dashed lines.

The sensors displayed in Figure 3.2 are located at FP1-F7, FZ-CZ, and FP2-F8 in the international 10-20 system and go from left to right in ascending order respectively. Similarly Figure 3.3 shows three sensors from the clip `chb08` in the caudal area of the brain again from left to right in ascending order located at sensor areas P7-O1, CZ-PZ, and FP2-F8. Seizure location is also highlighted for all three sensors for the 60 minute clips.

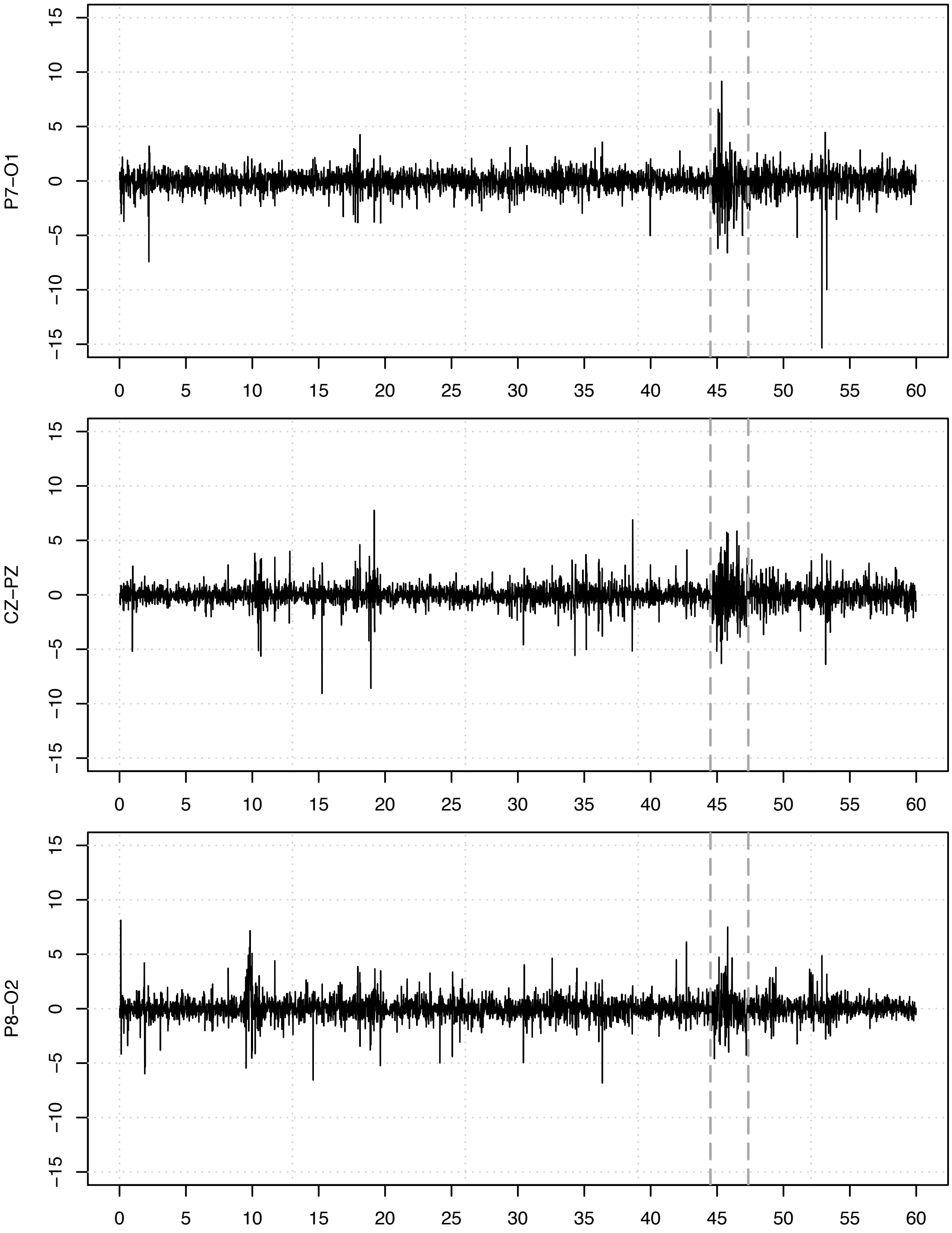

Figure 3.3. Three individual EEG sensor voltages in the caudal area of the brain for a 60 minute segment. In particular, the sensors displayed are located at P7-O1 (top), CZ-PZ (middle), and FP2-F8 (bottom) based on the internal 10-20 system. The clips displayed from top to bottom are representative of going from the left hemisphere across the medial and to the right hemisphere respectively. Seizure locating is bounded by dashed lines.

Initial plots for clip `chb05.13` in Figure 3.4 highlight the variance of the 22 sensors at each time point, or one point per second and $3,600$ in total for the hour. Figure 3.4 also displays the logarithm of the variance of the sensors, which converts the heavily truncated at 0 volatility displayed in the top plot to a Gaussian normal distribution. Notable in both the top and middle plots of Figure 3.4 is the persistence of the EEG variance during the seizure event compared to the non-seizure events. The lower plot in Figure 3.4 shows the composite density of all $3,600$ points along with the non-seizure and seizure densities divided by the number of points they represent in the observations. Because of the small numbers of actual seizure points in comparison to the non-seizure points, it seems to be only a small change in the overall density, however there is clearly a mixture of two densities occurring when looking at the overall composite density.

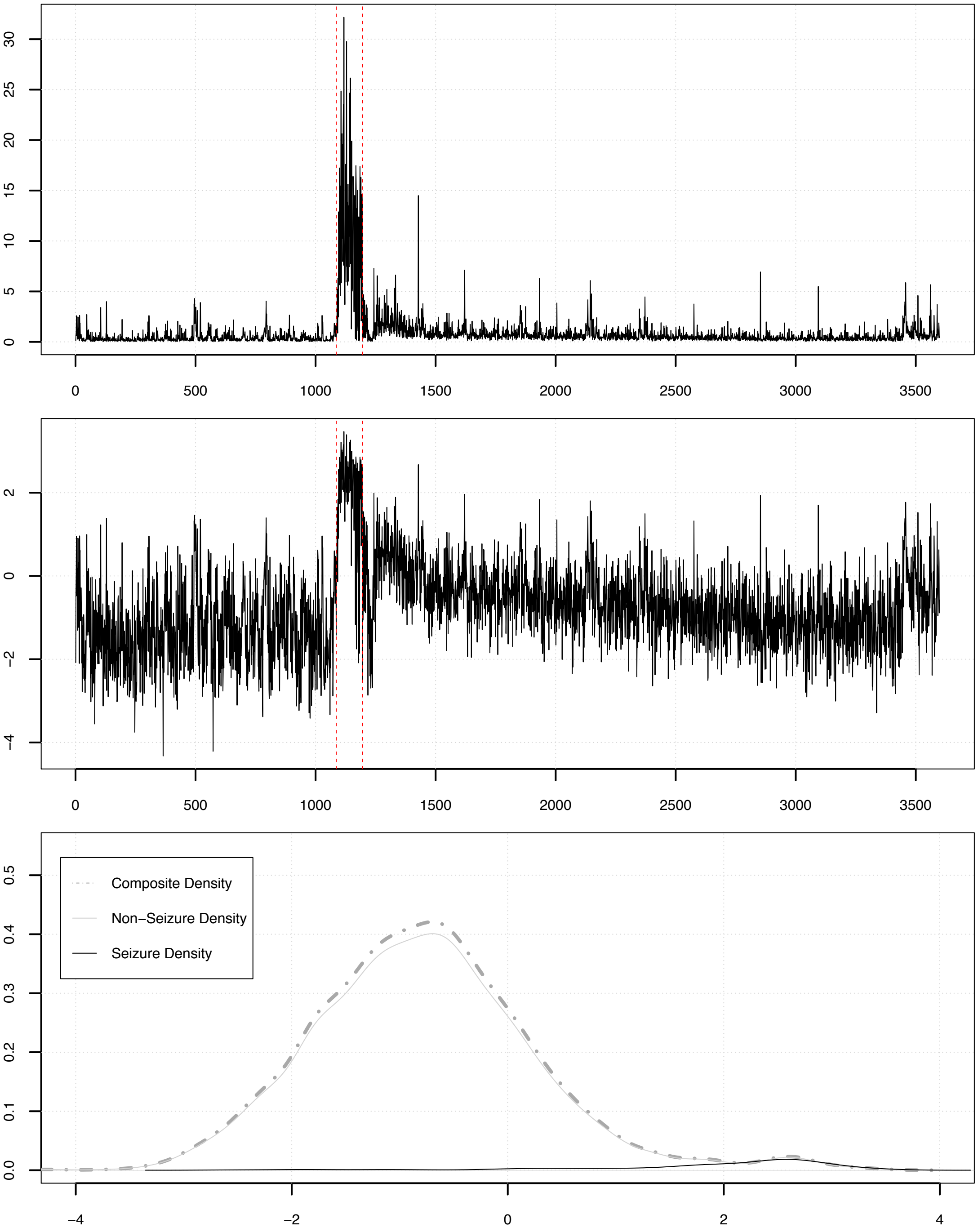

Figure 3.4. Initial plots for clip `chb05.13`. The variance of all 22 sensors at each time point (top), and logarithm of the variance (middle) are displayed with dashed vertical lines representing the seizure location boundaries. The densities (bottom) for the clip are segmented by non-seizure, seizure, and composite where the densities are divided by the total number of points they represent in the clip.

In a similar manner, Figure 3.5 displays the variance of the sensors, the logarithm of the sensor variance, and the densities for clip `chb08.02`. Figure 3.6 presents the same items for clip `chb22.38`. All three of these clips seem to suggest a mixture of normal densities, where the seizure data points follow a different mean and standard deviation than the non-seizure points. This knowledge will be helpful as we start modeling mixtures later in the chapter. The sensor variance can also be predicted by a DCC-GARCH model and will help serve as a validation for the univariate measure of sensor variance derived from the multivariate EEG data available for each clip. This potential univariate measure of the sensor variance will be discussed soon, and offers a potentially strong way to efficiently and computationally downsize EEG data while still keeping the overall integrity of the data.

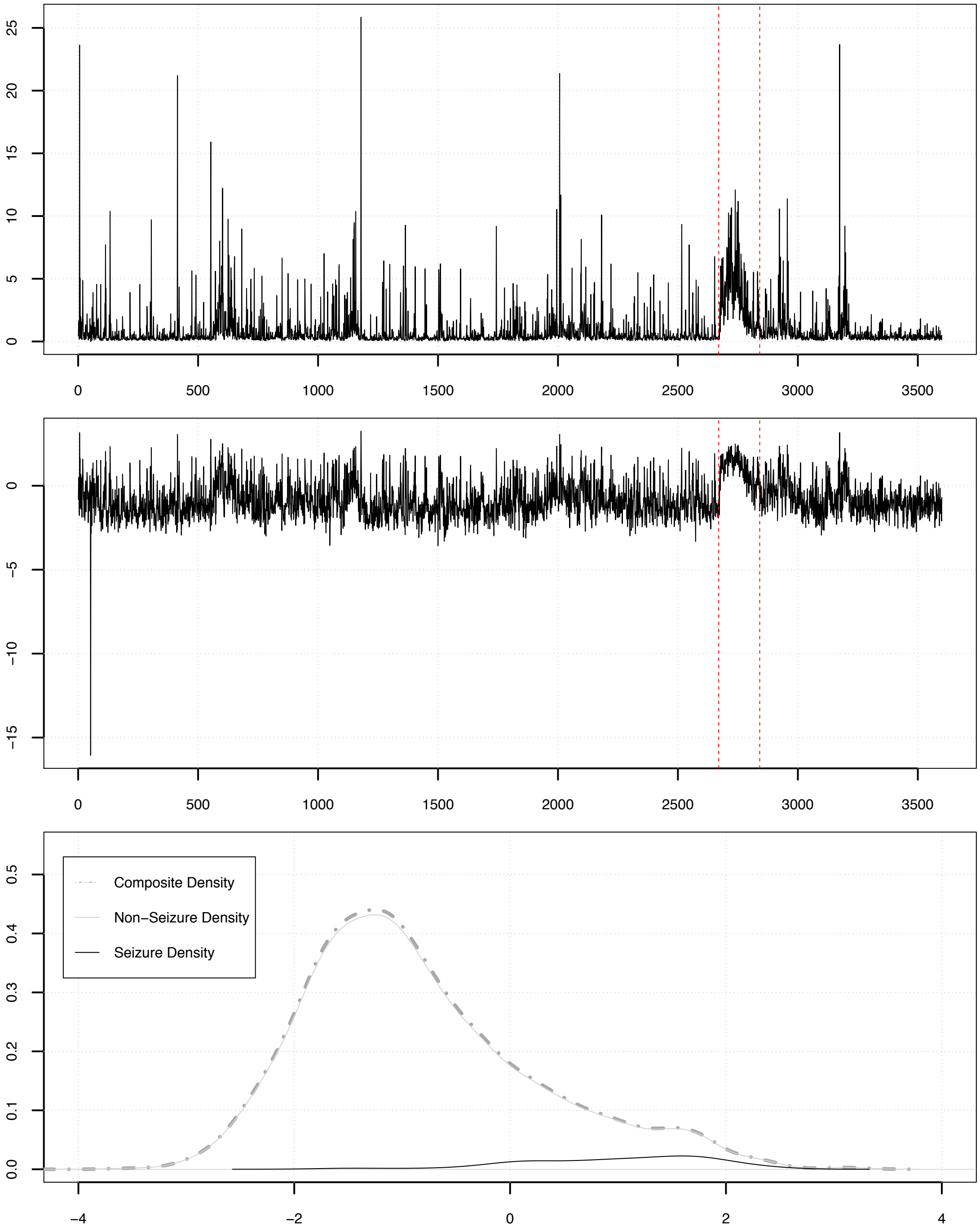

Figure 3.5. Initial plots for clip `chb08.02`. The variance of all 22 sensors (top) at each time point, and logarithm of the variance (middle) are displayed with dashed vertical lines representing the seizure location boundaries. The densities (bottom) for the clip are segmented by non-seizure, seizure, and composite where the densities are divided by the total number of points they represent in the clip.

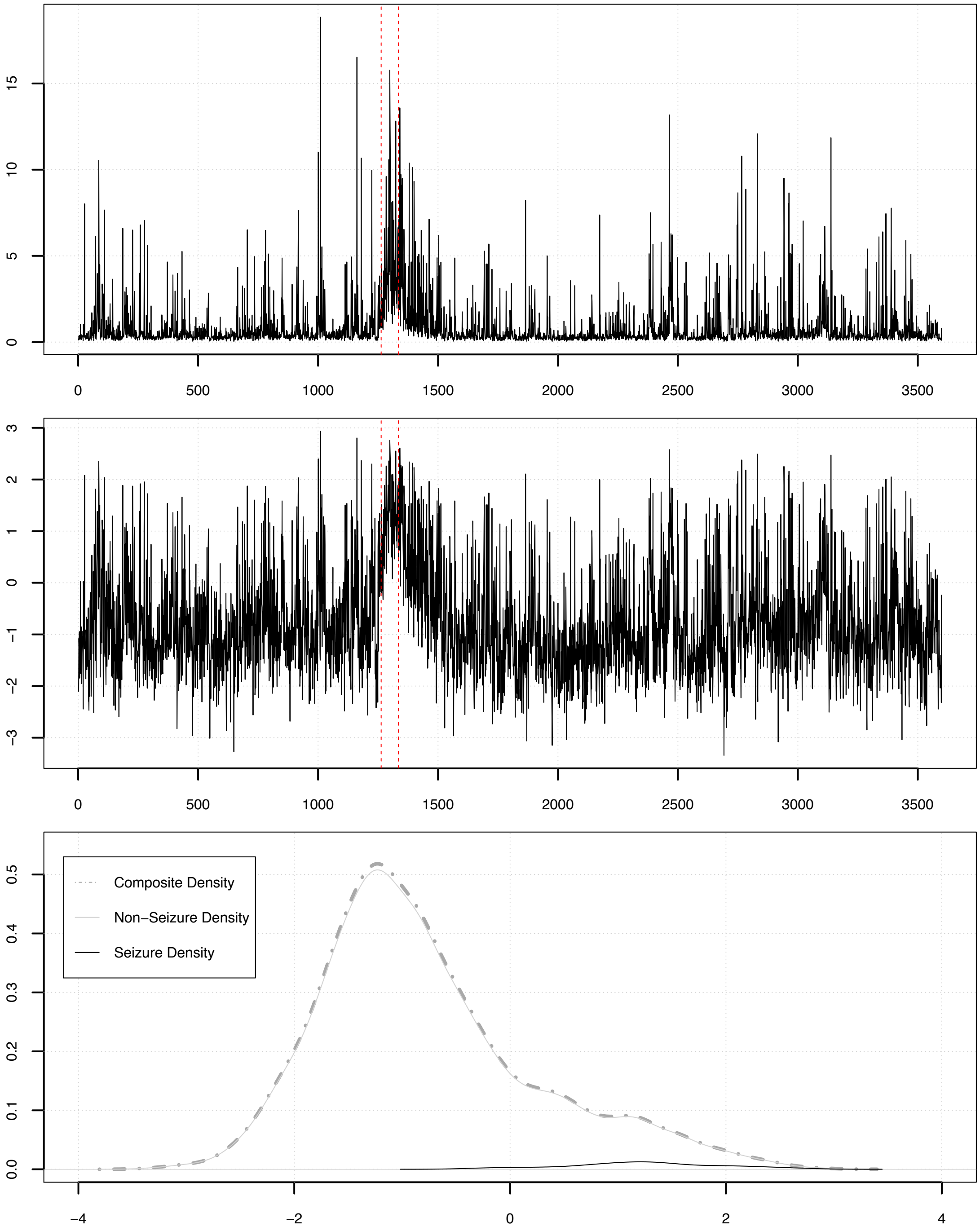

Figure 3.6. Initial plots for clip `chb22.38`. The variance of all 22 sensors (top) at each time point, and logarithm of the variance (middle) are displayed with dashed vertical lines representing the seizure location boundaries. The densities (bottom) for the clip are segmented by non-seizure, seizure, and composite where the densities are divided by the total number of points they represent in the clip.

3.2 DCC-GARCH Modeling

A relatively unused approach to modeling EEG sensor data is the use of a DCC-GARCH model. The multivariate DCC-GARCH can not only serve as a method all in its own, but can also help serve as validation for the univariate modeling of EEG sensor variance with a Markov regime switching model and change point models as demonstrated later in this chapter. The DCC-GARCH allows for a relatively efficient assessment of very large complex multivariate data sets such as that in the CHB-MIT scalp database. It is worth noting that the DCC-GARCH model might not be as beneficial as the univariate measure presented earlier simply because of the computational power still required and the large amount of estimated model parameters. However, in a post-hoc fashion, the DCC-GARCH serves as a viable approach to modeling volatility in EEG data, particularly with the idea of brain desynchronization in mind. Changes in the overall correlation and covariance structure of the data should also be reflected in the DCC-GARCH model analysis.

The DCC-GARCH in this research will assess all 22 sensors during each one hour clip, with a one point per second specification as will be used in the forthcoming Bayesian models. The major difference is that the DCC-GARCH will not use the sensor variance at one measure per second, but will utilize the actual standardized EEG recordings. Thus there will be 22 sensors modeled at a total of $3,600$ data points, or correspondingly one point for every second.

Specification for DCC-GARCH models allows for varied parameters of not only the ARMA(p, q) parameters but also for the GARCH(p, q) model parameters. Use of the Akaike information criterion (AIC) and Bayesian Information Criterion (BIC) values will be used for the models to determine the best model fit along with the most parsimonious fit. Typically, GARCH$(1, 1)$ models fit volatile data best, so the GARCH parameters will be set at $(1, 1)$, however the autoregressive and moving average time series parameters will be varied to find the best model fit. Therefore each clip analyzed will be run with a GARCH$(1, 1)$ along with a combination of four alternative models of ARMA$(0, 0)$, ARMA$(0, 1)$, ARMA$(1, 0)$, and ARMA$(1, 1)$. Because of the very high number of parameters required for a DCC-GARCH model with a data matrix of $3,600$ rows by 22 columns, it is unlikely that ARMA specifications greater than 1 will actually be better fitting that models with 0 or 1 simply because of parsimony.

DCC-GARCH Analysis for the three clips presented above will look at the individual model parameters, the predicted model based on the parameters, the corresponding model fit in terms of standardized mean squared error (MSE), and the predicted model correlation and covariance matrices in comparison to the actual matrices.

Use of the `rmgarch` package in `R` was used to fit the four various DCC-GARCH models for each of the three clips to be examined throughout this thesis. Use of only one point per second, or $3,600$ time points for 22 sensors, still requires heavy computational usage and thus it would be ill-advised to use the entire $921,600$ points for each sensor contained in the clips. There is also heavy autocorrelation because of the use of 256 points per second, and thus downsampling the data seems to be a potential remedy.

The various model specifications for the DCC-GARCH are highlighted in Table 3.1. Both `chb08.02` and `chb22.38` were modeled most efficiently with an ARMA(0,0) and GARCH(1,1), however clip `chb05.13` performed better with an ARMA(1,0) and GARCH(1,1) based on AIC and BIC values. It is of note that the ARMA(0,0)-GARCH(1,1) DCC model for `chb05.13` was not too much higher on the AIC and BIC values than the ARMA(1,0)-GARCH(1,1) model, suggesting that either model could potentially be used. Note that another 22 parameters are required when using an ARMA(1,0) specification versus an ARMA(0,0) and thus parsimony could still side with an ARMA(0,0) model. For demonstration purposes, an ARMA(0,0)-GARCH(1,1) model will be utilized for `chb08.02` and `chb22.38` and an ARMA(1,0)-GARCH(1,1) will be used for `chb05.13`.

Clip	ARMA	GARCH	AIC	BIC
`chb05_13`	(0,0)	(1,1)	8.635	9.187
	(0,1)	(1,1)	14.864	15.453
	(1,0)	(1,1)	7.879	8.469
	(1,1)	(1,1)	14.928	15.556
`chb08_02`	(0,0)	(1,1)	13.502	14.054
	(0,1)	(1,1)	14.233	14.822
	(1,0)	(1,1)	15.060	15.649
	(1,1)	(1,1)	17.127	17.755
`chb22_38`	(0,0)	(1,1)	12.435	12.986
	(0,1)	(1,1)	15.432	16.022
	(1,0)	(1,1)	13.158	13.748
	(1,1)	(1,1)	17.093	17.721

Table 3.1. AIC and BIC model comparison for varied specifications for the three clips analyzed in this research. The AIC and BIC values are divided by the total of points used as specified in the package used.

Model diagnostics are presented in Figure 3.7 for `chb08.02`. The standardized mean square error (MSE) values for each of the 22 sensors fall well below 0.10 and are adequate according to typical standards (see Figure 3.7). The predicted values for the sensors are also standardized by sensor and the variance amongst the sensors is then computed and plotted against the actual sensor variance (see Figure 3.7) as seen originally in Figure 3.4. The residuals from the actual and predicted sensor variance are also displayed in Figure 3.7, with very few large deviations from 0 suggesting a relatively good model fit overall.

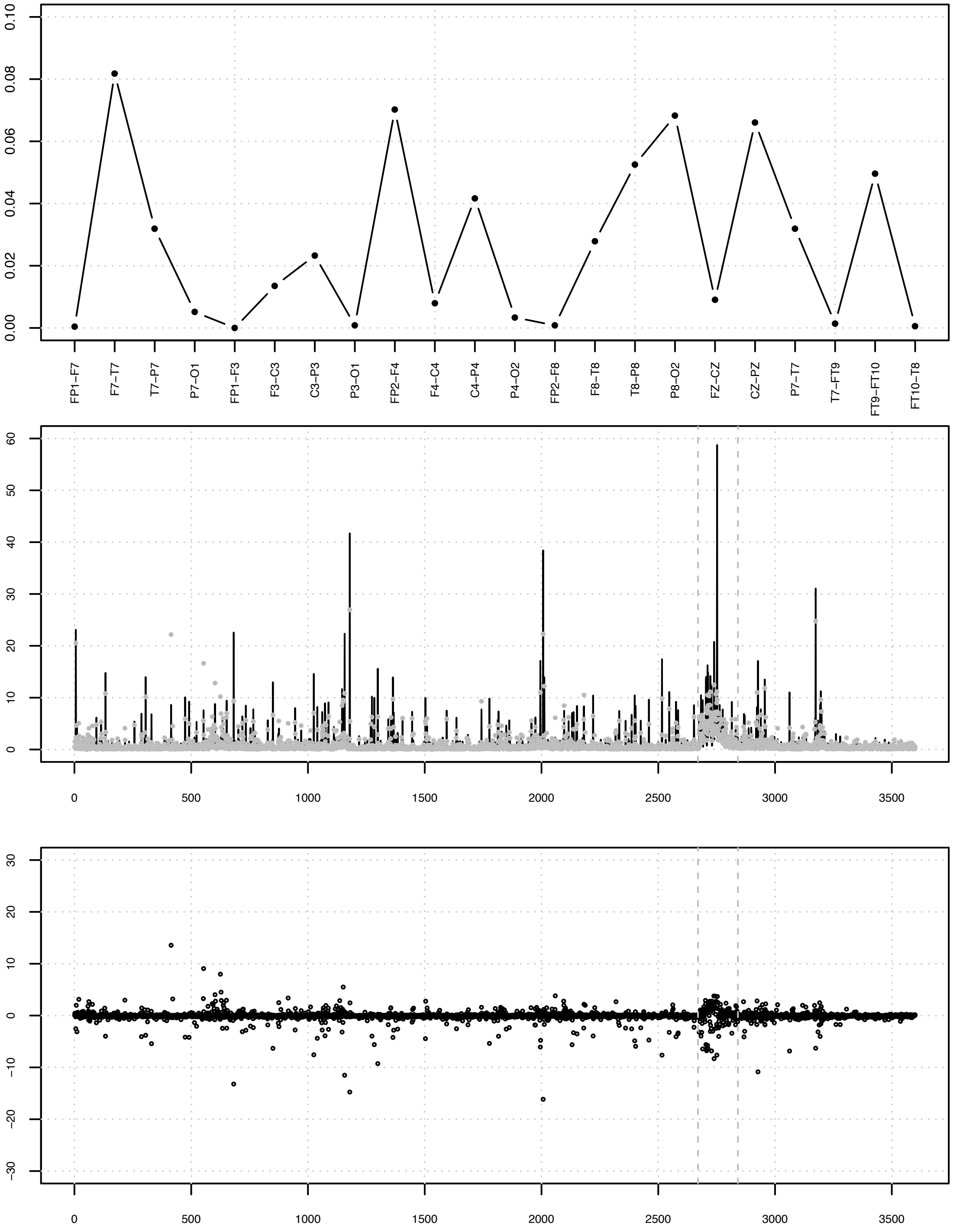

Figure 3.7. Standardized MSE values from the DCC-GARCH model for all 22 of the individual sensors for clip `chb08.02` (top). Plot of the actual sensor variance in black and the predicted sensor variance in grey circles from an ARMA(0,0)-GARCH(1,1) DCC model (middle). Residuals of the actual sensor variance compared to the predicted sensor variance from those displayed in the middle panel (bottom). Dashed grey lines represent seizure boundaries.

The covariance matrices of the actual and predicted model can be plotted using a heat-map image to help visualize the matrices for `chb08.02` (see Figure 3.8), with brighter colors closer to white representing higher covariance values in the matrices. To visually aid the difference between the actual and predicted covariance matrices for the sensor variance, the absolute value of the difference between the matrices is also presented in Figure 3.8. Ideally, we'd like to see darker colors in the absolute difference of the covariances, since the model should be a relatively close predictor of the actual covariance of the 22 sensors and we'd expect values closer to 0, which are represented as dark colors in the image plots.

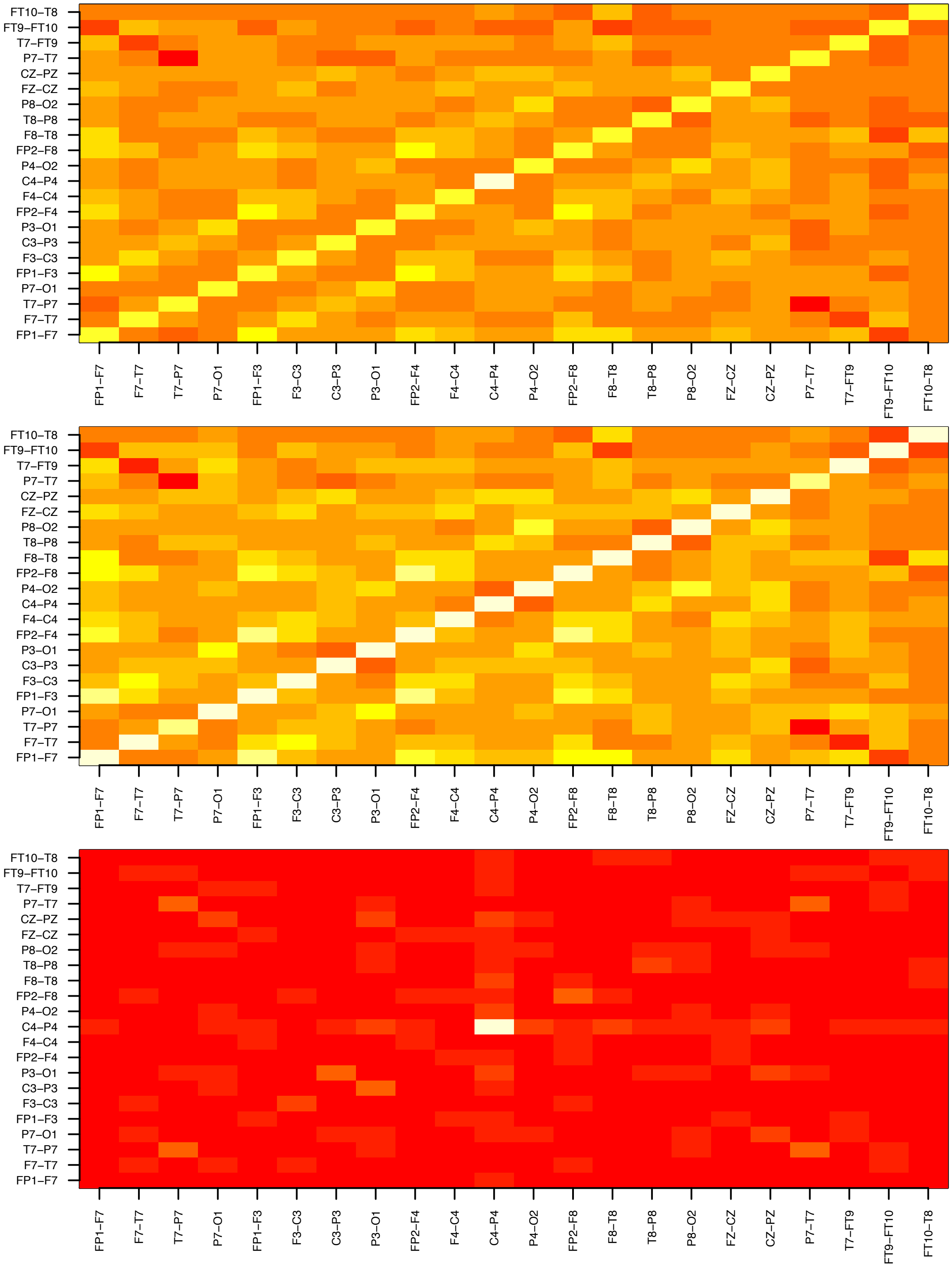

Figure 3.8. Heat-map images of the covariance matrices for the DCC-GARCH predicted covariance matrix (top), actual covariance matrix of the data (middle), and the absolute value of the difference between the predicted and actual covariance matrices (bottom) for clip `chb08.02`.

A total of 321 estimated parameters are necessary for the model used with `chb08.02`. The estimated values of the sensor means (μ) are presented in Table 3.2. The values for ω (see Table 3.3), α (see Table 3.4), and β (see Table 3.5) are also presented. The parameter estimates are also presented with standard errors (SE), t-values, and corresponding p-values for all of the sensors individually based on the DCC-GARCH model used.

Sensor	Estimate	SE	t	p
FP1-F7	0.00716	0.93865	0.00763	0.99391
F7-T7	-0.01182	1.11098	-0.01064	0.99151
T7-P7	0.02981	4.38913	0.00679	0.99458
P7-O1	-0.02568	0.01570	-1.63587	0.10187
FP1-F3	0.01309	0.01948	0.67163	0.50182
F3-C3	0.00087	0.02684	0.03225	0.97427
C3-P3	0.01812	0.01288	1.40702	0.15942
P3-O1	-0.03194	0.01476	-2.16369	0.03049
FP2-F4	0.03237	0.01217	2.66049	0.00780
F4-C4	-0.01198	0.01577	-0.75959	0.44750
C4-P4	0.01157	0.00003	409.95981	< 0.00001
P4-O2	-0.00146	0.01341	-0.10903	0.91318
FP2-F8	0.00336	0.01248	0.26950	0.78754
F8-T8	0.00904	0.01445	0.62556	0.53160
T8-P8	-0.03266	0.01350	-2.41919	0.01556
P8-O2	0.01461	0.01522	0.96004	0.33703
FZ-CZ	-0.00778	0.01456	-0.53437	0.59308
CZ-PZ	-0.01923	0.01733	-1.10960	0.26717
P7-T7	-0.02981	4.43820	-0.00672	0.99464
T7-FT9	-0.01773	0.01475	-1.20242	0.22920
FT9-FT10	0.00810	0.01517	0.53409	0.59328
FT10-T8	0.03175	0.01146	2.77025	0.00560

Table 3.2. Values of μ for a DCC-GARCH model with specifications of ARMA(0,0) and GARCH(1,1) for clip `chb08.02`.

The diagnostics for `chb05.13` are presented for an ARMA(1,0)-GARCH(1,1) DCC model in Figure 3.9. The covariance matrices for the actual sensor variance and predicted are displayed in Figure 3.10. Similarly, the diagnostics for

Sensor	Estimate	SE	t	p
FP1-F7	0.00095	0.00726	0.13064	0.89606
F7-T7	0.02706	3.04233	0.00890	0.99290
T7-P7	0.05351	3.23891	0.01652	0.98682
P7-O1	0.03637	0.01259	2.88918	0.00386
FP1-F3	0.00138	0.00148	0.93270	0.35097
F3-C3	0.03119	0.48788	0.06392	0.94903
C3-P3	0.01489	0.00838	1.77715	0.07554
P3-O1	0.02780	0.01446	1.92230	0.05457
FP2-F4	0.00038	0.00045	0.84089	0.40041
F4-C4	0.02327	0.01136	2.04808	0.04055
C4-P4	0.01019	0.00002	640.19030	< 0.00001
P4-O2	0.02559	0.01355	1.88936	0.05884
FP2-F8	0.00011	0.00020	0.55080	0.58177
F8-T8	0.04897	0.02429	2.01579	0.04382
T8-P8	0.03691	0.01128	3.27134	0.00107
P8-O2	0.08500	0.04458	1.90685	0.05654
FZ-CZ	0.09608	0.03264	2.94382	0.00324
CZ-PZ	0.03268	0.01273	2.56631	0.01028
P7-T7	0.05351	3.22703	0.01658	0.98677
T7-FT9	0.10183	0.04616	2.20579	0.02740
FT9-FT10	0.04093	0.01578	2.59399	0.00949
FT10-T8	0.00921	0.00917	1.00468	0.31505

Table 3.3. Values of ω for a DCC-GARCH model with specifications of ARMA(0,0) and GARCH(1,1) for clip `chb08.02`.

Sensor	Estimate	SE	t	p
FP1-F7	0.00957	0.01169	0.81896	0.41281
F7-T7	0.07400	6.15450	0.01202	0.99041
T7-P7	0.08644	4.80732	0.01798	0.98565
P7-O1	0.10453	0.04695	2.22621	0.02600
FP1-F3	0.00924	0.00168	5.51662	< 0.00001
F3-C3	0.03711	0.46366	0.08004	0.93620
C3-P3	0.10451	0.03118	3.35184	0.00080
P3-O1	0.07193	0.02263	3.17897	0.00148
FP2-F4	0.01190	0.00097	12.21913	< 0.00001
F4-C4	0.06138	0.02235	2.74597	0.00603
C4-P4	0.15551	0.00024	636.93804	< 0.00001
P4-O2	0.04329	0.01231	3.51624	0.00044
FP2-F8	0.01296	0.00103	12.53598	< 0.00001
F8-T8	0.13550	0.06015	2.25251	0.02429
T8-P8	0.10288	0.02271	4.53006	< 0.00001
P8-O2	0.12002	0.05282	2.27237	0.02306
FZ-CZ	0.20169	0.05564	3.62454	0.00029
CZ-PZ	0.09708	0.02237	4.34066	< 0.00001
P7-T7	0.08644	5.48248	0.01577	0.98742
T7-FT9	0.15562	0.05462	2.84914	0.00438
FT9-FT10	0.10102	0.02785	3.62769	0.00029
FT10-T8	0.08610	0.01317	6.53931	< 0.00001
Joint DCC	0.02496	0.00050	49.63762	< 0.00001

Table 3.4. Values of α for a DCC-GARCH model with specifications of ARMA(0,0) and GARCH(1,1) for clip `chb08.02`.

Sensor	Estimate	SE	t	p
FP1-F7	0.98943	0.00112	881.83231	< 0.00001
F7-T7	0.89867	9.26699	0.09698	0.92275
T7-P7	0.84417	8.11891	0.10398	0.91719
P7-O1	0.86213	0.04244	20.31643	< 0.00001
FP1-F3	0.98943	0.00007	13898.08917	< 0.00001
F3-C3	0.92669	1.01301	0.91478	0.36031
C3-P3	0.89024	0.03169	28.09611	< 0.00001
P3-O1	0.89855	0.03654	24.59215	< 0.00001
FP2-F4	0.98710	0.00052	1906.59168	< 0.00001
F4-C4	0.91415	0.03150	29.01955	< 0.00001
C4-P4	0.84093	0.00148	569.26242	< 0.00001
P4-O2	0.92985	0.02411	38.57383	< 0.00001
FP2-F8	0.98604	0.00035	2853.52707	< 0.00001
F8-T8	0.82897	0.06619	12.52458	< 0.00001
T8-P8	0.85203	0.03141	27.12959	< 0.00001
P8-O2	0.78791	0.09174	8.58830	< 0.00001
FZ-CZ	0.69275	0.07360	9.41296	< 0.00001
CZ-PZ	0.87087	0.03171	27.46688	< 0.00001
P7-T7	0.84417	8.93943	0.09443	0.92477
T7-FT9	0.74528	0.09091	8.19843	< 0.00001
FT9-FT10	0.86589	0.03368	25.70883	< 0.00001
FT10-T8	0.91290	0.02672	34.17107	< 0.00001
Joint DCC	0.95637	0.00014	6670.19110	< 0.00001

Table 3.5. Values of β for a DCC-GARCH model with specifications of ARMA(0,0) and GARCH(1,1) for clip `chb08.02`.

an ARMA(0,0)-GARCH(1,1) DCC model for `chb22.38` are illustrated in Figure 3.11 and the corresponding covariance matrices are displayed in Figure 3.12. For compactness, the parameters of both models will be excluded.

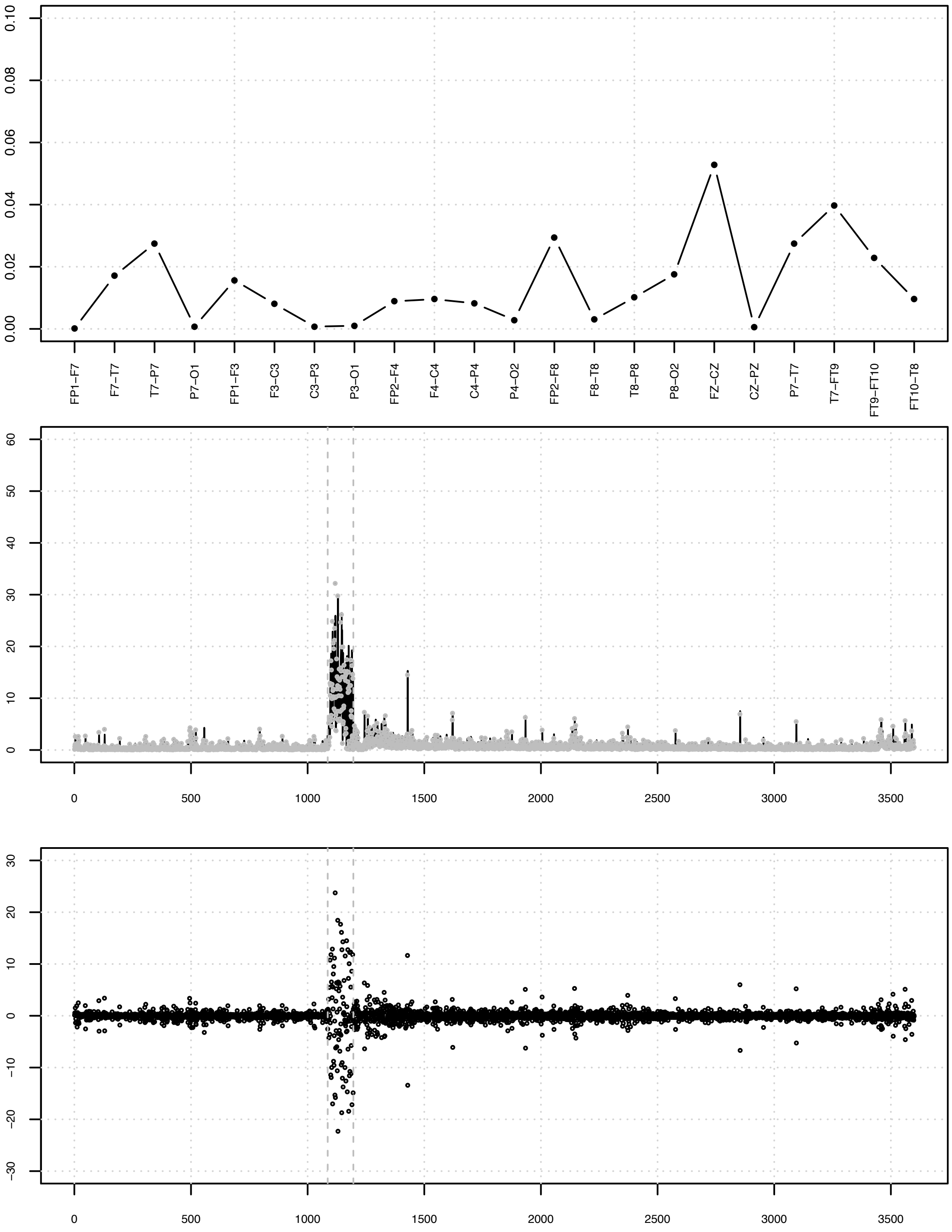

Figure 3.9. Standardized MSE values from the DCC-GARCH model for all 22 of the individual sensors for clip `chb05.13`. Plot of the actual sensor variance in black and the predicted sensor variance in grey circles from an ARMA(1,0)-GARCH(1,1) DCC model (middle). Residuals of the actual sensor variance compared to the predicted sensor variance from those displayed in the middle panel (bottom). Dashed grey lines represent the seizure boundaries.

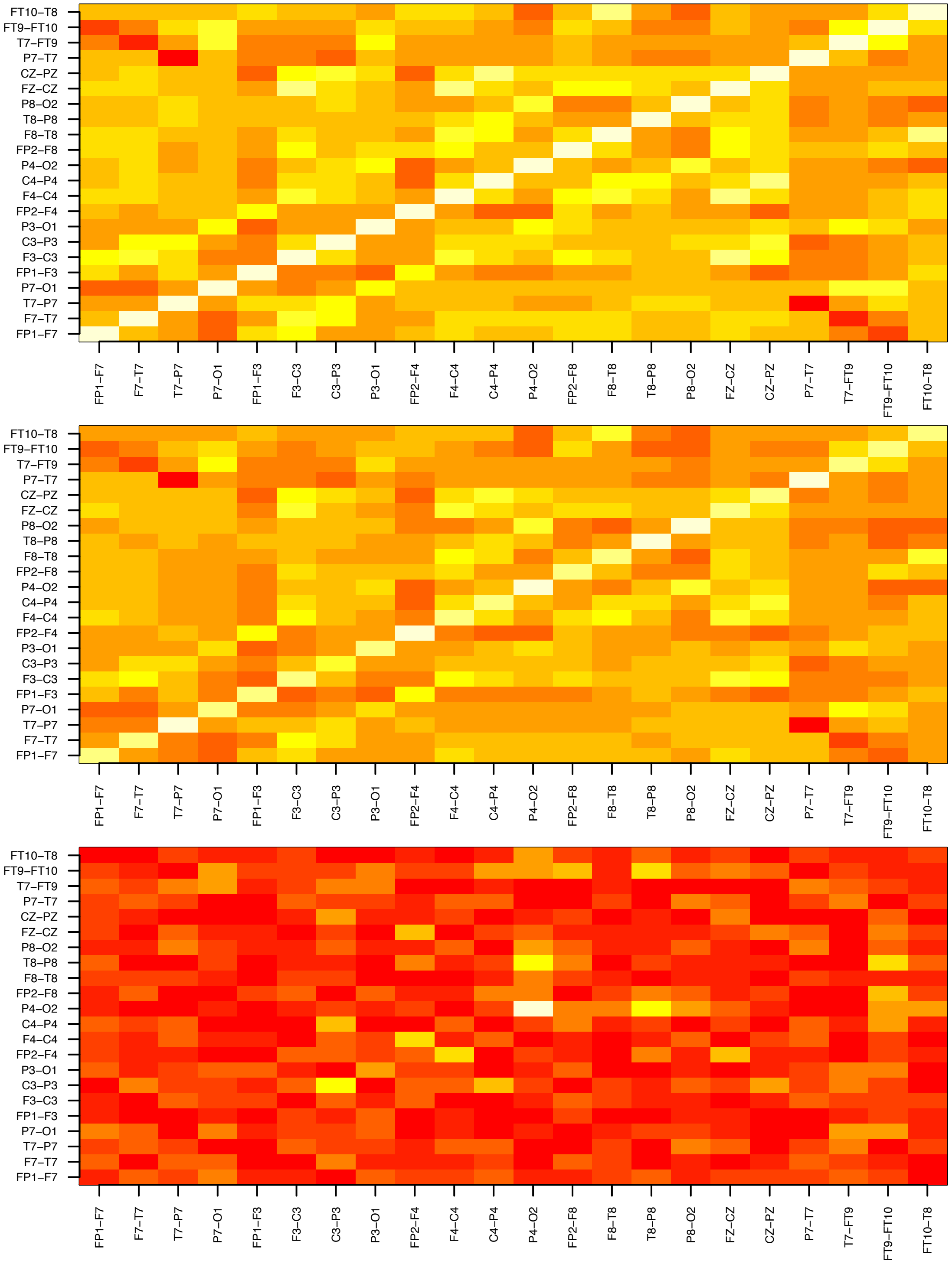

Figure 3.10. Heat-map images of the covariance matrices for the DCC-GARCH predicted covariance matrix (top), actual covariance matrix of the data (middle), and the absolute value of the difference between the predicted and actual covariance matrices (bottom) for clip `chb05.13`

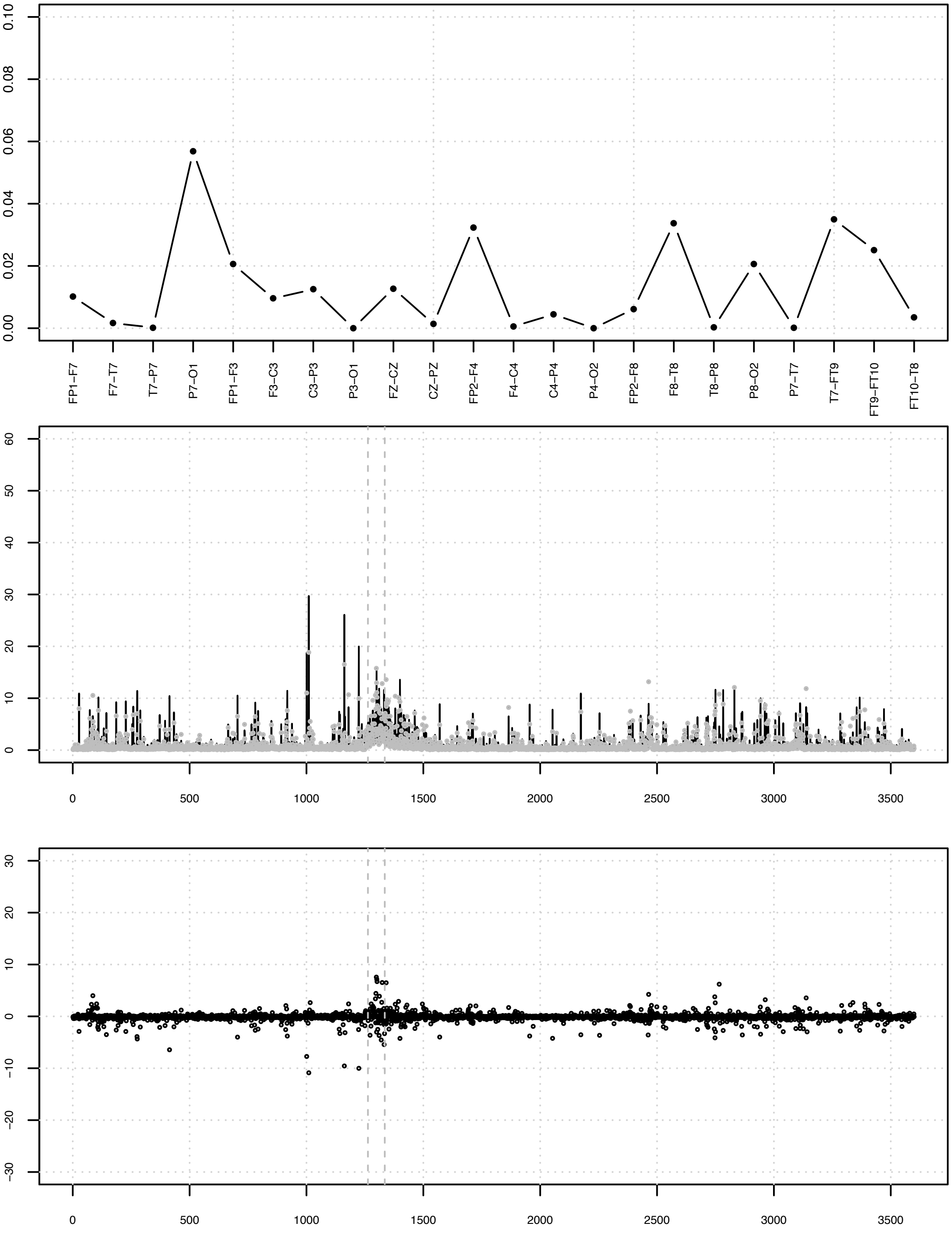

Figure 3.11. Standardized MSE values from the DCC-GARCH model for all 22 of the individual sensors for clip `chb22.38`. Plot of the actual sensor variance in black and the predicted sensor variance in grey circles from an ARMA(0,0)-GARCH(1,1) DCC model (middle). Residuals of the actual sensor variance compared to the predicted sensor variance from those displayed in the middle panel (bottom). Dashed grey lines represent the seizure boundaries.

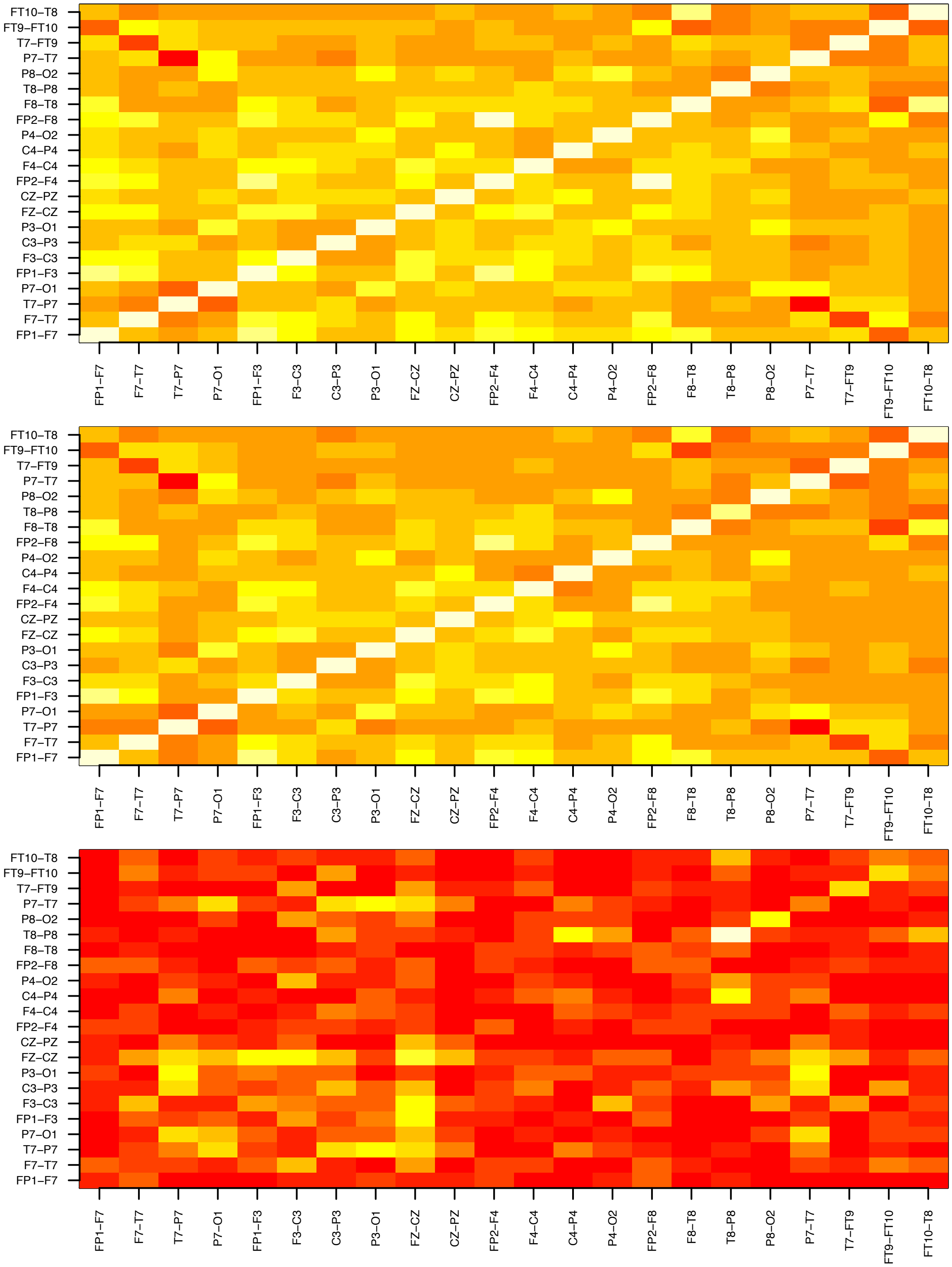

Figure 3.12. Heat-map images of the covariance matrices for the DCC-GARCH predicted covariance matrix (top), actual covariance matrix of the data (middle), and the absolute value of the difference between the predicted and actual covariance matrices (bottom) for clip `chb22.38`.

3.3 EEG Sensor Variance

Common measures of EEG data in regards to synchronization, desynchronization, and re-synchronization in regards to seizures have focused on the correlations between only a few sensors [7, 8, 9, 10]. It could then be expected that the patterns of (and lack thereof) synchrony could be seen in a much larger spectrum covering more spatial area in the brain by utilizing the variance of the sensors at each time point. The variance (s^2) addresses the amount of variation with a vector of sample values commonly as:

$$s^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1}$$

where n represents the sample size and y_i represents the data in $y_1, \ldots, y_n$ with a sample mean of $\bar{y}$. This gives an unbiased variance of the sample of values which is a valid representation on the expected population variance (σ^2) of:

$$\begin{aligned} \hat{\sigma^2} &= \mathrm{E}\left[X^2\right] - \left(\mathrm{E}\left[X\right]\right)^2 \\ &= \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N} \end{aligned}$$

where N is the complete set of discrete values of x_i and a population mean of μ. Variance is always positive and also equivalent to the population standard deviation (σ) squared, or in the sense of a sample (s) squared.

In terms of a multivariate time series, assume a matrix $\mathbf{M}$ of size $r \times c$ with r rows and c columns. When there is a total of T time points in the discrete series, we can presume r consecutive values from $1, 2, \ldots, T$ in $\mathbf{M}$. Let each column from

$1, 2, \ldots, c$ in $\mathbf{M}$ be the values of each sensor respectively in the time series from $1, 2, \ldots, T$. Thus each c is a vector of consecutive values from $1, 2, \ldots, T$ for one sensor.

Iteratively, one could now take each r in $\mathbf{M}$ as an independent sample of the EEG state at a given time point. Let Y' be a storage vector of size r where:

$$Y_i = \frac{\sum_{i=1}^{r} (r_i - \bar{r_i})}{c - 1}$$

when r_i represent the vector of EEG sensor values at each given time point i. While it is not necessarily, it is preferable and likely that there are no missing values and that all the sensors represented as c have the same amount of values for r. It is also noteworthy that this same process could simply be extrapolated for the transpose of $\mathbf{M}$ where r represents the individual sensors and c the time points from $1, 2, \ldots, T$.

A new approach to analyzing the multivariate data available in the CHB-MIT scalp database is now demonstrated. Based on the idea of brain desynchronization, strong changes in the synchrony of the individual sensors can be extrapolated by looking at the variance of the sensors at every time point. In the case of the clip `chb08.02` highlighted, with one point per second and a total of $3,600$ points, there would also be $3,600$ univariate based sequential data points for the hour clip. For example, at time point t, the variance between all 22 sensors at time t is calculated. The same follows for all points in T. Figure 3.13 displays the univariate measure of brain synchrony based on the variance of the sensors at one point per second. This is the same as the sensor variance plotted in Figure 3.5 for `chb08.02`.

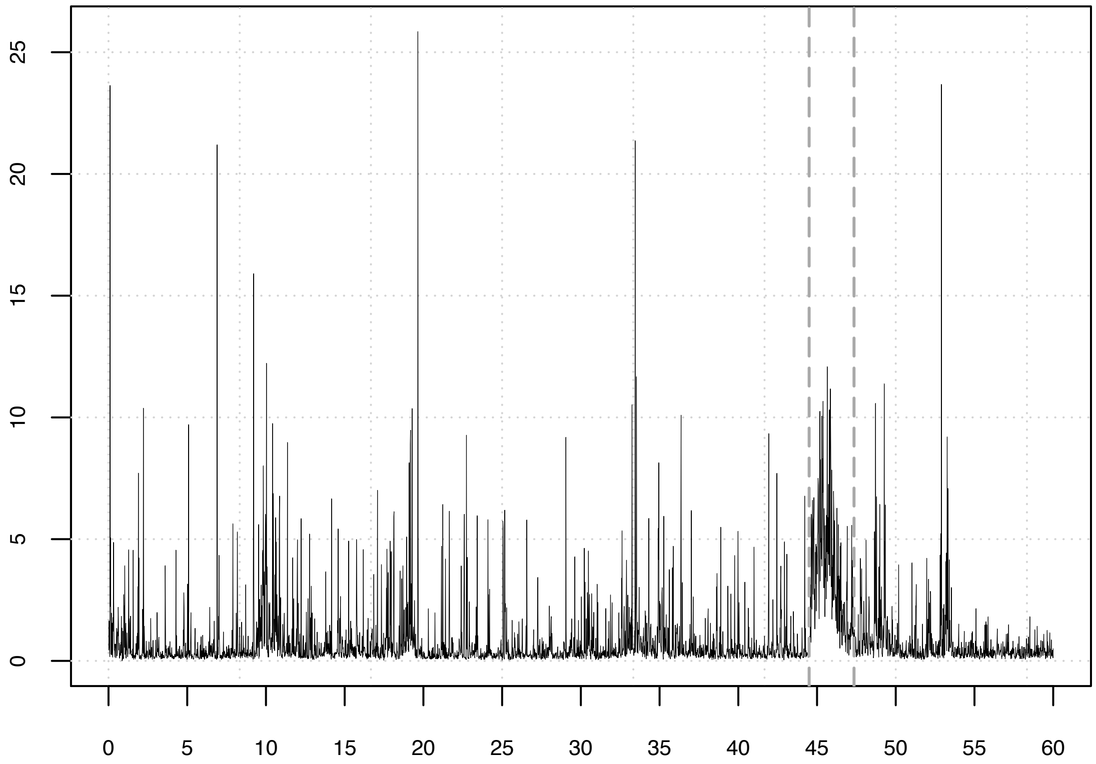

Figure 3.13. Time series plot of the variance of all 22 sensors sequentially, with one point per second. The dashed grey lines represent the professionally marked beginning and end of the seizure in this clip. Notably, the sensor variance is persistently larger during the seizure portion than the non-seizure portions.

The obvious benefit of this approach is that it can highlight desynchronization more simply from the complex multivariate measures available. Figure 3.13 displays a commonplace occurrence in not only the clip used for this example, but for many if not most of the seizure clips in the CHB-MIT database, with a strong persistence of sensor variance primarily during the seizure. The persistence of the variance to stay away from 0 seen in Figure 3.13 suggests that the variance between the 22 sensors is consistently larger than it is during the non-seizure points. Some occasional rises in variance with notably high peaks in the non-seizure portions can also be seen, but is presumed to be movement of the headset or other potential non-seizure phenomenon such as a sneeze that might cause some disturbance. The long term persistence during the 171 seconds of the professional marked seizure occurrence is definitely stark and quite easy to view simply by eye.

As with any time series based data, it is important to observe the autocorrelation between the time points. Figure 3.14 shows the autocorrelation function (ACF) of the univariate variance measures extracted from the clip `chb08.02`.

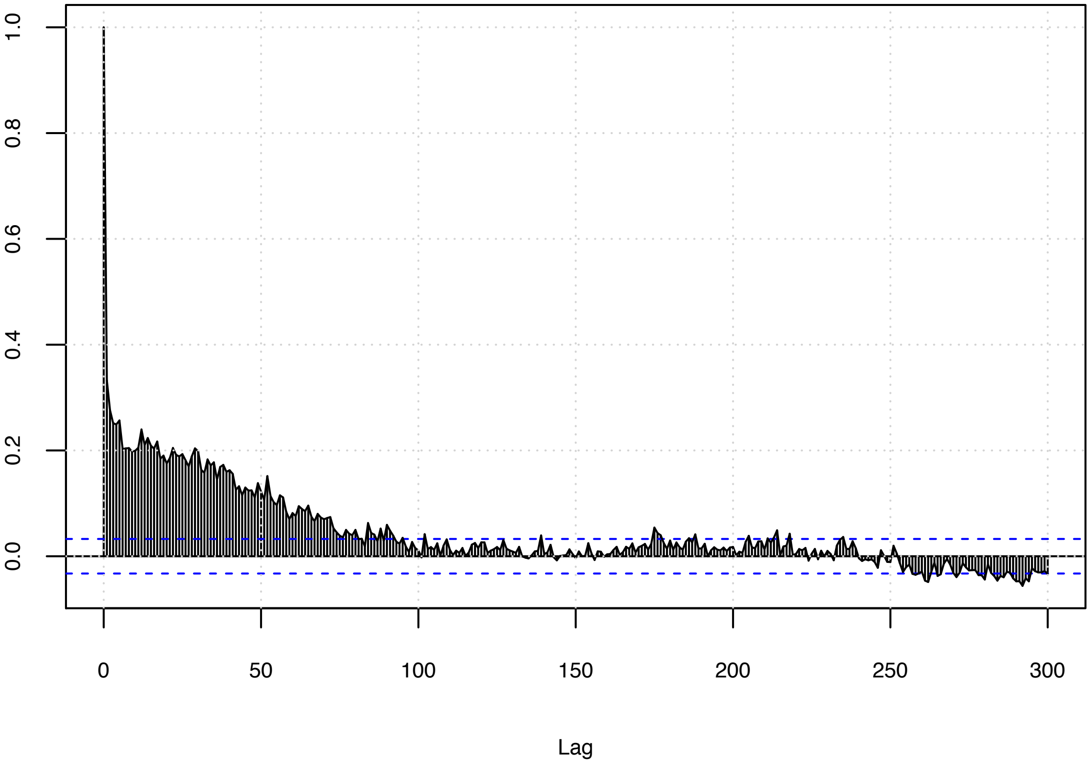

Figure 3.14. Autocorrelation function of the variance of the 22 sensors at one time point per second for a total of 60 minutes.

Because of the use of variance, all values are lower bounded by 0 and thus a logarithmic transformation can be applied to help make the truncated data more Gaussian based (see Figure 3.15). Use of the log of the variance will allow for a mixture of normals to be utilized, which would not be advised for the original variance of the sensors as seen in Figure 3.14 because of the heavy density of the data at and near 0. Future avenues could utilize a mixture of Gamma distributions however this research will opt to work with normal mixtures based on the log-transformation of the sensor variance.

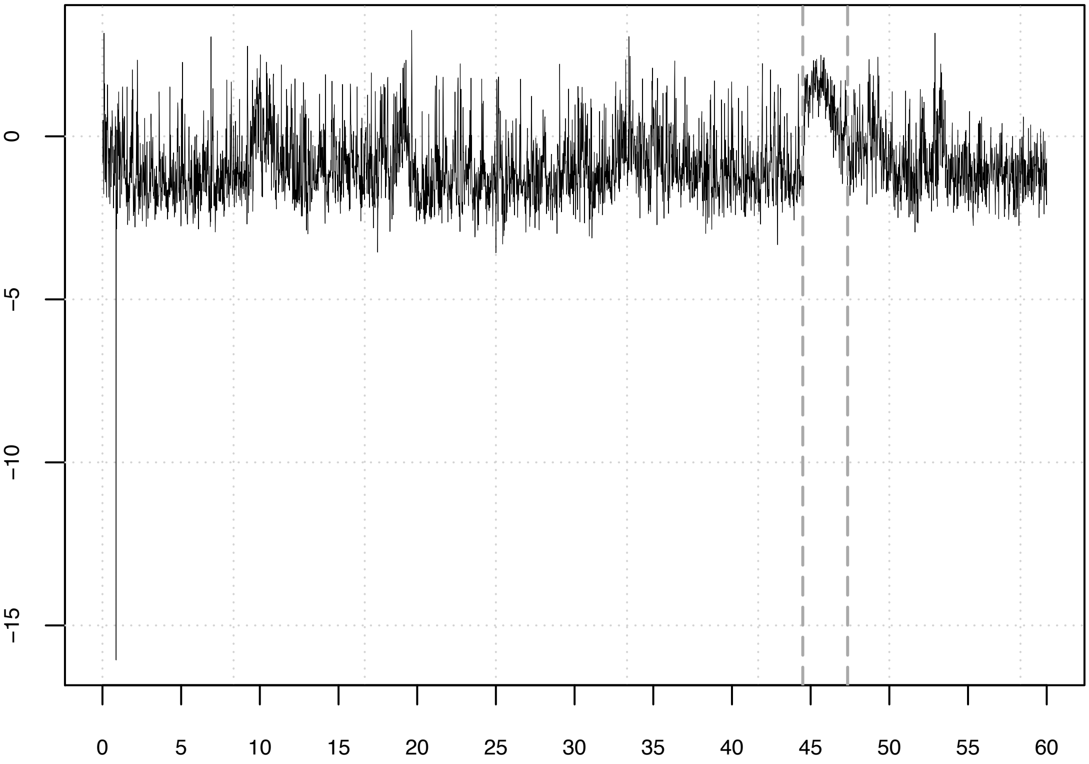

Figure 3.15. Logarithm of the variance of the 22 sensors from the same data as displayed in Figure 3.10.

Figure 3.16 displays the kernel density estimate of the logarithmic based sensor variance with a solid black line. The data does have a slight bump towards the right hand of the density, thus suggesting it is not normal but perhaps a mixture of several normals. It is worth noting there is one extreme outlier at about -16.06 which is only one point and is likely noise, not seizure based. All of the remaining points reside between -3.58 and 3.26 and have a relatively Gaussian based density. Notably the density appears to be bimodal with the right hand side of the density in Figure 3.16 showing a slight increase in the density when compared to its left hand equivalent, a result of the seizure sensor variance persistence. Figure 3.16 is an example of what a density for a normal mixture model could look like, and thus use of a Markovian mixture model could be applied.

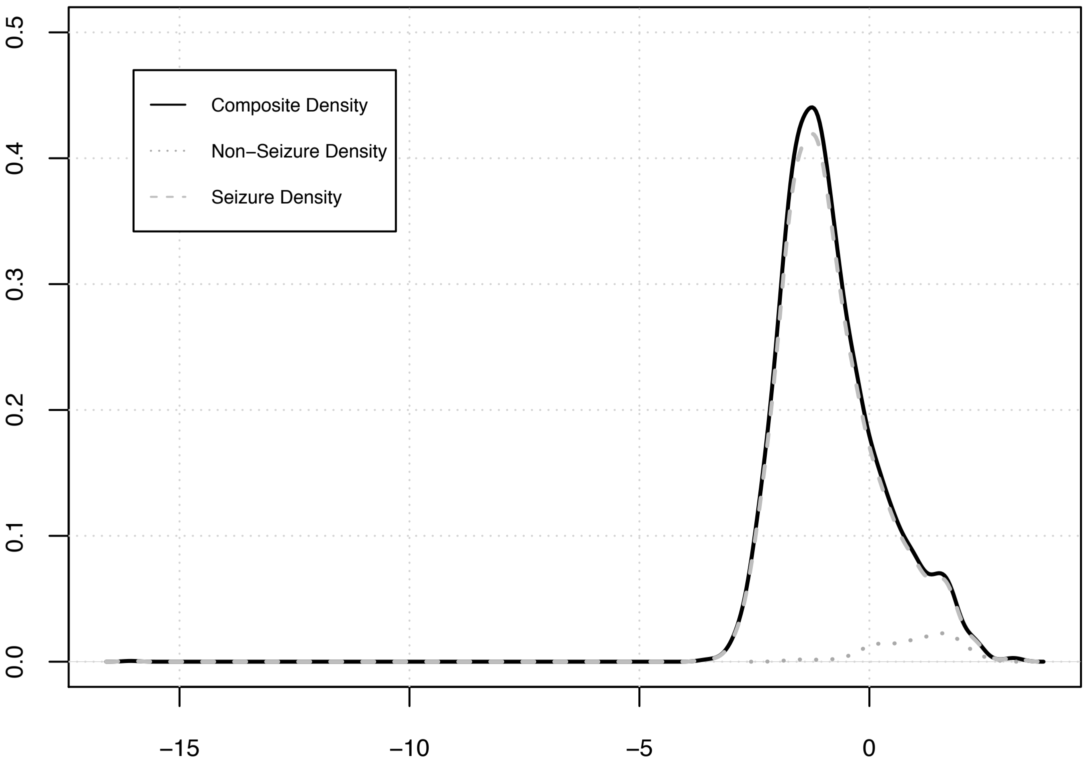

Figure 3.16. Complete density of all 3,600 points from clip `chb08.02`. The density of the non-seizure points are highlighted (dashed) as well as the seizure point density (dotted). Both the non-seizure and seizure densities are scaled according to the number of points they represent out of the total of 3,600 points.

The box-plots of the points from clip `chb08.02` in Figure 3.17 highlight the one extremely low point which can also be seen in Figure 3.15, however the rest of the points are in general well distributed. There are a handful of potential outliers in the non-seizure box-plot that are higher peaks in the data, which again are not actual seizure points but movement noise. Notably, all of the seizure points fall within a suitable box-plot, which also shows an overall larger set of values than the non-seizure box-plot and the composite box-plot.

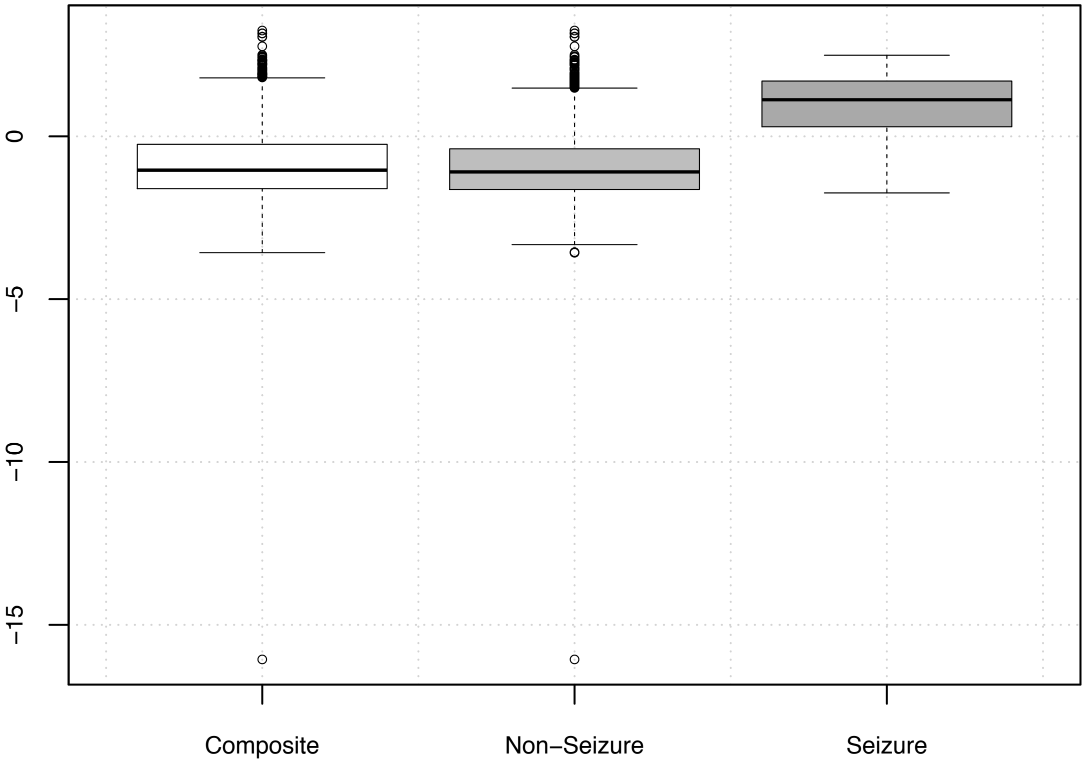

Figure 3.17. Boxplots for the full composite mixture density (left), non-seizure data points (middle) and seizure data points (right).

3.4 Markov Regime Switching Mixture Models

The initial analyses for the three clips (Figure 3.4-3.6) have so far suggested two component models would be adequate, which fits ideally with the idea of a non-seizure and seizure based model. Furthermore, the two states will be referred to in a binary format as 0 (non-seizure) and 1 (seizure). Modeling of the two states or regimes, can be done by assessing the point probabilities of being a seizure and approximated with a Gibbs sampling algorithm. A prior for the Dirichlet distribution will be utilized and the number of transitions across the $3,600$ sequential time points will be needed. The transitions for a standard model require four prior values as well, that of transitioning from non-seizure to seizure (01), seizure to non-seizure (10), and staying in a non-seizure (00) or seizure (11) state between $t-1$ and t. Similarly this can be extended to three points of $t-2$, $t-1$, and t with eight total combinations of transitions available (000, 001, 010, 100, 101, 110, 011, 111). Further extensions are also possible following the same concept outlined. Because it seems unlikely to have a state go from non-seizure to seizure and non-seizure (010) as similarly for seizure to non-seizure to seizure (101) in 3 seconds, the transitions will only be modeled using two points (00, 01, 10, 11).

A somewhat strong prior can be justified for the mixture model when using post-hoc data such as that in the CHB-MIT scalp database since the proportions of non-seizure and seizure points can be directly calculated. For example, clip `chb08.02` is approximately 0.95 non-seizure and 0.05 seizure and since we are basing the data on one point per second for one hour, we would like to keep our priors for the transitions based on about $3,600$ points. Assuming a prior too concentrated can outweigh the actual data, we will opt for a slightly weaker yet informative prior of $\pi(00) = 3,400$, $\pi(01) = \pi(10) = 5$, and $\pi(11) = 190$. The other two clips don't fluctuate much from the same ratio of seizure to non-seizure points and thus all three clips will use the same prior.

The Bayesian priors for the model were set at $\mu_0 = \mu_1 = 0$ for the mean and $\tau_0 = \tau_1 = 1$ for the precision. The Gibbs priors for the conditionals of μ_0, τ_0, ν_0, and s_0^2 were all set as 1. These are considered weak priors and have very little bearing on the overall model as the iterations of i increase.

Unconstrained random permutation sampling (RPS) for all three clips are highlighted in Figure 3.18. Allowing the dummy coded 0 and 1 labels to randomly fluctuate and plotting the two μ values as the x and y axes helps determine if there is proper separation in the mixture densities. Because of the low frequency of seizure points, the separation of the points in Figure 3.18 isn't terribly exaggerated, however it is clear there are two components ($k = 2$) at work. This is a good indicator that a constrained RPS should be used. Therefore, the Markov regime switching models will be run with a constraint for the two component means s.t. $\mu_0 < \mu_1$.

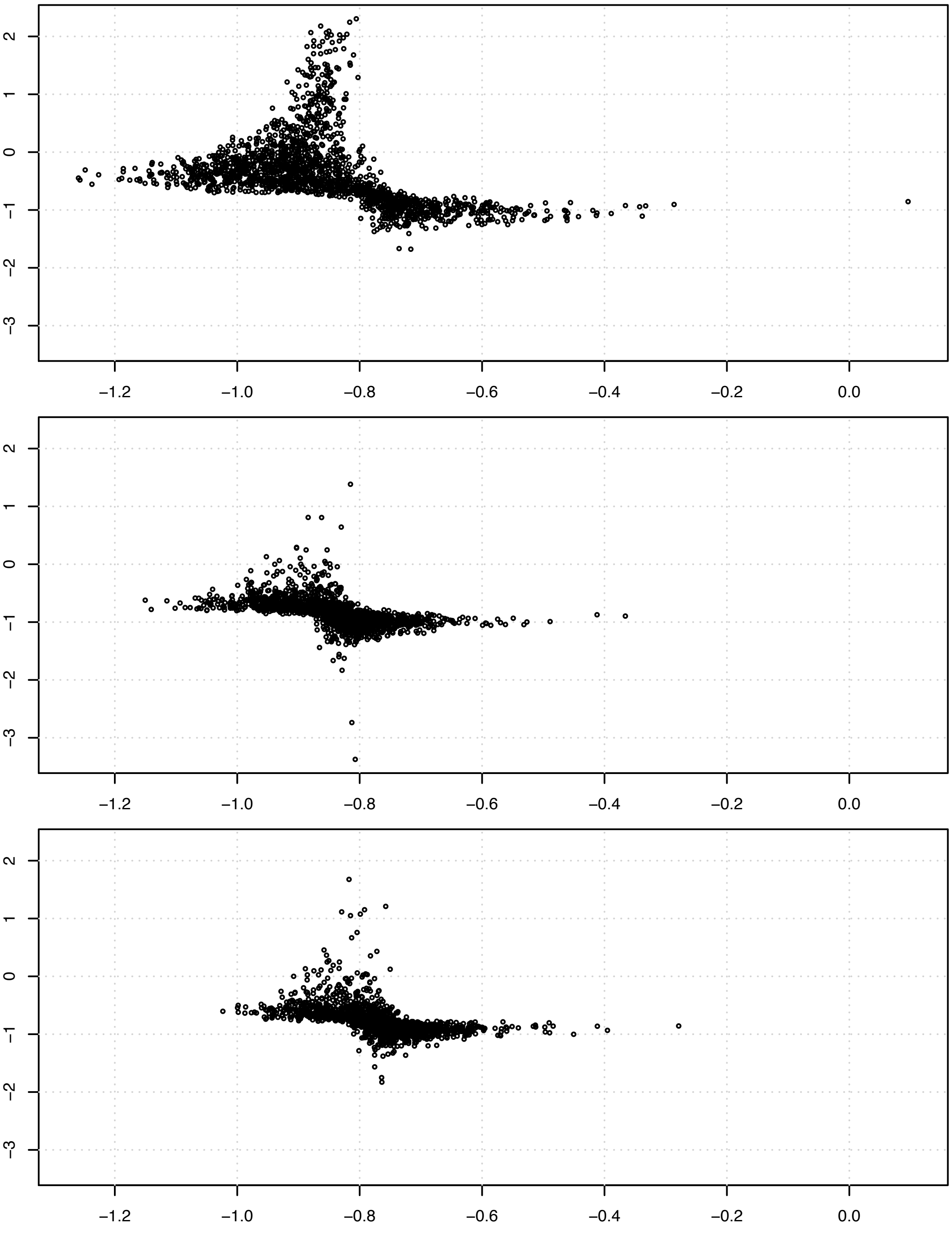

Figure 3.18. Plot of μ_0 (x) against μ_1 (y) for $1,800$ Gibbs sampling iterations with an unconstrained random permutation sampling scheme and a 200 iteration burn-in. Three clips are analyzed, `chb05.13` (top), `chb08.02` (middle), and `chb22.38` (bottom).

Each of the three clips were run with a Gibbs sampler of $20,000$ iterations and a burn in of $2,000$ which removes some of the instability that can occur at the initialization of the model. To remove the autocorrelation of the sampling, a thinning of 1 point per 18 will be used to generate $1,000$ total points to be analyzed. The constrained permutation sampling will require that $\mu_0 < \mu_1$ for all iterations of the sampler. Figure 3.19 highlights the point probabilities for the thinned samples of $1,000$ for all three clips, with the actual seizure boundaries marked in dashed red lines.

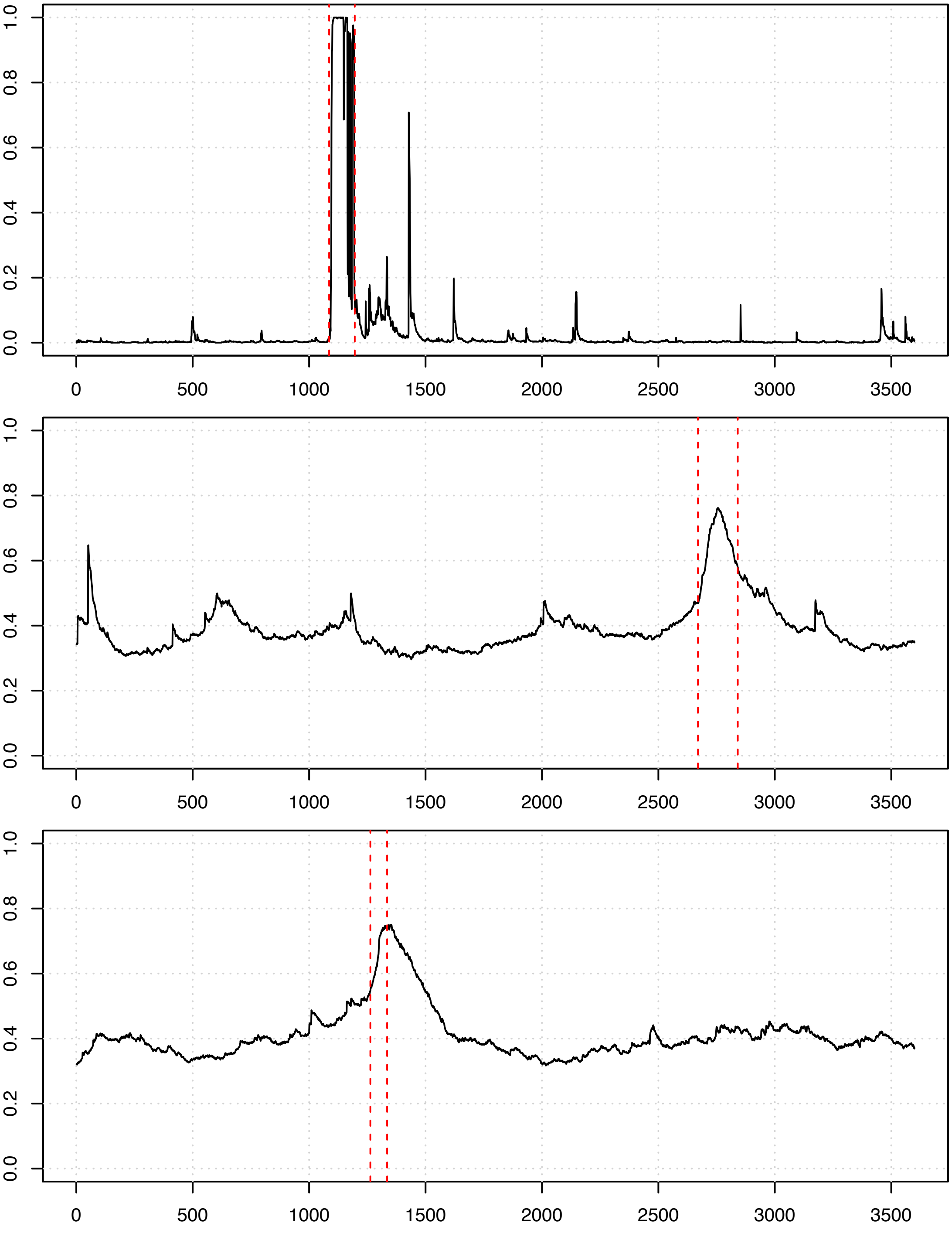

Figure 3.19. Plot of the Gibbs sampling based Markov switching point probabilities based on 18, 000 points thinned to 1 point per 18, or a total of 1, 000 points after a burn-in of 2, 000 iterations. Three clips are analyzed, `chb05.13` (top), `chb08.02` (middle), and `chb22.38` (bottom).

All three of the regime switching models in Figure 3.19 show relatively good success. The model for `chb05.13` looks different than the models for `chb08.02` and `chb22.38`, which could be the result of more noise in the data for the latter two clips. Nonetheless, the point probability plots all seem to model the variance of the 22 sensors quite well, with the seizure areas peaking above 0.6 for both `chb08.02` and `chb22.38` and the non-seizure points in general staying below 0.4-0.5. Considering the EEG files used are raw and unfiltered data, these results still suggest strong modeling capabilities of the Markov regime switching model mixed with a proper Dirichlet prior. The plot for `chb05.13` is much more distinct, with the point probabilities for the seizure area staying consistently close to 1.0, while the non-seizure points stay close to 0.0.

Figure 3.19 highlights the notion that not every patient will be modeled quite the same, particularly because the amount of noise can vary as can seizure location. Use of the Markov regime switching model with the addition of a threshold of perhaps 0.5-0.6 could have important use in future ambulatory seizure treatment research. An added benefit of the use of Bayesian modeling is that further research can better adjust the prior information going into the model, thus enhancing the model even more.

3.5 Change Points Models

Gibbs sampling was used to determine the most likely change points in the models, or where the state goes from non-seizure to seizure (change point 1) as well as from seizure back to non-seizure (change point 2). The iterations of the model converge relatively fast compared to the Markov regime switching models, and thus only $1,000$ iterations are assessed after a burn-in of 100.

Because of the specific application of these clips, only two change points will be determined, however initial starting points are necessary as well. To keep good separation of the two change points but to also prevent starting them at the very beginning and end, a moderately informative starting point for the change points is utilized. Change point 1 was set at $1,000$ and change point 2 was initialized at $2,600$ based on the $3,600$ possible points available. Again, convergence can occur quickly, however poor starting points for the change points can cause some convergence issues.

The Bayesian priors for the model were set at $\mu_0 = \mu_1 = 0$ for the mean and $\tau_0 = \tau_1 = 1$ for the precision as they were in the previous models. The Gibbs priors for the conditionals were the same as those used in the Markov regime switching model, or $\mu_0 = \tau_0 = \nu_0 = s_0^2 = 1$.

The trace plots of the estimated change points for all three clips are shown in Figure 3.20. The means for change point 1 and 2 for clip `chb05.13` are $1,075.915$ and $1,452.240$ respectively. The actual marked change points for the clip are $1,086$ and $1,196$, suggesting change point 1 was very close and change point 2 was about 256 seconds later than the actual end of the seizure. Clip `chb08.02` resulted in means of $2,673.895$ and $2,964.980$ when the actual change points are $2,670$ and $2,841$. The results for this clip were much more promising as change point 1 was almost exact, however change point 2 still overshot the end of the seizure by over 100 seconds. Finally, clip `chb22.38` has mean change points of $1,251.978$ and $1,473.323$ compared to the actual changes of $1,263$ and $1,335$.

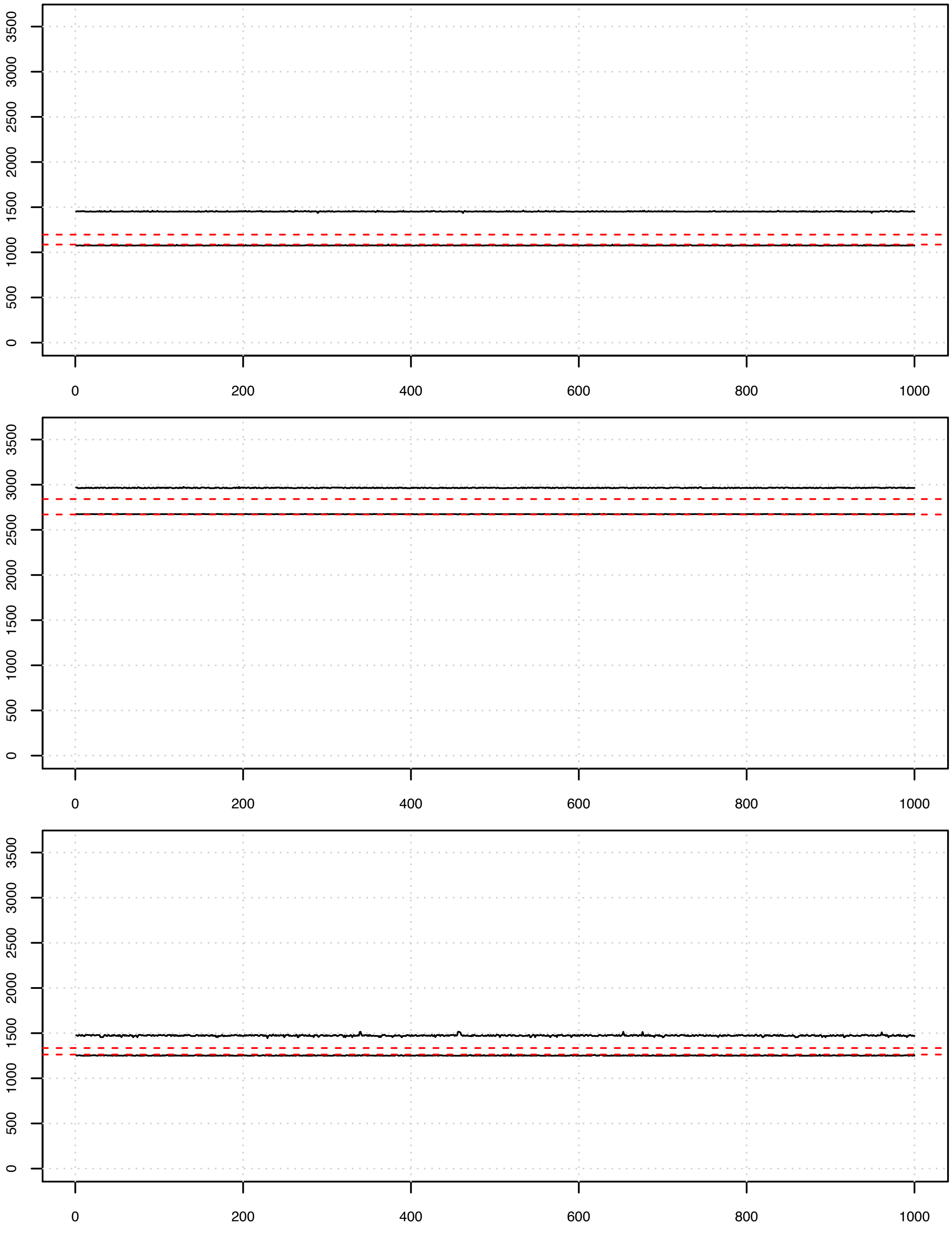

Figure 3.20. Plot of the Gibbs sampling based change points for $1,100$ iterations with a burn-in of 100 for a total of $1,000$ iterations. Three clips are analyzed, `chb05.13` (top), `chb08.02` (middle), and `chb22.38` (bottom).

All three change point models had promising results for the change point from non-seizure to seizure, however the seizure to non-seizure change point tended to be slightly delayed. It is possible the lingering effects of the seizure are carrying beyond the marked end of the seizure and some excess headset movement might also occur.

List of References

[1] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[2] A. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment," Ph.D. dissertation, Massachusetts Institute of Technology, September 2009.

[3] K. H. Lee and N. Verma, "A low-power processor with configurable embedded machine-learning accelerators for high-order and adaptive analysis of medical-sensor signals," *Solid-State Circuits, IEEE Journal of*, vol. 48, no. 7, pp. 1625–1637, 2013.

[4] J. Chiang and R. K. Ward, "Energy-efficient data reduction techniques for wireless seizure detection systems," *Sensors*, vol. 14, no. 2, pp. 2036–2051, 2014.

[5] H. Khammari and A. Anwar, "A spectral based forecasting tool of epileptic seizures," *IJCSI International Journal of Computer*, 2012.

[6] Y. U. Khan, O. Farooq, and P. Sharma, "Automatic detection of seizure onset in pediatric EEG," *International Joural of Embeded Systems and Applications*, vol. 2, no. 3, pp. 81–89, 2012.

[7] F. Mormann, T. Kreuz, R. G. Andrzejak, P. David, K. Lehnertz, and C. E. Elger, "Epileptic seizures are preceded by a decrease in synchronization," *Epilepsy research*, vol. 53, no. 3, pp. 173–185, 2003.

[8] P. Olejniczak, "Neurophysiologic basis of EEG," *Journal of clinical neurophysiology*, vol. 23, no. 3, pp. 186–189, 2006.

[9] P. R. Carney, S. Myers, and J. D. Geyer, "Seizure prediction: methods," *Epilepsy & Behavior*, vol. 22, pp. S94–S101, 2011.

[10] P. Jiruska, M. de Curtis, J. G. Jefferys, C. A. Schevon, S. J. Schiff, and K. Schindler, "Synchronization and desynchronization in epilepsy: controversies and hypotheses," *The Journal of physiology*, vol. 591, no. 4, pp. 787–797, 2013.

CHAPTER 4

Discussion

4.1 Model Summary

In all, three clips from the CHB-MIT database were examined using three different time-series modeling techniques. The initial analysis focused on multivariate DCC-GARCH modeling, which also served as a potential validation for the use of the sensor variance in a univariate modeling approach. After calculation of the sensor variance for all three clips, a univariate Markov regime switching model was explored, followed by a change point model of the sensor variance. All of the models aimed to predict the location of seizures in clips with knowledge of there being a seizure evident in the one hour clips.

The DCC-GARCH models performed very well, even with heavy computational requirements still necessary. The MSE values were low for all three clips, suggesting the predicted models were fairly accurate and the residuals were not too large. Plotting of the predicted sensor variance against the actual sensor variance (Figures 3.7, 3.9, and 3.11) help support the use of EEG sensor variance in a univariate fashion as a measure of volatility as again the predicted and actual sensor variance follow each other rather closely in all three clips. The heat-map images of the difference between the actual and predicted covariance matrices (Figures 3.8, 3.10, 3.12) also support the use of DCC-GARCH modeling for multivariate EEG sensor modeling as most of the differences were only negligible for all three clips.

What is a downfall for EEG modeling with the DCC-GARCH resides in the large number of parameters. This very large set of parameters all need estimating and make overall analysis of such a large DCC-GARCH model daunting. The DCC-GARCH model could potentially be modified for future analysis by using a smaller number of the 22 sensors available. Ideas include using one sensor in each quadrant of the brain, or a left, middle, right hemisphere scheme. All 22 sensors were used here primarily because of the desire to examine the univariate modeling of the sensor variance for all available sensors.

With the very adequate modeling results of the DCC-GARCH, analysis moved to univariate measures based on the variance of the sensors at all $3,600$ time points. The exact same time points were used for all three models in this study, the only change was from using all 22 sensors in the DCC-GARCH to only using a single vector of values for the two Bayesian models. The ideal was to capture the persistence of the volatility from the sensors that was very apparent in the time-series plots. While many spikes and peaks occurred at random through the clips, the more important concept was the variance of the sensors to stay away from 0 for the duration of the seizure, which was not nearly as noticeable in the non-seizure portions.

The Markov regime switching model was the first univariate Bayesian analysis presented. Results for all three clips were again very promising. The initial step was to use unconstrained RPS to help verify that $k = 2$ was adequate for the models. Figure 3.18 suggests that a mixture of two components is valid and thus the constraint of $\mu_0 < \mu_1$ was employed, meaning the mean of the non-seizure points should be less than the seizure points. This also makes sense conceptually as the plots of the logarithms of the sensor variances showed the seizure points as markedly higher than the non-seizure points (Figures 3.4-3.6).

Because of potential autocorrelation in the Gibbs sampling for a Markovian model, a large amount of iterations $(20,000)$ were used with a burn-in or $(2,000)$ and a thinning of only one point per eighteen. A total of $1,000$ samples could then be used to calculate the point probabilities for all $3,600$ time points based on their averages as displayed in Figure 3.19. Clip `chb05.13` had very strong predictive power during the entire seizure and very few large spikes elsewhere. Clips `chb08.02` and `chb22.38` also had good predictive power during seizures, however they only neared 0.8 whereas `chb05.13` was very close to 1.0. Also noticeably different is the higher probabilities for the non-seizure points of the two bottom clips in Figure 3.19, while `chb05.13` had most non-seizure points much closer to 0. This suggests that the models behave slightly differently, yet the main results would suggest that the seizures generally occurred when the point probabilities exceeded 0.6 with a few exceptions.

Finally, the change point analysis utilized the univariate sensor variance similar to the Markov regime switching model. The goal of the change point analysis was to determine the most likely change points in the models without the model actually knowing the change points. Figure 3.20 highlighted the predicted and actual change points of non-seizure to seizure and seizure to non-seizure for all three clips. The first change point, indicating non-seizure to seizure state, was fairly accurate for all three models. The second change point, representing the change of state from seizure to non-seizure, was not as promising for all three clips. What is worth mentioning is that even in the Markov regime switching models (see Figure 3.19) there is a still a slightly higher predicted seizure probability than directly before the seizure starts. This could in turn be reflected by the change point analysis, particularly with the second change point occurring later than the actual change point for all three models.

All three models showed fair to good results in modeling seizure location for the clips presented from the CHB-MIT database. The upside is that the multivariate DCC-GARCH supplied sufficient modeling of the data without summation, and also gave credence to using the sensor variance as opposed to all 22 channels simultaneously. The Bayesian Markov regime switching model showed a strong capability for modeling the sensor variance and could be a potentially useful tool for online seizure detection. Because clips `chb08.02` and `chb22.38` had higher point probabilities in general, a threshold of perhaps 0.5 or 0.6 could be very useful as a signal detector for the start of a seizure. In terms of implanted EEG shocks to prevent seizures, a small probability spike in the clips periodically wouldn't do any harm to the patients. Another option would be for a threshold of several seconds worth of 0.6 or higher probabilities to prevent sporadic electrical stimulation. Clip `chb05.13` had low to almost 0 point probabilities for the pre-seizure and post-seizure areas with the probabilities during seizure reaching almost 1, thus even a threshold of say 0.6 would still be applicable.

The Bayesian change point models required a little more precision with the initialization of values and could certainly be a potential hinderance to online usage. Not to mention the knowledge of there being a seizure in the clip and thus two change points was assumed and most certainly wouldn't always be the case. That withstanding, the change points from non-seizure to seizure were very close to the actual seizure start, which could certainly help in an ambulatory electric shock situation where the beginning of the seizure is much more important than the end. The change points from seizure to non-seizure were delayed quite a bit when compared to the actual end of the seizure, however there seems to be some carryover persistence in the sensor variance directly after the seizures. This can also be seen in the Markov regime switching models, where the point probabilities have a slow decline back to the approximate area of the non-seizure points, particularly for clips `chb08.02` and `chb22.38.`

Overall, in a post-hoc manner, the DCC-GARCH models might actually be a valid approach if an EEG researcher is willing to work with a large amount of parameters and slow computation. For a faster and more particularly ambulatory type use, the Markov regime switching model seems to be more logical and demonstrated better results than the Bayesian change point models. Time wise, the regime switching models were calculated the fastest of the three models. While the change point models are more simple in scope, they require constant back and forth point by point iteration which caused for slower computation than the regime switching models. The Markov regime switching models did converge rather quickly, and it is worth noting the models could be run efficiently without $20,000$ iterations and thinning if needed.

4.2 Future Directions

The goal in this thesis was to assess several models and there potential application to ambulatory EEG usage. The models were all run in a post-hoc fashion with knowledge of there being a seizure, so this research was of a potential direction for models that could be used in the future. Modifications to all of the models would be necessary to apply the techniques into a realtime environment, but they certainly could be possible.

Primarily, the Bayesian Markov regime switching model might have the best application potential. What would be needed is a particle filtering routine that constantly updates the probabilities with the information being collected and the information already collected [1]. A threshold would also need to be applied to determine when the probabilities are great enough to cause electrical stimulation if that is the purpose of the ambulatory research.

For research interested in ambulatory online EEG data collection without treatment of stimulation, the regime switching model could be highly beneficial. Again, a particle filter type routine would be necessary to replace the FFBS algorithm used in this study. While prior information is needed for the Bayesian models, it would be assumed that those parameters could be fine tuned with analysis of hundreds of various EEG clips. Keeping the priors informative but not too strong would be advised to prevent the EEG readings from spiking excessively or not at all.

What does seem to be a viable and so far unused diagnostic for EEG data regardless of model is the sensor variance. The initial plots of the data displayed stark persistence in the sensor variance when the seizure were occurring versus when they were not. This becomes a potentially useful univariate measure with plenty of applications and further types of analysis if needed. Because of the grand scale of data points in EEG studies, sensor variance as a measure of brain synchronization has merit.

The DCC-GARCH model has a benefit in that it doesn't require downsizing the data into a univariate measure and it also doesn't require Bayesian priors. Potential work could always look at the use of only several sensors in an online recording scenario, however the parameters would need to be constantly updated which again seems somewhat tedious. The option does exist however, and again the DCC-GARCH does model the volatility of the 22 sensors sequentially quite well.

The change point analysis models might be improved with tuning parameters that were not specified in this study. Regardless, the change point models have a downside similar to the DCC-GARCH in that they really work with discrete data in a post-hoc manner. Application of this type of model seems less versatile, however its ability to capture the start of the seizure by change point from non-seizure to seizure might give it some interest for future work with added model specifications adapted for online data acquisition.

In terms of analysis, future work could certainly look at inter-cranial EEG research. It is known that scalp measures can be more a measure of the physical muscle contractions during a seizure. Inter-cranial measures would also need investigation to determine the appropriateness of sensor variance usage in EEG modeling because they actually measure synaptic functioning much better than scalp EEG recordings. Future work could also address whether 22 or more sensors are even necessary in terms of variance. Perhaps only several sensors are needed to capture brain desynchronization via sensor variance.

The age of the patients used could also be an issue of contention in that a child's brain might not function the same as an adult. There are some patients in their teens and even early 20's in the CHB-MIT database, so they could be examined as well. However it might be more advantageous to utilize the modeling techniques here with several various databases and a much larger age range to determine overall usefulness of the models presented.

4.3 Limitations

The models presented all have the major limitation of post-hoc usage. It is known which clips have seizures in the CHB-MIT scalp database which makes prediction the goal but not with the ideal of potential ambulatory usage in the future. Many studies, including this research, utilize statistical techniques of modeling or machine learning which have very little realistic applications. The DCC-GARCH and change point models utilized in this study both served very useful purposes with generally good results, but would likely not work well in realtime environment. On the other hand, models such as the Markov regime switching model have potential realtime application, but further additions would be needed. It is thus suggested that EEG seizure prediction research should always try and keep potential ambulatory usage in mind when doing their research.

The downfall of the data used in this research is that they are post-hoc based and we are left relying on the reliability of others. The actual seizure boundaries are noted by trained doctors, however that doesn't guarantee the placement of the seizures are exact. The data is also collected in a hospital, which doesn't reflect a potential ambulatory usage situation where the patients are free to go about their day while still being recorded. More noise to the data could be introduced with increased motion and daily activities with ambulatory usage, and the models used might not be wholly reflective of that. The CHB-MIT data is all scalp EEG based, and thus modeling as done in this thesis still lacks evidence of working with inter-cranial EEG recordings.

No knowledge of the patients used in he CHB-MIT database is available other than age or gender. We don't know any of the history or background of the patients, and thus we don't know the actual scope and severity of their seizures. We get some idea of the frequency of their seizures from their clips, but we are still left without a more complete history. It is also unknown if the patients suffer from any other ailments or diseases that could be at work. This is a downfall of any post-hoc analysis of readily available data, and thus research as such lacks some practicality until put into a realtime collection where more of the patient information is available and can be accounted for.

4.4 Conclusions

Three models for three clips where a seizure occurs have been examined in this research. In a multivariate analysis, the DCC-GARCH performed very well at modeling the 22 sensors simultaneously. The predicted variance of the sensors mapped well to the actual variance of the 22 sensors and supported the notion of

EEG sensor variance as a strong univariate tool. The Markov regime switching model also performed very well at assessing the individual point probabilities of the sensor variance in a Bayesian framework. These two relatively unused approaches to EEG modeling have both been shown to have interest for future work.

In an approach to predict the actual start and end of a seizure, a Bayesian change point model was used. With an ability to capture the beginning of the seizure quite well, the change point models also have some interest going forward. For use with discrete post-hoc data, the change point analysis might be beneficial with some added model specifications and modifications. This type of model could be extremely useful for locating the seizure boundaries when they may not be known.

Since the ideal going forward is to have online ambulatory use, the Markov regime switching model seems the most viable of the three models presented. To make the switching model applicable to online recording, a particle filtering algorithm will need to be put into place to keep the point probabilities continuously updating. The use of particle filtering is not new to EEG data acquisition, so it seems logical that the incorporation of modeling sensor variance through a Markov regime switching model might be a worthwhile next step.

The modeling techniques used in this research also highlight alternatives to the more commonplace machine learning methods utilized in EEG research. Modeling offers a more streamlined approach with realtime implications than machine learning does, and no training data is required for modeling. Some AR and/or MA time-series parameters could also be introduced to the Markov regime switching model to make it even more effective at modeling EEG sensor variance in terms of seizure prediction. These ARMA parameters can in turn help fine tune the particle filtering of the incoming data for use in ambulatory EEG modeling.

Ultimately, the goal is to utilize a mix of statistics, engineering, and computers to help prevent seizures before they occur in epileptic patients. A synergy will be needed between the three to make it viable and practical. While computers continue to process faster and engineering increases the ability to capture EEG in realtime, statistical modeling will be a necessity to actual locate the occurrence of seizure activity. While many methods for prediction exist, it should become the goal to pinpoint models that would viably work in an ambulatory scenario. The prediction of seizures without actual application to reality seems wasteful. Thus, statistical models for seizure activity should focus a little less on almost perfect predictive power and more so on strong sensitivity and specificity mixed with an actual application potential. The use of sensor variance and modeling it with a Markov regime switching model as demonstrated in this research, seems to be a good example of a solid mix of prediction and realtime use and both should be explored more.

List of References

[1] R. Prado and M. West, *Time series: modeling, computation, and inference.* CRC Press, 2010.

APPENDIX

Appendix A

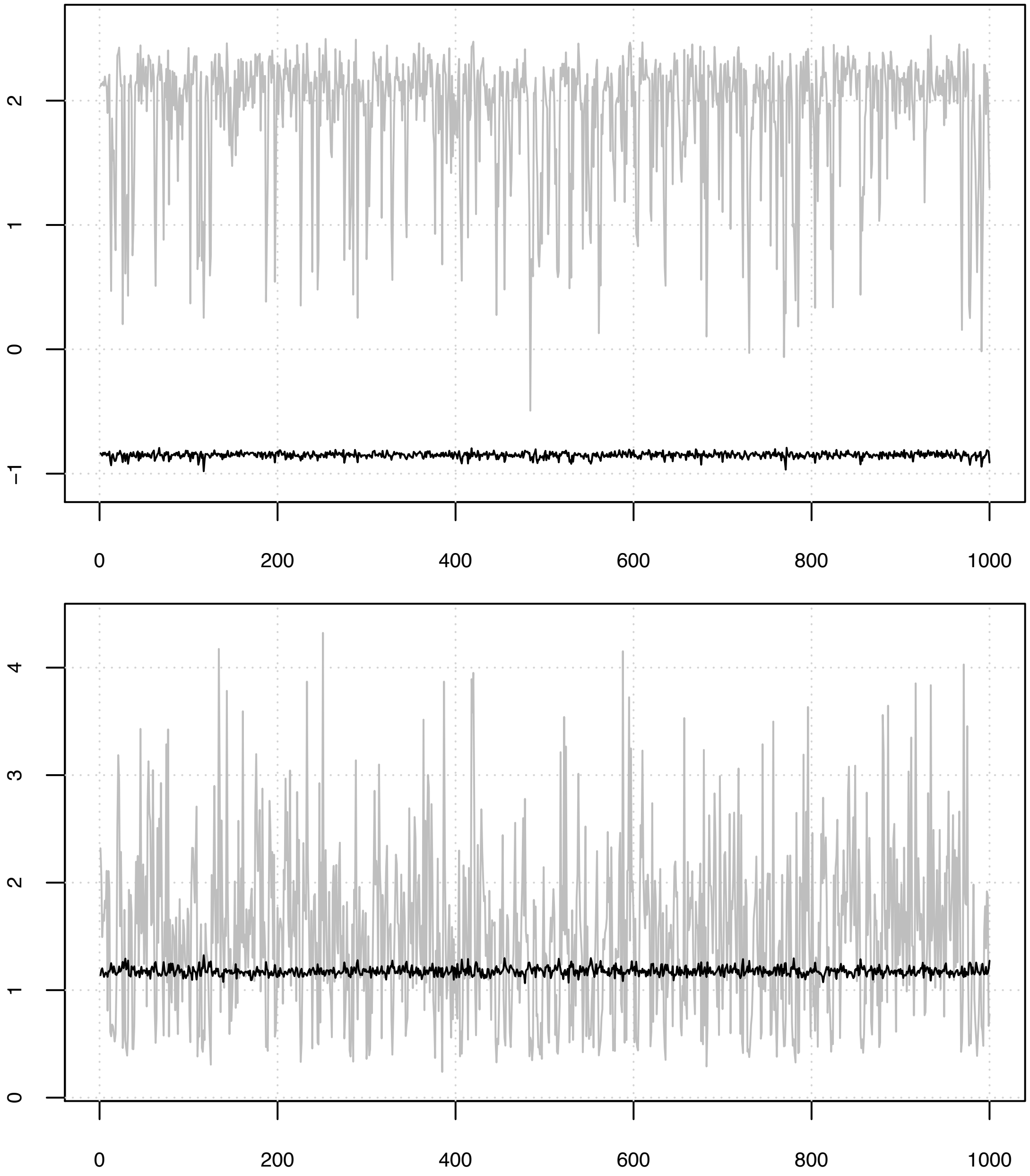

Figure A.1. Gibbs sampling trace plots for clip `chb05.13`. Traces for μ (top) and τ^2 (bottom) are displayed, with black lines representing the non-seizure components and the grey lines the seizure components.

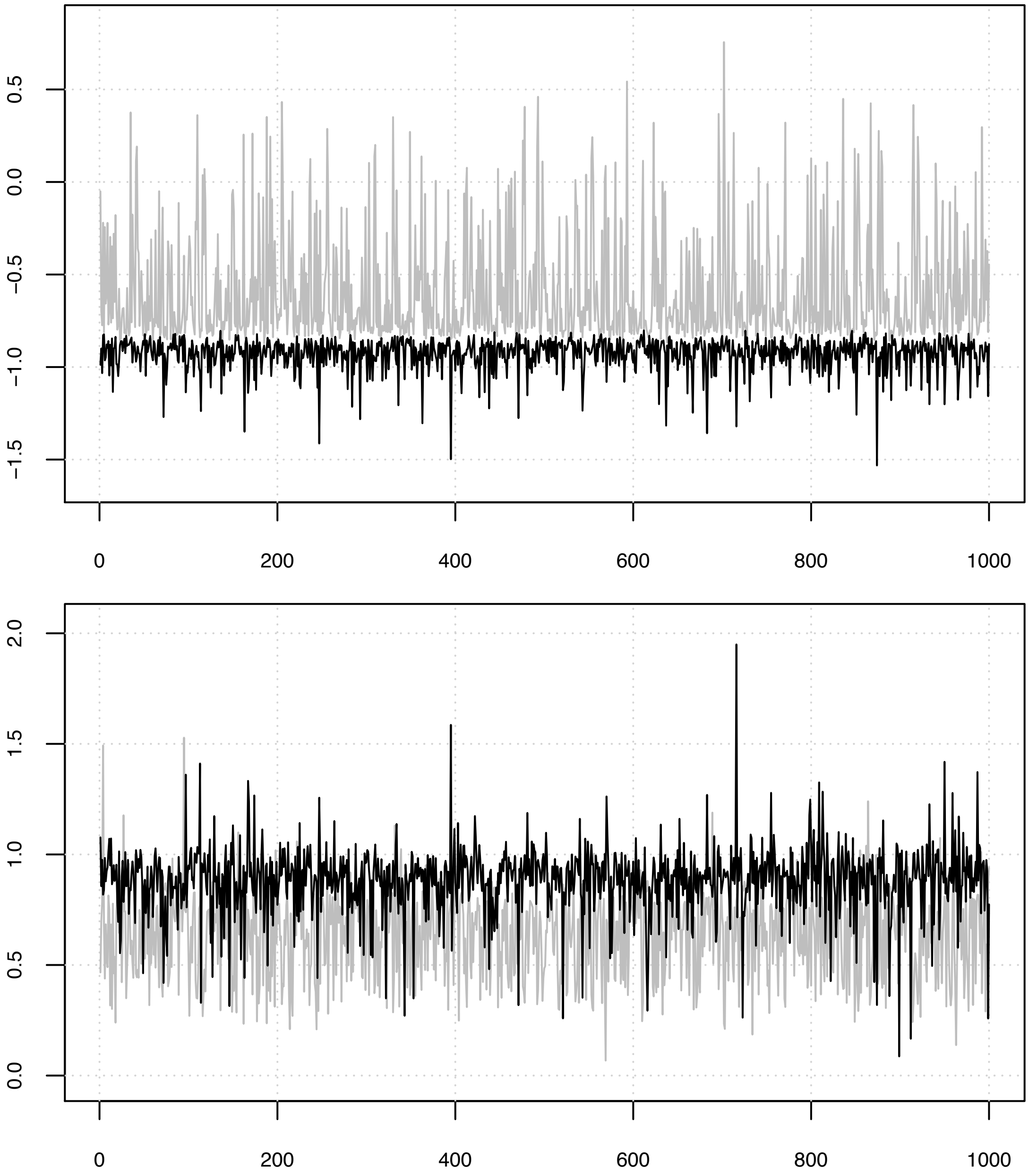

Figure A.2. Gibbs sampling trace plots for clip `chb08.02`. Traces for μ (top) and τ^2 (bottom) are displayed, with black lines representing the non-seizure components and the grey lines the seizure components.

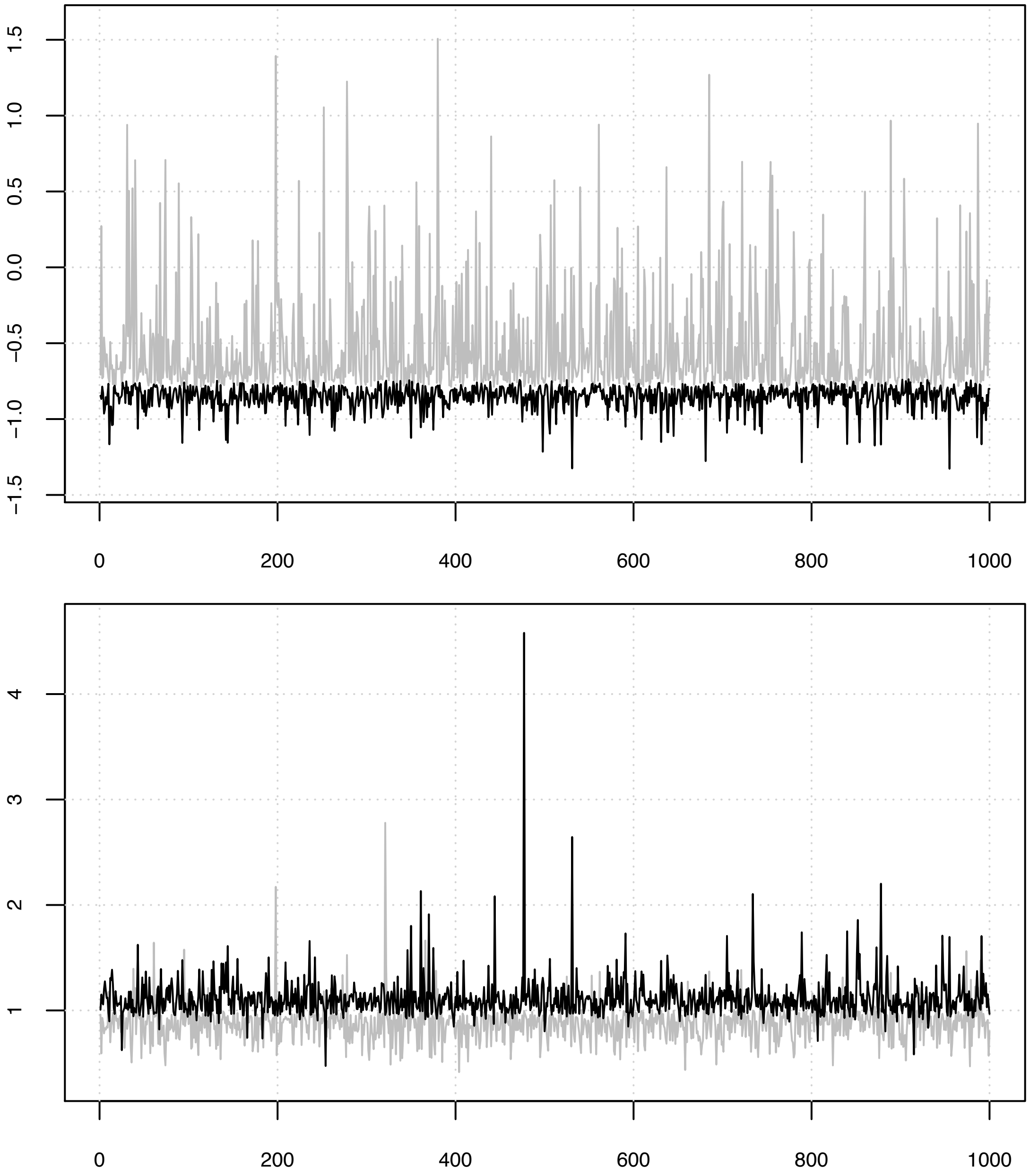

Figure A.3. Gibbs sampling trace plots for clip `chb22.38`. Traces for μ (top) and τ^2 (bottom) are displayed, with black lines representing the non-seizure components and the grey lines the seizure components.

BIBLIOGRAPHY

D. Barry and J. A. Hartigan. A bayesian analysis for change point problems. *Journal of the American Statistical Association*, 88(421):309–319, 1993.

W. G. Besio, K. Koka, R. Aakula, and W. Dai. Tri-polar concentric ring electrode development for laplacian electroencephalography. *Biomedical Engineering, IEEE Transactions on*, 53(5):926–933, 2006.

M. Billio, M. Caporin, and M. Gobbo. Flexible dynamic conditional correlation multivariate GARCH models for asset allocation. *Applied Financial Economics Letters*, 2(02):123–130, 2006.

T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.

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

G. Casella and E. I. George. Explaining the gibbs sampler. *The American Statistician*, 46(3):167–174, 1992.

J. Chiang and R. K. Ward. Energy-efficient data reduction techniques for wireless seizure detection systems. *Sensors*, 14(2):2036–2051, 2014.

S. Chib. Calculating posterior distributions and modal estimates in markov mixture models. *Journal of Econometrics*, 75(1):79–97, 1996.

J. Diebolt and C. P. Robert. Estimation of finite mixture distributions through bayesian sampling. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 363–375, 1994.

R. Engle. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20(3):339–350, 2002.

R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, pages 987–1007, 1982.

C. Erdman, J. W. Emerson, et al. bcp: an r package for performing a bayesian analysis of change point problems. *Journal of Statistical Software*, 23(3):1–13, 2007.

S. Frühwirth-Schnatter. Markov chain monte carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association*, 96(453):194–209, 2001.

S. Frühwirth-Schnatter. *Finite Mixture and Markov Switching Models: Modeling and Applications to Random Processes*. Springer, 2006.

A. Galka, K. Wong, and T. Ozaki. Generalized state-space models for modeling nonstationary EEG time-series. In *Modeling Phase Transitions in the Brain*, pages 27–52. Springer, 2010.

A. Galka, O. Yamashita, and T. Ozaki. Garch modelling of covariance in dynamical estimation of inverse solutions. *Physics Letters A*, 333(3):261–268, 2004.

R. L. Gilmore. American-electroencephalographic-society guidelines in electroencephalography, evoked-potentials, and polysomnography. *Journal of Clinical Neurophysiology*, 11(1):1–142, 1994.

A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley. Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000.

P. D. Hoff. *A first course in Bayesian statistical methods.* Springer, 2009.

P. Jiruska, M. de Curtis, J. G. Jefferys, C. A. Schevon, S. J. Schiff, and K. Schindler. Synchronization and desynchronization in epilepsy: controversies and hypotheses. *The Journal of physiology*, 591(4):787–797, 2013.

T. D. Johnson, R. M. Elashoff, and S. J. Harkema. A bayesian changepoint analysis of electromyographic data: detecting muscle activation patterns and associated applications. *Biostatistics*, 4(1):143–164, 2003.

H. Khammari and A. Anwar. A spectral based forecasting tool of epileptic seizures. *IJCSI International Journal of Computer*, 2012.

Y. U. Khan, O. Farooq, and P. Sharma. Automatic detection of seizure onset in pediatric EEG. *International Joural of Embeded Systems and Applications*, 2(3):81–89, 2012.

K. H. Lee and N. Verma. A low-power processor with configurable embedded machine-learning accelerators for high-order and adaptive analysis of medical-sensor signals. *Solid-State Circuits, IEEE Journal of*, 48(7):1625–1637, 2013.

S. Mihandoost, M. C. Amirani, and B. Z. Varghahan. A new approach for feature extraction of EEG signal using GARCH variance series. In *Application of Information and Communication Technologies (AICT), 2011 5th International Conference on*, pages 1–5. IEEE, 2011.

F. Mormann, T. Kreuz, R. G. Andrzejak, P. David, K. Lehnertz, and C. E. Elger. Epileptic seizures are preceded by a decrease in synchronization. *Epilepsy research*, 53(3):173–185, 2003.

P. Olejniczak. Neurophysiologic basis of EEG. *Journal of clinical neurophysiology*, 23(3):186–189, 2006.

L. Orosco, A. G. Correa, and E. Laciar. Review: A survey of performance and techniques for automatic epilepsy detection. *Journal of Medical and Biological Engineering*, 33(6):526–537, 2013.

R. T. Pivik, R. J. Broughton, R. Coppola, R. J. Davidson, N. Fox, and M. R. Nuwer. Guidelines for the recording and quantitative analysis of electroencephalographic activity in research contexts. *Psychophysiology*, 30(6):547–558, 1993.

R. Prado and M. West. *Time series: modeling, computation, and inference*. CRC Press, 2010.

F. Sharbrough, G. Chatrian, R. Lesser, H. Lüders, M. Nuwer, and T. Picton. American electroencephalographic society guidelines for standard electrode position nomenclature. *J. Clin. Neurophysiol*, 8(2):200–202, 1991.

A. Shoeb. *Application of Machine Learning to Epileptic Seizure Onset Detection and Treatment*. PhD thesis, Massachusetts Institute of Technology, September 2009.

K. F. K. Wong, A. Galka, O. Yamashita, and T. Ozaki. Modelling non-stationary variance in EEG time series by state space GARCH model. *Computers in biology and medicine*, 36(12):1327–1335, 2006.

World Health Organization, January 2015. http://www.who.int/mediacentre/factsheets/fs999/en/.

D. F. Wulsin. *Bayesian Nonparametric Modeling of Epileptic Events*. PhD thesis, University of Pennsylvania, 2013.