A Modeling Approach to Analyze Epileptic Brain Data Using Univariate and Multivariate Dynamic Linear Models

by Samrachana Adhikari

under the Direction of

Professor Janice Gifford and Professor Michael Lavine

A Thesis

Submitted to the Faculty of Mount Holyoke College

in Partial Fulfillment of the Requirements for the Degree of

Bachelor of Arts with Honors

Department of Mathematics and Statistics

Mount Holyoke College

South Hadley, MA 01075

May 2011

To the lady who's sure all that glitters is gold.

ACKNOWLEDGMENTS

ABSTRACT

Epileptic patients with severe cases of epilepsy usually undergo more than one surgery to locate and remove the epileptic tissue from the brain. It is hard to localize epileptic tissue as it is defined by abnormal excitability and synchronization of neurons, and does not have any distinguishing visual characteristics. During the first surgery a piece of cranium is removed and an array of EEG electrodes is placed on the surface of the brain, after which the patient is monitored for several days. During the second surgery, the brain is stimulated electrically and the result is monitored. Patients are traditionally monitored by EEG electrodes. However in the study done by Hochman and Haglund optical imaging technique is used to study the impact of the stimulus.

In this project, we use data from a study conducted by Haglund and Hochman. The brain is divided into different regions of interest and illuminated with light of the desired wavelength, which is then photographed at intervals of approximately .2 to .3 seconds. The data is the series of average intensities for each region recorded at different times. Interpreting this time series data is challenging because of the noise contributed by heartbeat and respiration. We use a Bayesian dynamic linear model (DLM) to remove the noise artifacts. We first develop a univariate DLM to analyze each region independently. We observe that some regions responded to stimulus in a similar manner. To account for the correlation between these regions

we further develop a multivariate DLM. These models help us to study the response of stimulus in noisy data, and will hopefully enable us to identify distinguishing properties of epileptic tissues in the brain.

Contents

8 Discussion **42**

9 APPENDIX **46**

List of Figures

1   Introduction

1.1   Optical Imaging Technique: An introduction

Epileptic patients with severe cases of epilepsy usually undergo two surgeries to locate the tissues responsible for origination and spread of seizure, and to remove these tissues from the brain. During the first surgery a piece of cranium is removed, and an array of electrodes is placed on the surface of the brain under the **dura mater**[1] to activate the neurons around the electrode. This process of activation of neurons by electrodes is called the **electrical stimulation** of the brain.[5] The patient is then monitored for several days. Second surgery involves the removal of the grid array and the definitive resection of the epileptogenic tissues from the brain.[2][3] The process of locating the tissues where epileptic seizures originate is called **localization** and the process of removal of these tissues from the brain is called **resection**. Tissues where the seizures originate are called **epileptogenic tissues**.

   The surgical outcomes of resection in patients suffering from **neocortical epilepsy**[2] are not as successful as the surgical outcomes from resections in epileptic patients with **mesial temporal epilepsy**.[3][3] It

---

[1]**Dura mater** is the outermost, toughest, and most fibrous of the three membranes covering the brain and the spinal cord.[2]

[2]**Neocortex** is the outer layer of cerebral hemisphere in mammalian brain. It is involved in higher functions such as sensory perception, generation of motor commands, spatial reasoning, conscious thought and language. The epilepsy that occurs in neocortex part of the brain is called **neocortical epilepsy**.

[3]**Mesial Temporal Epilepsy**: Mesial temporal lobe epilepsy is the most common form of human epilepsy, and usually occurs in hippocampal region of the brain.[6]

is hard to localize neocortical epileptogenic tissues accurately, as they are characterized by abnormal excitability and synchronization of neurons, and do not necessarily have any distinguishing visual characteristics.[9]

Traditionally, neuronal activity induced by electrical stimulation in the patients after first surgery is monitored by Electroencephalography (EEG) electrodes. However, in a study conducted by Hochman and Haglund, an **intra-operative imaging of optical signals technique**, discussed below, is used to monitor the stimulation in the patients during the surgery. (For more information regarding the technical aspects of the study refer to [3])

The amount of light absorbed and scattered by brain tissue changes in response to neuronal activity.[3][8][4] Optical responses evoked by neuronal activity are often called **intrinsic optical signaling (IOS)**. [8] These optical responses arise from changes in the optical properties of the tissue itself. Imaging of intrinsic optical signals (ImIOS) is a technique that involves illuminating neuronal tissue with light at various wavelengths and recording its IOS with an imaging detector. These activity evoked optical changes are generated by changes in **blood oxygenation** and **blood volume** in the brain, which can independently affect the amount of light absorbed by cortical tissues.[3][4]

As there is change in the density of the red blood cells with the change in the volume of blood, alterations in blood volume affect the optical properties or light absorbing properties of the blood. Similarly,

blood oxygenation, the level of oxygen saturation in the blood, affecting optical change is a consequence of difference in absorption spectra of oxyhemoglobin and deoxyhemoglobin.[1][8]When the blood is saturated with oxygen it is called oxyhemoglobin and when it is desaturated, it is called deoxyhemoglobin. Hochmann and Haglund pointed to evidences of the illumination of the cortex at specific wavelengths enabling ImIOS to selectively record changes in either blood volume or blood oxygenation.[3]

Light with 535nm wavelength detects changes in blood volume whereas that with 660nm detects the changes in blood oxygenation.[8] Optical change measured at 535nm is negative going. This is because as blood cells in neurons absorb more light due to increased activity, brain tissue becomes darker. Hence the intensity of IOS is lower with increased activity in the neurons. Conversely, optical change at 660 nm wavelength light is positive going. Optical changes recorded at this wavelength represent the change in ratio of oxyhemoglobin to hemoglobin.[3] Since oxyhemoglobin absorbs significantly less light than hemoglobin, at this wavelength, the tissue, and particularly the larger veins overlying active neocortex, becomes brighter during increased neuronal activity.[3] So intensity of optical imaging is higher with increased activity.

As visible and near-infrared lights are not harmful to cells and tissues, they are used for the purpose of illumination.[8] The data of the response of brain to external stimulus collected by optical imag-

ing technique could be a potential tool for localizing functional and epileptic cortical regions in operating rooms (intra-operative settings), and providing guidance for the resection of tissues during neurosurgical problems.[8][10]

Possible advantages of ImIOS compared to other brain mapping modalities, as discussed by Haglund and Hochman in their paper [3] are that ImIOS is an inexpensive technique, can provide significantly greater spatial and temporal resolution, and is capable of on-line intra-operative imaging. Consequently, ImIOS has the potential to become a powerful and widely applicable tool for both the clinical and surgical treatment of epilepsy.[3] However, this technique has been of limited use because the data obtained is noisy and hard to interpret. In this thesis, we develop a Bayesian model that will help us extract the response of brain to external stimulus from the noisy data and will hopefully be useful in distinguishing regions that behave abnormally.

1.2   Overview of the Thesis

In this thesis, I use data from a study conducted by Haglund and Hochman. Collection and structure of the data is explained in Chapter 2. In chapter 3 & 4, I talk about the specifics of univariate dynamic linear model used to model different regions independently and the methods used to estimate unknown parameters. Chapter 5 investigates the results obtained from univariate dynamic linear model. I continue my analysis with multivariate dynamic linear model in Chapter 5 where

there is correlation among regions that respond to stimulus in a similar way. Chapter 6 is followed by the results from the multivariate analysis in Chapter 7. I wrap up my thesis with a section on discussion in Chapter 8.

Lavine et al. has developed a Bayesian, univariate dynamic linear model for the analysis of human ImIOS data and verified its robustness in data sets derived from multiple replicates and cortical regions from multiple subjects in [8] and [9]. In this thesis I begin my analysis following Lavine et al.'s modeling approach and extend their model by introducing a multivariate dynamic linear model for regions that show similar dynamics. Dynamic linear models have also been used to model the brain imaging data of other animals by few other researchers. For example Myers, Brockwell and Eddy [10] used univariate dynamic linear model to study the optical imaging data from cats' brain.

All the analysis are done using statistical software R.[13] Package dlm in R provides excellent setting for running computationally complicated algorithms and estimating different components of dynamic linear model.[11] We use this package extensively for all of our computation. We use package ggplot2 and other graphics in R to illustrate our findings graphically.[14]

2   Data

For the purpose of the study, the brain was divided into different re-
gions of interest (as shown in the figure 1 below), and illuminated with
light of desired wavelength. It was then photographed at intervals of
approximately .2 to .3 seconds for about 200 seconds, with a total sam-
ple size of $t = 1015$. We will refer to each t as a time point when the
data was collected and 5 time points ≈ 1 second.
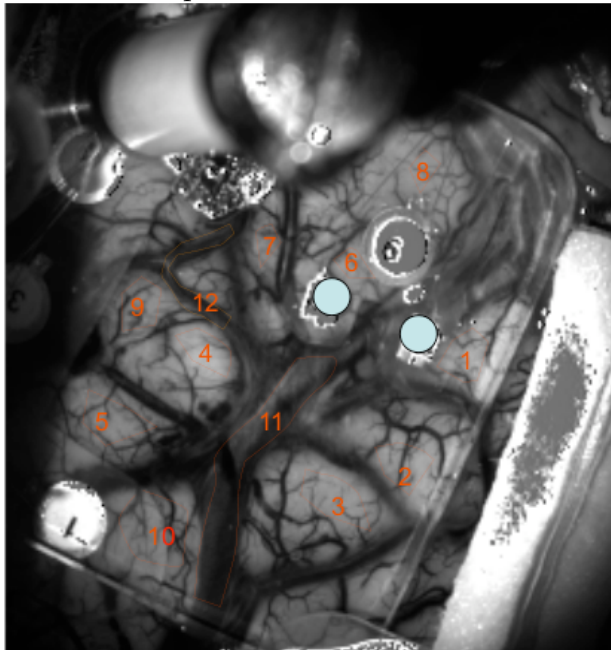
Figure 1: The image above is a human brain during surgery, placed under
the glass plate for stability. Several regions of the brain are numbered in
red. These are the regions from which data is collected. Most of the image
is gray matter overlain with a network of dark blood vessels.

The raw data is the series of intensities for each pixel recorded at
different time. We use the average of the intensities of all the pixels
in a region for a given time period. Since there are variations in the
illumination profile across the field of view, we are interested in relative

change of intensities across each region of interest. The average of the intensities within each region is divided by the regional average at time zero. So the data is the time series of average response of each region of interest of the brain to the stimulus and is denoted by y_t, where $t \geq 0$ is the time of recording of each intensity. This method is consistent with the method used by Lavine et al. in [8][9].

Thus the data we work with is:

$$y_t = \frac{\text{regional average intensity at time } t}{\text{regional average intensity at time } 0}$$

Interpreting this time series data of optical imaging is challenging because of the noises in the data contributed primarily by heartbeat and respiration.[3][10]
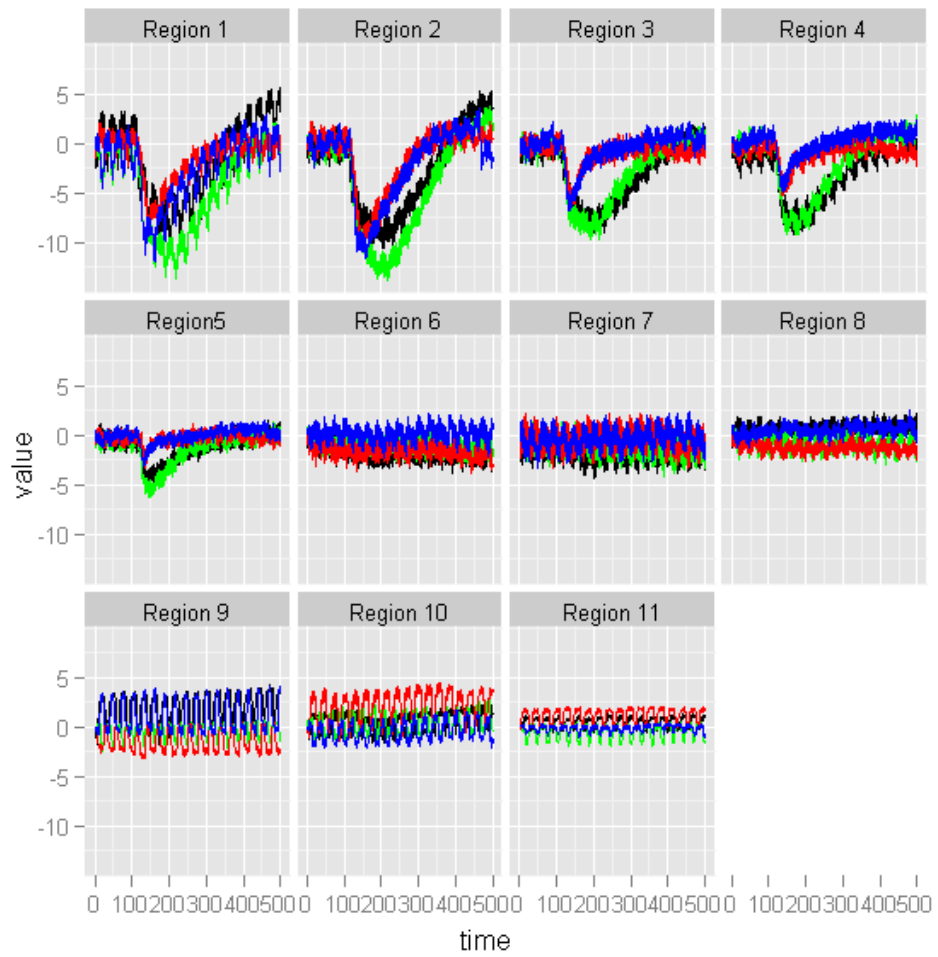
Figure 2: Graphical Representation of the time series of intensity of light recorded by optical imaging for regions 1-11 in patient 4. Different color represents different replicate of the study in each region. Along Y-axis is the intensty (y_t) recorded at each t and along X-axis is the time point t.
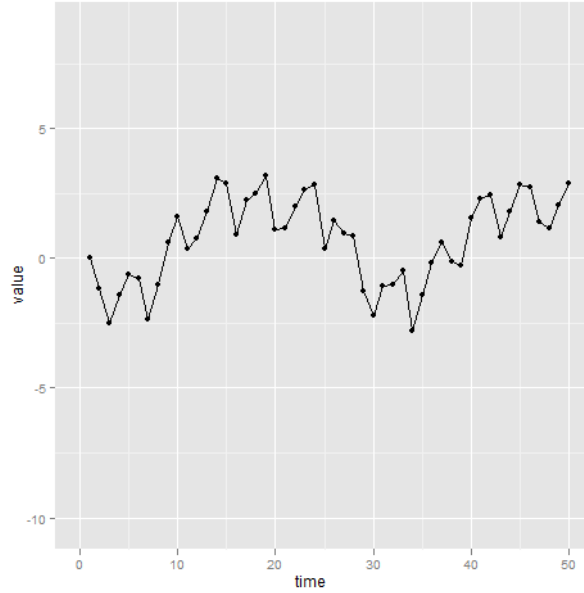
Figure 3: Region 1 in Higher Resolution

Figure 2 represents plots of the time series data for 11 regions of the brain of Patient 4 illuminated with 535nm light. For each region four different replicates of the study were conducted. Hence different colors in the plot represent data from different replicates of the study. From the figure we can observe the general patterns of response in the data. When stimulus is applied at time $\approx$ 100 in patient 4, the time series falls rapidly in regions 1-5. In Figure 3, we observe larger fluctuations at every 40 time points or so. These fluctuations are accounted by respiration. Also, in the same figure on a smaller time scale we observe fluctuations of one cycle per 5 time points or so. These smaller fluctuations are due to heartbeat. To study the response of the neurons to external stimulus, we need a method that allows us to extract the signal by accounting for the noise of heart beat and respiration.

18

3 Investigation I: Univariate Analysis of IOS data

Bayesian Dynamic Linear Model, a Gaussian linear sub-class of Space State model, is used to identify (separate out) the signal of the response of neurons to external stimulus by accounting for the cyclic component of heart beat and respiration. The model used is of the form:

$$y_t = s_t + h_t + r_t + v_t \tag{1}$$

where: s_t is a smooth curve for the response, h_t is for heart beat, r_t is for respiration and v_t is an error term.

In this chapter, I will briefly introduce State Space model, followed by an explanation of the properties of dynamic linear model, as a special case of State Space model. I will continue the chapter by talking about specific types of dynamic linear models used in our analysis and the results thus obtained.

3.1 State Space Models

State space models are used for modelling systems that evolve with time.[8] They are characterized by unobservable Markov chain θ_t, also referred to as state process or a latent process for, $t \in \{0, 1, 2, ..T\}$ and observed process y_t which is an imprecise measurement of θ_t.[12] The dependence structure of y_t and θ_t is a part of the definition of space

state models. Conditional on θ_t, the $y'_t s$ are independent and y_t depends on θ_t only. Therefore, state space model is completely specified by the initial distribution $\pi(\theta_o)$ and the conditional densities $\pi(\theta_t|\theta_{t-1})$ and $\pi(y_t|\theta_t)$.[12]

Definition 3.1. Markov Chain

$(\theta_t)_{t>1}$ is called a Markov chain if for any $t > 1$,

$$\pi(\theta_t|\theta_{1:t-1}) = \pi(\theta_t|\theta_{t-1}).$$

[12]

From definition 3.1 we see that conditional on θ_{t-1}, θ_t and θ_{t-2} are independent, which means the information carried about θ_t by θ_1 up to θ_{t-1} is same as that carried by θ_{t-1}.

3.2   Dynamic Linear Model

The dynamic linear model (dlm) is a special sub-class of State Space model. Like a state space model, it is characterized by the observable process $(Y_t : t = 1, 2, ...)$ which is thought of as determined by a latent process $(\theta_t : t = 1, 2, ...)$, up to Gaussian random errors. The observation Y_t depends only on θ_t at time t, and θ_t depends only on the previous position θ_{t-1}, through a linear relationship, up to Gaussian random errors.[12]

The dynamic linear model is specified by:

- A Normal prior distribution for the p-dimensional state vector at time $t = 0$, $\theta_0 \sim N_p(m_0, C_0)$,

- Observation equation of the form

$$Y_t = F\theta_t + v_t, \text{with } v_t \sim N_m(0, V_t) \text{ and,}$$

- State equation or evolution equation of the form:

$$\theta_t = G\theta_{t-1} + w_t, \text{with } w_t \sim N_p(0, W_t).$$

- G and F are constant known matrices of order $p \times p$ and $m \times p$ respectively. v_t and w_t are two sequences of independent Gausian random errors with mean zero and variance matrices V and W .

We assume that the time series at each region evolves independently. So we model each region separately by univariate dlm. To model a complicated system, like ours, we combine many simple dlms. We use a polynomial model of order 2 to fit the smooth component (response of the neurons to external stimulus), and a cyclic Trigonometric model to fit the cyclic components of heartbeat and respiration.

3.3 Local linear trend model for the Smooth Component

A polynomial dlm of order 2 is used to describe the smooth development of series over time. At each time t, the state of the system is represented by its level and slope, i.e.

$$\theta_{s,t} = (level_t, slope_t.)$$

Only $level_t$ and not $slope_t$ contributes to what we observe at time t. Also, error in $\theta_{s,t}$ comes from $w_{slope,t}$ (error in slope component). It means that the slope can change over time.

This model is specified by the system with an observation equation of the form

$$s_t = (1,0)^t \theta_{s,t} = level_t$$

and state (evolution) equation of the form

$$\theta_{s,t} = G_s \theta_{s,t-1} + w_{s,t}.$$

State equation is written as:

$$\theta_{s,t+1} = \begin{bmatrix} level_{t+1} \\ slope_{t+1} \end{bmatrix} = \begin{bmatrix} level_t + slope_t \\ slope_t \end{bmatrix} + \begin{bmatrix} 0 \\ w_{slope,t+1} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} level_t \\ slope_t \end{bmatrix} + \begin{bmatrix} 0 \\ w_{slope,t+1} \end{bmatrix}$$

with,

$$G_s = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} ; \mathbf{w}_{s,t} = \begin{bmatrix} 0 \\ w_{slope,t} \end{bmatrix} \text{ and, } w_s \sim N(0, W_s) \text{ where } W_s = \begin{bmatrix} 0 & 0 \\ 0 & W_{slope} \end{bmatrix}.$$

3.4   Trigonometric Model for Heartbeat and Respiration

Components of heartbeat and respiration are modeled by trigonometric dynamic linear model which is described below.

Consider a sinusoidal function with period p. It can be thought as a point moving around a circle in anti-clockwise direction.[8] Its location at time t is given by $(b \cos \alpha_t, b \sin \alpha_t)$ where b is the amplitude and α_t is its angle, and the change in angle in one time-step is $\delta = \frac{2\pi}{p}$. Then the position of the point at time $t + 1$ can be written as:

$$\begin{bmatrix} b \cos(\alpha_{t+1}) \\ b \sin(\alpha_{t+1}) \end{bmatrix} = \begin{bmatrix} b \cos(\alpha_t + \delta) \\ b \sin(\alpha_t + \delta) \end{bmatrix} = \begin{bmatrix} b(\cos \alpha_t \cos \delta - \sin \alpha_t \sin \delta) \\ b(\cos \alpha_t \sin \delta + \sin \alpha_t \cos \delta) \end{bmatrix}$$

$$= \begin{bmatrix} \cos \delta & -\sin \delta \\ \sin \delta & \cos \delta \end{bmatrix} \begin{bmatrix} b \cos \alpha_t \\ b \sin \alpha_t \end{bmatrix}.$$

Our dlm for heartbeat consists of this equation for sinusoidal func-

tion plus the additional noise. Heartbeat at time t can be represented by a two dimensional state vector, $\theta_{h,t} = (\theta_{h,1,t}, \theta_{h,2,t})^t = (b_t \cos \alpha_t, b_t \sin \alpha_t)^t$. The state equation is of the form

$$\theta_{h,t+1} = G_h \theta_{h,t} + w_{h,t+1}.$$

G_h is a rotation matrix and

$$G_h = \begin{bmatrix} \cos \delta_h & -\sin \delta_h \\ \sin \delta_h & cos\delta_h \end{bmatrix}$$

where, δ_h is a heartbeat frequency given by $\delta_h = \frac{1}{2\pi \times \text{seconds per heartbeat}}$, $w_{h,t} = (w_{h,1,t}, w_{h,2,t})^t$ is an error term with $w_{h,t} \sim$ iid $N(0, W_h I_2)$, and I_2 is an identity matrix of dimension 2×2. G_h ensures that the expected shape of heartbeat cycle is sinusoid of frequency δ_h, and $w_{h,t}$ allows for each cycle to deviate in shape, amplitude and duration. Also note that at time t we observe only $h_{1,t}$ and not $h_{2,t}$.

In figure 3, we observed that the respiration had slighly longer sine waves and more fluctuations than heart beat. So, for modeling respiration component, we use a slightly more complicated and flexible family of shapes in trigonometric dlm. We use a trigonometric model with two harmonics.[8]

Observation equation for respiration is $r_t = (\theta_{r,1,t}, \theta_{r,3,t})^t$, with state

equation of the form,

$$\theta_{r,t+1} = \begin{bmatrix} \theta_{r,1,t+1} \\ \theta_{r,2,t+1} \\ \theta_{r,3,t+1} \\ \theta_{r,4,t+1}) \end{bmatrix} = \begin{bmatrix} b_{1,t+1}\cos\alpha_{1,t+1} \\ b_{1,t+1}\sin\alpha_{1,t+1} \\ b_{2,t+1}\cos\alpha_{2,t+1} \\ b_{2,t+1}\sin\alpha_{2,t+1} \end{bmatrix}$$

$$= \begin{bmatrix} \cos\delta_r & -\sin\delta_r & 0 & 0 \\ \sin\delta_r & \cos\delta_r & 0 & 0 \\ 0 & 0 & \cos 2\delta_r & -\sin 2\delta_r \\ 0 & 0 & \sin 2\delta_r & \cos 2\delta_r \end{bmatrix} \begin{bmatrix} b_{1,t}\cos\alpha_{1,t} \\ b_{1,t}\sin\alpha_{1,t} \\ b_{2,t}\cos\alpha_{2,t} \\ b_{2,t}\sin\alpha_{2,t} \end{bmatrix} + \begin{bmatrix} w_{r,1,t+1} \\ w_{r,2,t+1} \\ w_{r,3,t+1} \\ w_{r,4,t+1} \end{bmatrix}.$$

At time t, we observe $r_t = b_{1,t}\cos\alpha_{1,t} + b_{2,t}\cos\alpha_{2,t}$ and the $w's$ are mean zero Gaussian error with variance matrix W_r.

Thus our model for each region is an eight parameter model with two parameters for smooth component, two parameters for heartbeat, and four for respiration. The state vector is an eight dimensional θ_t such that

$$\theta_t = (level_t, slope_t, \theta_{h,1,t}, \theta_{h,2,t}, \theta_{r,1,t}, \theta_{r,2,t}, \theta_{r,3,t}, \theta_{r,4,t})^t$$

and the state and observation equations for θ_t are given by:

$$\theta_{t+1} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \cos \delta_h & -\sin \delta_h & 0 & 0 & 0 & 0 \\ 0 & 0 & \sin \delta_h & \cos \delta_h & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cos \delta_r & -\sin \delta_r & 0 & 0 \\ 0 & 0 & 0 & 0 & \sin \delta_r & \cos \delta_r & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cos 2\delta_r & -\sin 2\delta_r \\ 0 & 0 & 0 & 0 & 0 & 0 & \sin 2\delta_r & \cos 2\delta_r \end{bmatrix} \theta_t + \begin{bmatrix} 0 \\ w_{slope,t+1} \\ w_{h,1,t+1} \\ w_{h,2,t+1} \\ w_{r,1,t+1} \\ w_{r,2,t+1} \\ w_{r,3,t+1} \\ w_{r,4,t+1} \end{bmatrix}$$

$$y_t = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix} \theta_t + v_t = s_t + h_t + r_t + v_t.$$

4 Estimation and Forcasting

4.1 Estimation of Unknown Variances

Variances (W and V) used in the dynamic linear models are not usually known. Different Bayesian methods like Markov Chain Monte Carlo, modern sequential Monte Carlo, and non-Bayesian methods like maximum likelihood estimation can be used to estimate these unknown parameters.

We use maximum likelihood estimation (MLE) method to estimate unknown variance parameters in our DLMs. Let Λ denote the vector of unknown variance parameters. These parameters are constant over time for our model.[12] Also, let $y_1, ..., y_n$ be the vector of observed random time-series. Joint density of the observed values given particular value of Λ is given by

$$lik(\Lambda, y_1, ..., y_n) = \pi(y_1, ..., y_n; \Lambda) = \prod_{t=1}^{n} \pi(y_t|y_{t-1}; \Lambda)$$

where, $\pi(y_t|y_{t-1}; \Lambda)$ is the conditional density of y_t given the data up to time $t-1$ and Λ as particular value of the unknown parameter. $\pi(y_t|y_{t-1}; \Lambda)$ are Gaussian densities with mean and variance that depend on Λ.[12][10] Thus the likelihood function is numerically maximized to obtain the maximum likelihood estimates of the parameters,i.e.

$$\hat{\Lambda} = \arg\max_{\Lambda} LogLik(\Lambda, y_1, ..., y_n).$$

While computational time required for MLE is dramatically less than that for the Bayesian approaches, the numerical algorithm for computing the MLE might be sensitive to starting values and might not always converge.[12] To make sure that the MLE converges to one value (global maxima) we tried different starting values for the parameters. We also used simulated annealing algorithm for MLE which took comparatively longer to execute, to make sure that estimates of variances are global maxima. Simulated annealing introduces stochasticity to the optimization algorithm, thus making sure that parameters converge to global maxima and do not get stuck in the local mode.[7]

For the computation of MLE and simulated annealing we use dlmMLE function built in package dlm in R.[11]

Estimation of frequency for heart beat and respiration is done by visual inspection. The number of peaks in large interval of time from the plots are counted and then we compute the number of cycles per each time point for both respiration and heart beat. In figure 3 notice that between time points 0 and 50 there are approximately 2 respiration maxima. So, the frequency of respiration cycle is estimated as 2 per 50 times points or $\frac{1}{5}$ cycles per second. Similarly, for heart beat the frequency is estimated as 5 cycles per 20 time points or simply, $\frac{5}{4}$ cycles per second.[1]

---

[1]Remember that 5 time points are approximately equal to 1 second, since around 5 images were recorded every second.

4.2 Filtering and Prediction

Our interest in this analysis lies in estimating states of different components of the dlm given the observed data. To estimate the state vector we compute the conditional densities $\pi(\theta_s | y_{1:t})$ where $s \leq t$, also called smoothing. For smoothing all the information provided by $y_1, ..., y_t$ is used to estimate the state equation at each time $1, ..., t$. Smoothed estimates help us to study behavior of the system underlying the observations. Smoothing is implemented in package dlm by recursive algorithm called Kalman filter.[12] Kalman filter provides the formula for updating our current inference on state vector as new data is available.

5   Results: Investigation I

We use the univariate model described in Chapter 3 for accurate estimation of the components of smooth response, heartbeat and respiration in each of the regions independently. We first use MLE to estimate the unknown variance parameters in the model. Different starting values are used for MLE until the variance parameters converge to constant values and we use the variance parameter with greater likelihood to estimate different components of the model. We do this for all four replicates of the study in each region.

Mean level$_t$ estimated at each time t along with estimated mean response to heartbeat, $h_{1,t}$, and estimated mean response to respiration, $r_{1,t} + r_{3,t}$, from each replicate of the study is summarized graphically in figures in the Appendix section. In the figures, the green line is the observed value at each time point. The blue solid line represents the level estimated by dlm along with the red line for estimated heartbeat response and the purple for estimated respiration response. The magnitudes of optical responses in the estimated level curves, and their shapes after the initial response seem to somewhat accurately match the observed data at most of the time points for each region.

However, the figure also displays inadequacies in our modeling approach. The smooth level for most of the regions in all the replicates is smoother than what it appears on observed data at intervals where the intensity is minimum and the slope of the intensity changes from negative to positive. (This trend is very clearly seen in figures 7 and 8.)

This suggests that the estimated levels do not account for the sudden steep changes in the intensity. The estimated level appears to start changing smoothly prior to the true response in the raw data and does not go as deep as the observed intensities. Also, note that the smooth levels in most of the regions for each replicate of the data, except at regions 1, 2 and 3 in replicate 1, do not start at zero and slope of the estimated level at starting point is steeper than in the raw data for these regions. This result is contradictory to what we observed in the raw data. In the raw data we had normalized the observed intensity to make sure that the starting intensity[2] was constant in each region and that the change in intensities for these regions was comparable. The plots show that this model were not capable of capturing the sharp, sudden change in the response.

Moreover, in most of these regions the estimated respiration signal shows a big fluctuation for some starting time period and then goes to being stable. It seems like the respiration signal is following the path of the smooth response signal in these regions or there exist some mysterious (weird) component that is not being accounted by the model. While this is an interesting and important topic to pursue, we leave it for future investigation and proceed to check our investigation of smooth signal with different modeling approach.

We also estimate the variance parameters using simulated annealing to make sure that the estimates converge to global maxima and com-

---

[2]This is the intensity at the beginning of the stimulation of brain as explained in the data section.

pare the parameters estimated by these two methods. The parameters estimated by both of these methods were almost similar. We proceed by using MLE estimates estimated by MLE algorithm as it less time computationally than simulated annealing algorithm.

One way to make sure that the estimated levels follow the sudden steep changes in the data would be to consider long tailed T distribution of error that allows for more variation.[9] This is also computationally challenging and we need to use MCMC for estimating variance by this method.[8][9] The other approach could be to allow for correlation among regions that show similar dynamics and use multivariate dlm to model them. We use the latter procedure to check whether allowing for correlation among the regions give us more information about the smooth response to the stimulus.

6 Investigation II : Multivariate Analysis of IOS Data

6.1 Correlation of errors for slope of local linear trend model

From the plots of smooth component in the figures in Appendix I, we observe that regions 1 to 5 respond to the stimulus in a similar manner. Since there is a similar dynamic in smooth component (θ_s) we suspect that the slope errors in θ_s are correlated for these regions. We test for correlation by calculating Pearson's correlation (refer table 1), and by plotting the errors of each region against each other (refer to the plots in Figure 4).

From table 1 we see that the correlation coefficients for regions 1 to 5 with each other is fairly high. Pearson's correlation coeffiecients between regions 1&2, 1&3, 2&3 and 3&4 range from 0.86 to 0.9. Similarly, the coefficient between regions 1&4, 2&4 and 4&5 are between 0.73 and 0.79. The correlation coefficients between regions 1&5, 2&5 and 3&5 are the lowest, ranging from 0.49 to 0.6. The plots in figure 4 display the correlation between the regions graphically.

We need to build a model that allows us to incorporate the correlation of slopes in our analysis. We believe that this approach will improve our understanding of reponse of tissues evoked by electric stim-

	Region1	Region2	Region3	Region4	Region5
Region1	1	0.9	0.89	0.79	0.51
Region2	0.9	1	0.86	0.79	0.49
Region3	0.89	0.86	1	0.88	0.6
Region4	0.79	0.79	0.88	1	0.73
Region5	0.508	0.49	0.6	0.73	1

Table 1: Pearson Correlation Coefficient of Slope Error for each Region

ulation in different parts of the brain more accurately and hence detect any abnormality in these tissues that might exist.

Next, we proceed to multivariate analysis of these timeseries for regions 1 to 5 by building a multivariate dlm.
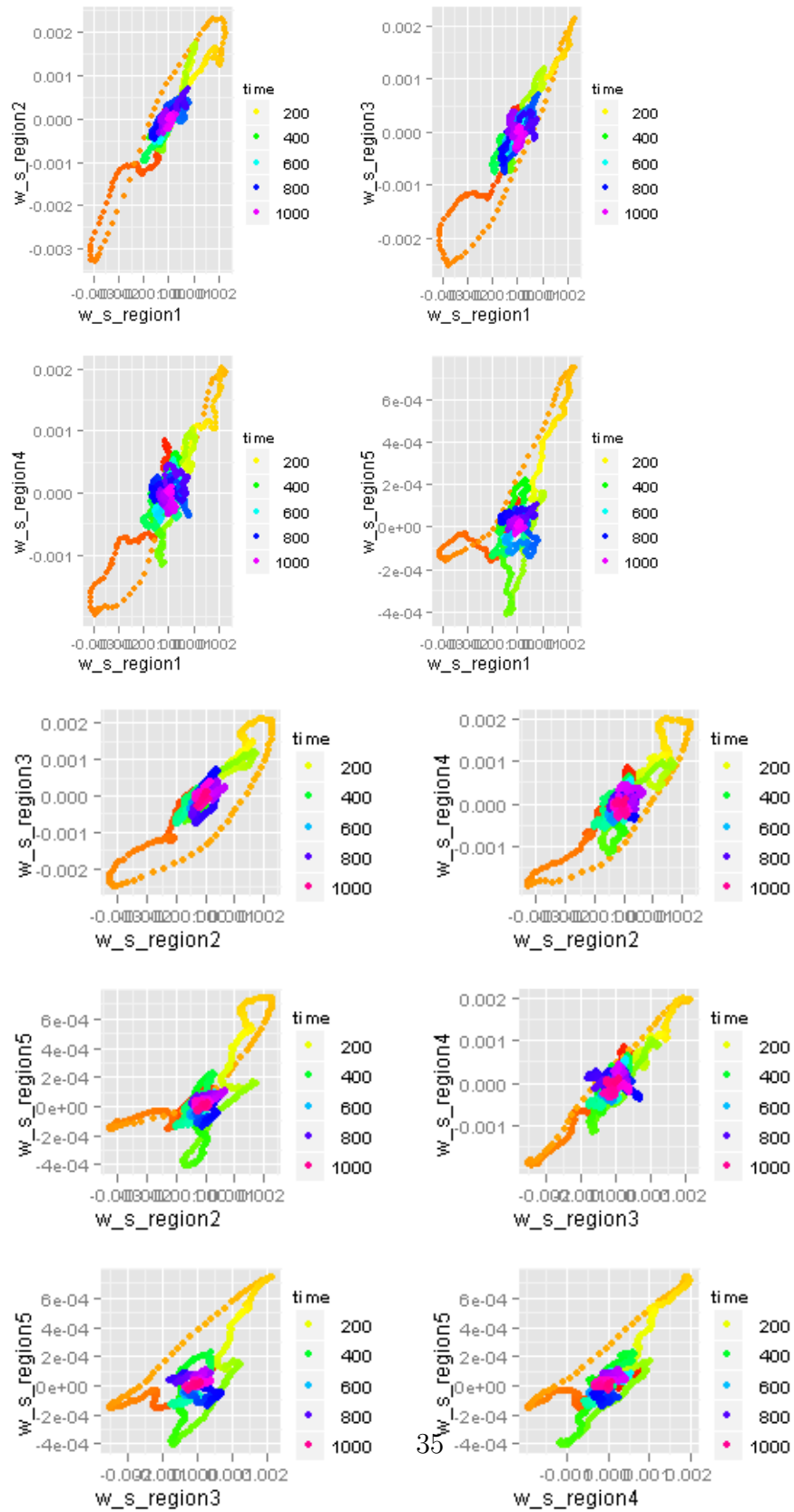
Figure 4: Scatterplot of Correlation of Errors

6.2 Multivariate DLM: Seemingly Unrelated Time Series Equations

We build a single model to analyze different regions that show similar dynamics by using seemingly unrelated time series equations (SUTSE). SUTSE are a class of models which specify the dependence structure among the state vectors.[12] This model is based on an assumption that all time series follow the same type of dynamics, and that the components of the state vectors have similar interpretations across the different DLMs.[12] We combine the individual DLMs that we created in the earlier section in order to build a multivariate model. We assume that the evolution of the slope of the state vector of smooth component is driven by correlated inputs. This means at any given time points the components of the system error corresponding to different slopes are correlated.

For the smooth component we can describe the joint evolution of the state vectors by grouping together all the levels and then all the slopes in an overall state vector

$$\theta_{s,t} = (level_{1,t}, ..., level_{n,t}, slope_{1,t}, ..., slope_{n,t})^t$$

where n is the number of regions included in the multivariate model.

6.3 Example of Multivariate Model for Time Series with Correlated Slope Errors

For the purpose of simplicity consider a multivariate model with 2 regions. Then the evolution equation at time $t + 1$ for smooth signal becomes

$$\begin{bmatrix} level_{1,t+1} \\ level_{2,t+1} \\ slope_{1,t+1} \\ slope_{2,t+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} level_{1,t} \\ level_{2,t} \\ slope_{1,t} \\ slope_{2,t} \end{bmatrix} + w_{s,t+1}$$

where $w_{s,t+1} \sim N(0, W_s)$ with

$$W_s = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & W_{1,s} & W_{1,2,s} \\ 0 & 0 & W_{2,1,s} & W_{2,s} \end{bmatrix}$$

$W_{1,s}$ is the variance of slope error of region 1, $W_{2,s}$ is the variance of slope error of region 2 and $W_{1,2,s} = W_{2,1,s}$ is the coviarance of slope error of these two regions.

We retain our assumption that the components for heart beat and respiration in different regions arise independently. The state vector

for heartbeat is

$$\theta_{h,t} = (\theta_{1,h,t}, \theta_{2,h,t})^t = (b_{1,t}cos\alpha_{1,t}, b_{2,t}cos\alpha_{2,t}, b_{1,t}sin\alpha_{1,t}, b_{2,t}sin\alpha_{2,t})^t$$

with evolution equation

$$\theta_{h,t+1} = G_h\theta_{h,t} + w_{h,t+1} = \begin{bmatrix} b_{1t}cos(\alpha_{1t} + \delta_{1h}) \\ b_{2t}cos(\alpha_{2t} + \delta_{2h}) \\ b_{1t}sin(\alpha_{1t} + \delta_{1h}) \\ b_{2t}sin(\alpha_{2t} + \delta_{2h}) \end{bmatrix} + w_{h,t+1}$$

$$= \begin{bmatrix} cos\delta_{1h} & 0 & -sin\delta_{1h} & 0 \\ 0 & cos\delta_{2h} & 0 & -sin\delta_{2h} \\ sin\delta_{1h} & 0 & cos\delta_{1h} & 0 \\ 0 & sin\delta_{2h} & 0 & cos\delta_{2h} \end{bmatrix} \begin{bmatrix} b_{1t}cos\alpha_{1t} \\ b_{2t}cos\alpha_{2t} \\ b_{1t}sin\alpha_{1t} \\ b_{2t}sin\alpha_{2t} \end{bmatrix} + w_{h,t+1}$$

where $w_{h,t}$ is normal with mean 0 and variance W_h and

$$W_h = \begin{bmatrix} W_{1,h,1} & 0 & 0 & 0 \\ 0 & W_{2,h,1} & 0 & 0 \\ 0 & 0 & W_{1,h,2} & 0 \\ 0 & 0 & 0 & W_{2,h,2} \end{bmatrix}.$$

Similarly, for modeling respiration with bivariate dlm we combine two univariate dlms with the state vector

$$\theta_{r,t} = (\theta_{1,r,t}, \theta_{2,r,t})^t$$

where, $\theta_{1,r,t}$ is the state vector from region 1 and $\theta_{2,r,t}$ is the state vector from region 2.

We use similar technique to build a dlm for modeling more than 2 regions together. The results obtained from a multivariate analysis are discussed in following section.

7 Results: Investigation II

We begin our multivariate analysis by considering different pairs of regions in each replicate. The results from the analysis are presented in Appendix II. We then build a model allowing for correlations among three regions at once. Result by allowing correlation among regions 2, 3 and 4 in data from replicate 1 is displayed in the figure 11 in Appendix II.

From the plots, we see that the multivariate model has improved our ability to model response signal as compared to the univariate model. The multivariate model is able to successfully capture the sudden deep changes in intensity that we could not capture in univariate model. Similarly, multivariate model is also accurate in initial state of the stimulus by being nearly stable at starting points before responding to the stimulus. This is an exciting finding. Because the multivariate model is using the information provided by all the regions included in the model to estimate the unknown variance parameters, we believe that these estimates are more accurate than those in univariate analysis.

This model is computationally more complicated than univariate model and estimation takes longer time. Ideally we would like to include all five regions in our analysis. However, we had difficulty with figuring out the MLE algorithm which kept on failing. It will be helpful to use more rigorous algorithms on a powerful machine for estimation like

this. The other issue with our results from multivariate analysis is that the weird component was still present in these new estimates. So it might be a good idea to include a component for mysterious element in our dlm for the analysis.

8   Discussion

This study addresses the issues that arise in estimating the smooth component from a noisy dataset, where the source of the noise is known and the noises have biological significance. Using dlm we were able to separate out the response of the brain to external stimulus from the response to heartbeat and respiration. While univariate dlm was able to estimate the smooth signal, allowing for correlation in the regions that show similar dynamics enabled us to estimate the smooth signal even more accurately. In that regard, dlm is a flexible model, which allows us to incorporate different aspects of the system.

However, there is a tradeoff between time taken to run the algorithm for maximizing variance estimates and accuracy of the estimates in implementing these two methods. For univariate dlm, we need to estimate only four unknown variances. So the computation time taken by dlmMLE to optimize the likelihood function is comparatively less. But for multivariate dlm, unknown variances and co-variances for all the regions for which we have allowed dependence needs to be estimated. And the time taken to execute the MLE is a lot longer than that for univariate model. Also, for some regions MLE of variances were sensitive to starting values and we kept on providing different values until the estimates converge. This might not be an effective method when time is a concern. Computation in simulated annealing took even longer than it did in MLE. (For univariate dlm, MLE algorithm took around 15 minutes to run while simulated annealing took around an

hour.) The time taken for executing the algorithm for multivariate analysis was even longer (more than 9 hours). While MCMC can be an alternative method for estimating variance, it might not be feasible in an intra-operative setting, that we eventually aim to use in, as it is computationally complicated and timely.

From our analysis we saw that allowing for correlation among regions that behave in a similar way improved our ability to accurately estimate smooth signal a lot. The results from multivariate analysis followed the raw data more accurately. We did multivariate analysis by allowing for correlation between two regions each at a time and among three regions at a time. We were also interested in modeling all five regions together. However, due to time constraints and requirement of large computational memory and space, we left it for future investigation. The other challenge lies in using this data in detecting any abnormal nature of the tissues so that we can localize epileptic tissues in the brain. We used the average of the intensities of all the pixels in a region for a given time period instead of using a data for each pixel in our study. By focusing in developing a model with average intensity value for each region we believe that we can replicate the model to study the time series from an individual pixel. This will help us in studying the behavior of the tissues in even smaller part of the brain for each region. Once we are successful in estimating the signals of response to stimulus for each pixel or pixels that show similar dynamics, we believe that we will be able to detect the abnormalities more accurately.

References

[1] Sonya Bahar, Minah Suh, Mingrui Zhao, and Theodore H. Schwartz. Intrinsic optical signal imaging of neocortical seizures: the epileptic dip, 2006.

[2] H. Blume. *The Surgical Treatment of Epilepsy*. Academic Press, San Diego, CA, 1997.

[3] Michael M. Haglund and Daryl W. Hochman. Optical imaging of epileptiform activity in human neocortex. *Epilepsia*, 45:43–47, 2004.

[4] Michael M. Haglund and Daryl W. Hochman. *Optical Imaging Techniques in Neocortical Epilepsy*. Thieme Medical Publishers, Inc., New York, NY, 2010.

[5] SMark H. Histed, Vincent Bonin, and R. Clay Reid. Direct activation of sparse, distributed ppulations of cortical neurons by electrical microstimulation. *Neuron*, 63(4):508–522, 2009.

[6] J. Engel Jr. Mesial temporal lobe epilepsy: What have we learned, 2001.

[7] S. Kirkpatrick, C.D. Gelatt, and M.P. Vecchi. Optimization by simulated annealing. *Science, New Series*, 220(4598):671–680, 1983.

[8] Michael Lavine, Michael M. Haglund, and Daryl W. Hochman. Dynamic linear model analysis of optical imaging data acquired from the human neocortex. 2008.

[9] Michael Lavine, Michael M. Haglund, and Daryl W. Hochman. Optical images of the brain during surgery: Detection of hemodynamics in response to electrical stimulus. 2008.

[10] Kary L. Myers, Anthony E. Brockwel, and William F. Eddy. State space models for optical imaging. *Stat MEd.*, 26:62–74, 2007.

[11] Giovanni Petris. An R package for dynamic linear models. *Journal of Statistical Software*, 36(12):1–16, 2010.

[12] Giovanni Petris, Sonia Petrone, and Patrizia Campagnoli. *Dynamic Linear Models with R.* useR! Springer-Verlag, New York, 2009.

[13] R Development Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2008. ISBN 3-900051-07-0.

[14] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis.* Springer New York, 2009.

9 APPENDIX

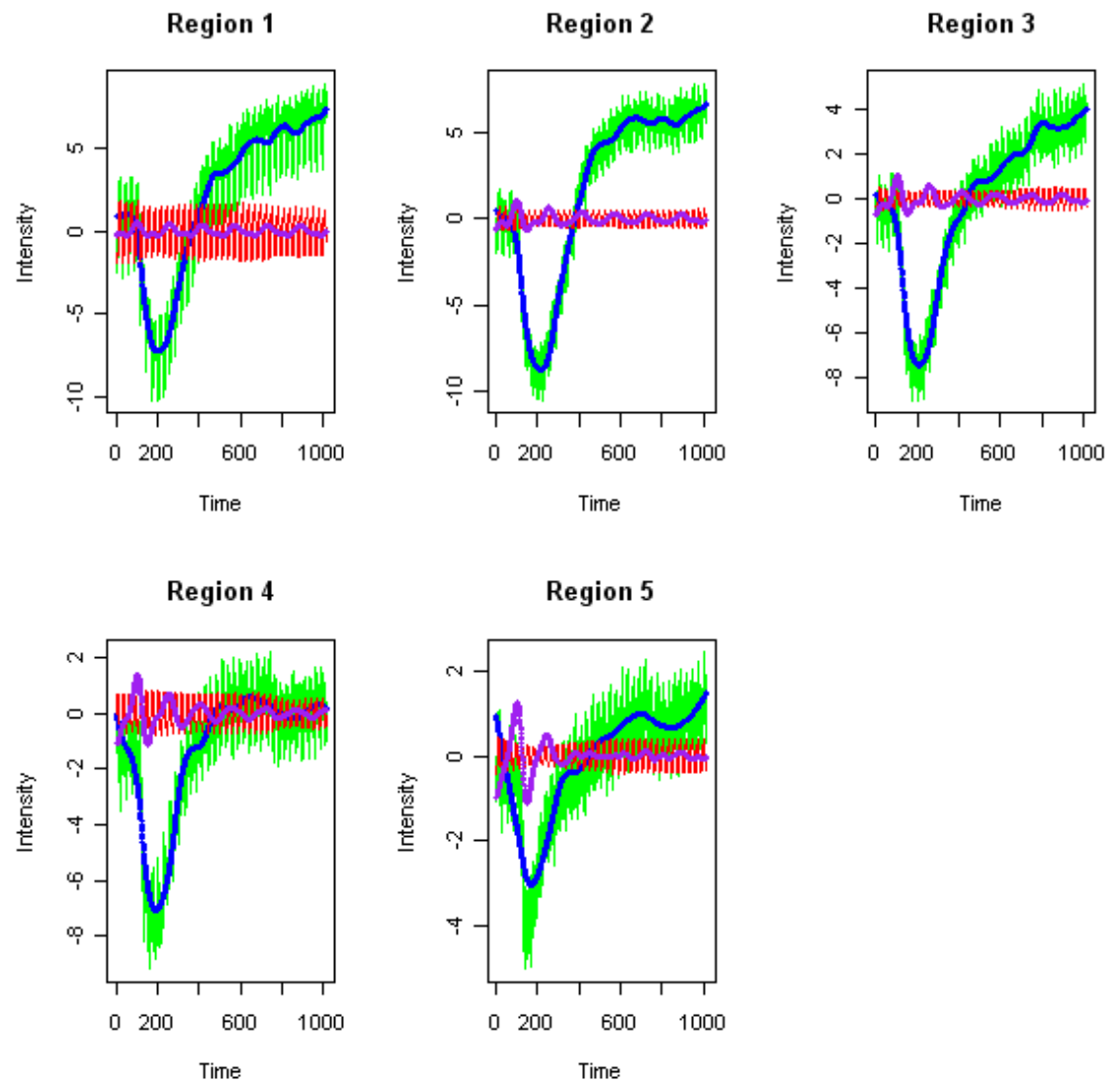## 9.1 APPENDIX I: Plots from Univariate Analysis

Figure 5: Observed Data and Estimated Smooth Level, Heartbeat and
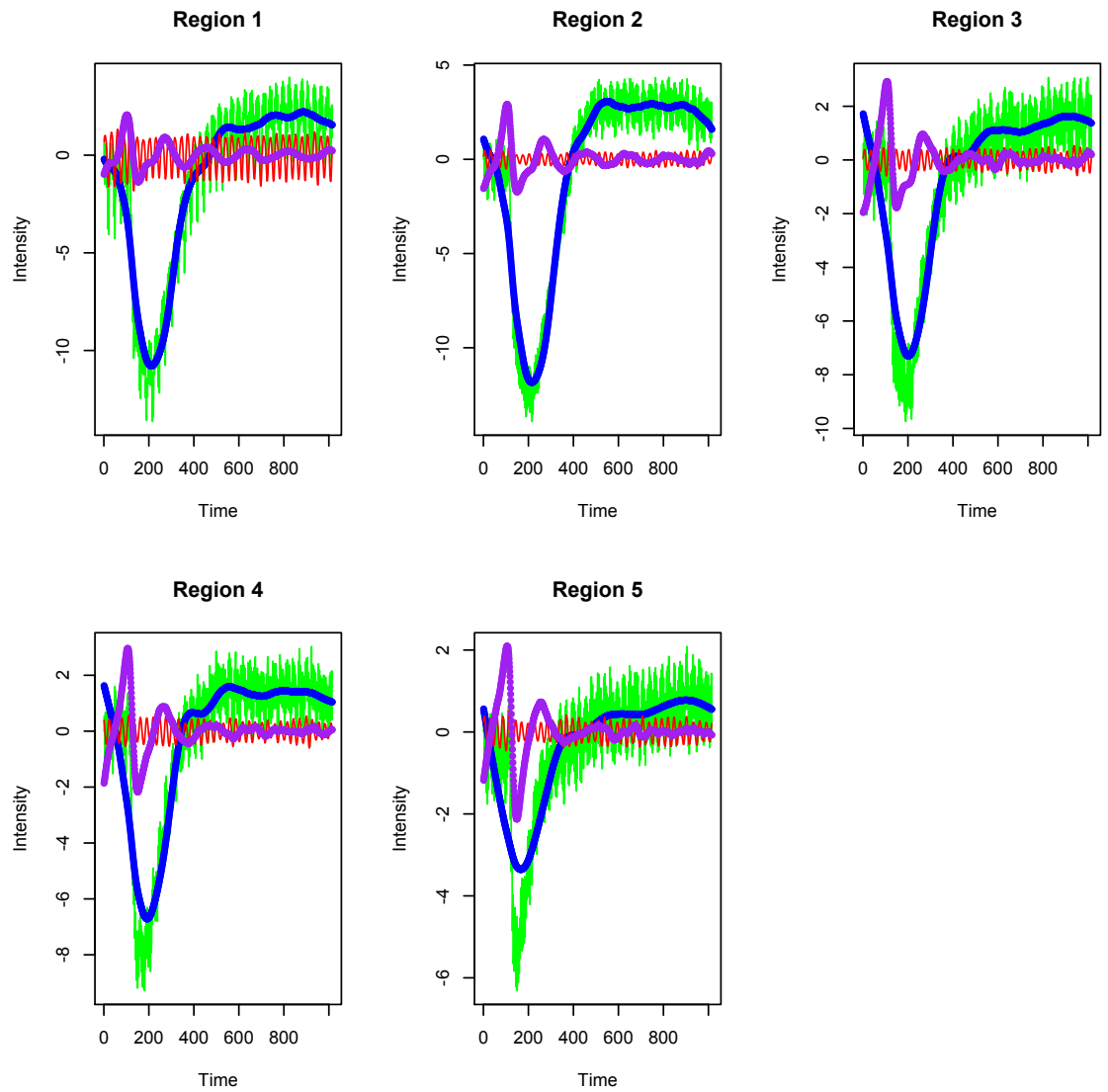Respiration for each Region from Replicate 1 of the Study

Figure 6: Observed Data and Estimated Smooth Level, Heartbeat and Respiration for each Region from Replicate 2 of the Study
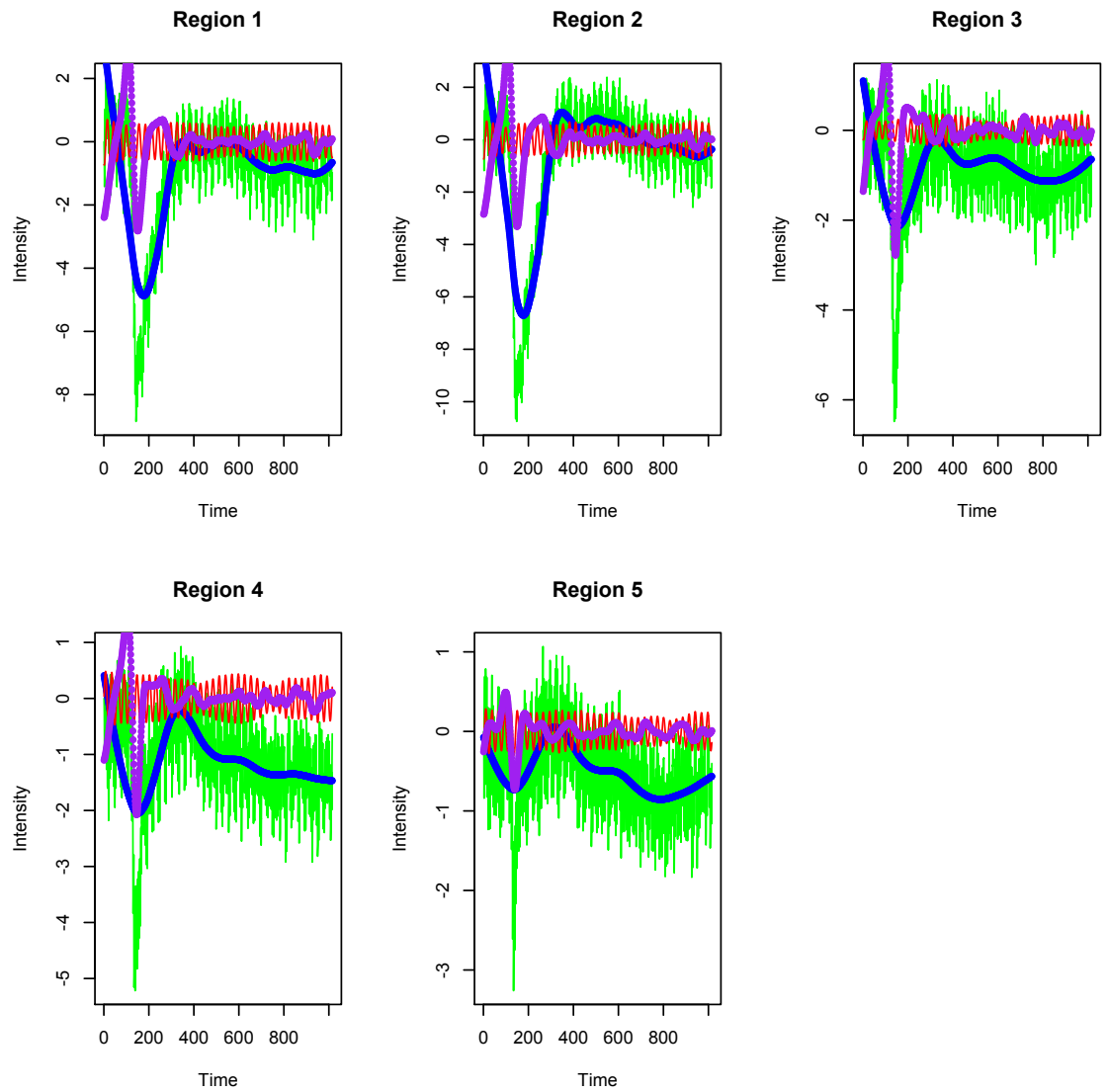
Figure 7: Observed Data and Estimated Smooth Level, Heartbeat and Respiration for each Region from Replicate 3 of the Study
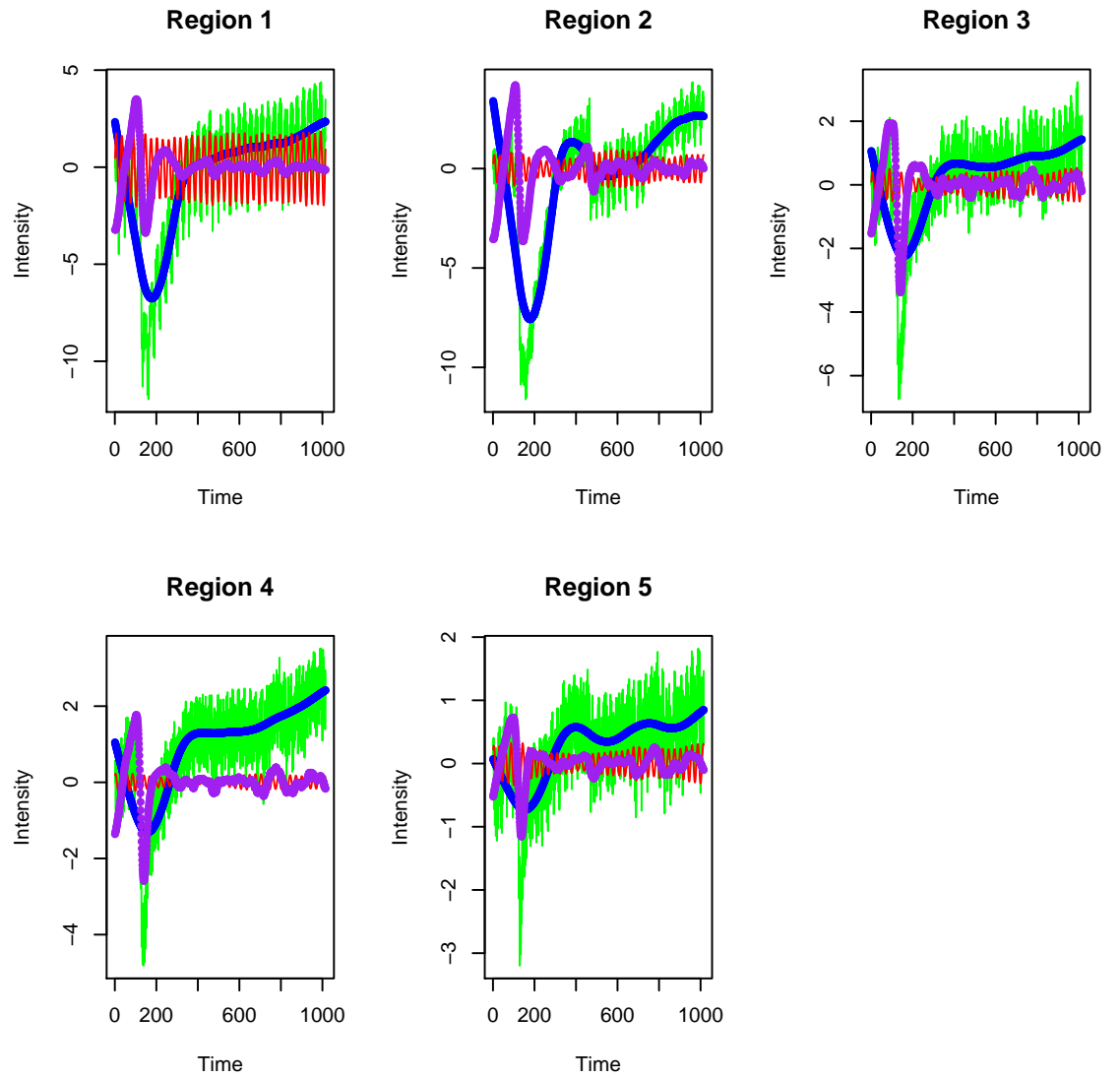
Figure 8: Observed Data and Estimated Smooth Level, Heartbeat and Respiration for each Region from Replicate 4 of the Study
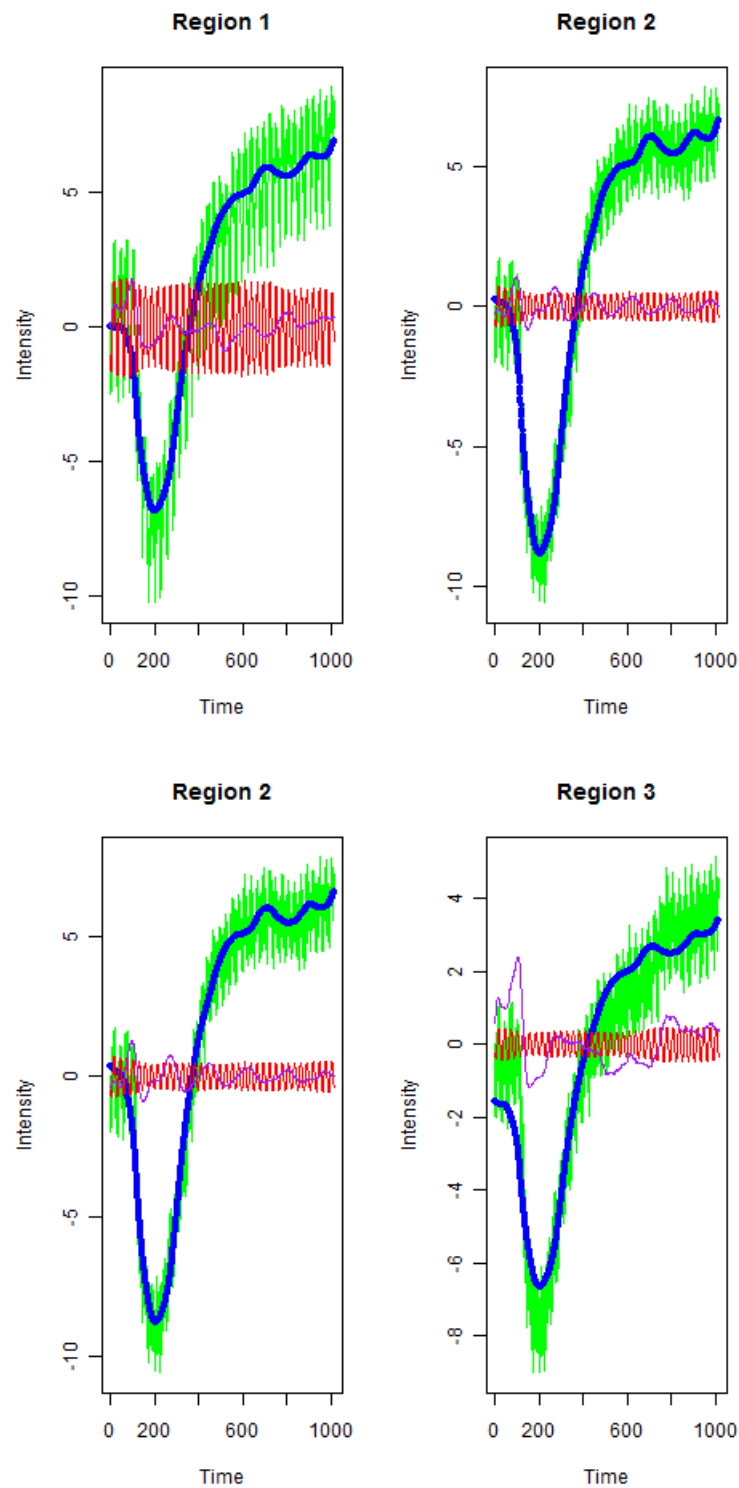
9.2 APPENDIX II: Plots from Multivariate Analysis

Figure 9: Observed Data and Estimated Smooth Level, Heartbeat and Respiration from Multivariate Analysis of Regions 1&2 and 2&3 Respectively from the dataset in Replcate 1
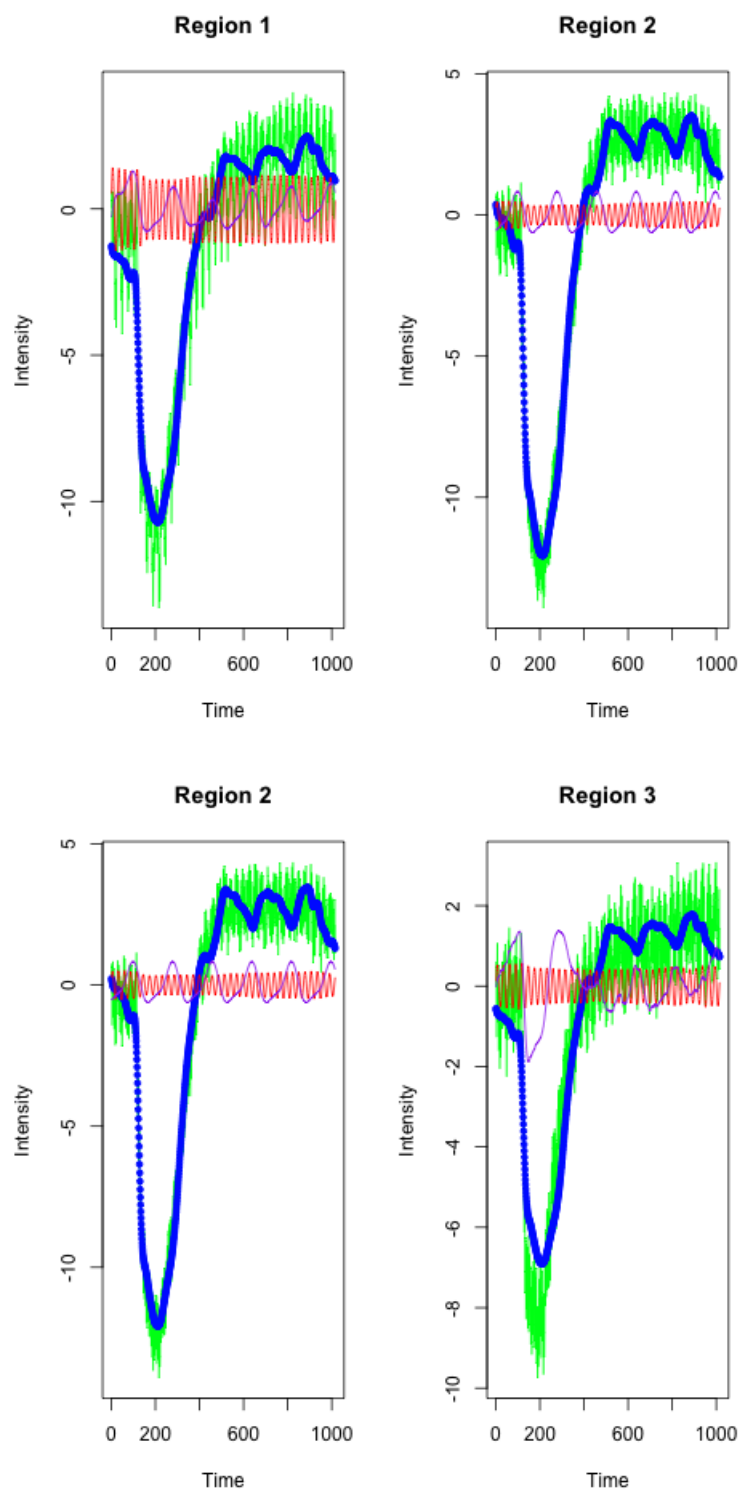
Figure 10: Observed Data and Estimated Smooth Level, Heartbeat and Respiration from Multivariate Analysis of Regions 1&2 and 2&3 Respectively from the dataset in Replcate 2
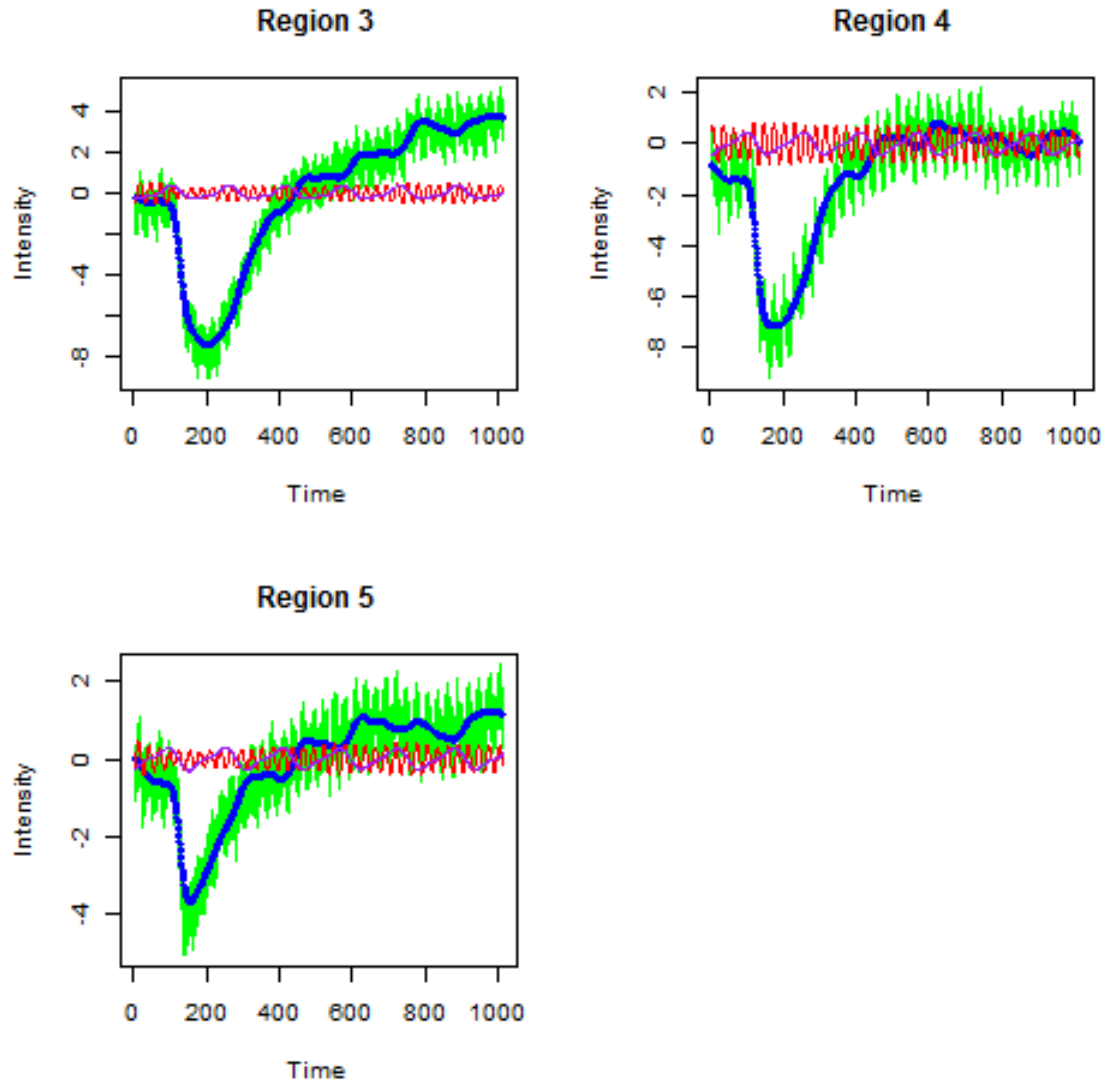
Figure 11: Observed Data and Estimated Smooth Level, Heartbeat and Respiration from Multivariate Analysis of Regions 2,3&4 from the dataset in Replcate 1