Transnational University Limburg is a unique collaboration of two universities in two countries: the University of Hasselt and Maastricht University.

UHASSELT
KNOWLEDGE IN ACTION

www.uhasselt.be
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

UHASSELT
KNOWLEDGE IN ACTION

Maastricht University

Faculty of Sciences
School for Information Technology

Master of Statistics

Masterthesis

Mathematical model to predict the disease course of Multiple Sclerosis (MS)

Oluyomi Modupe Adesoji

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics, specialization Epidemiology & Public Health Methodology

SUPERVISOR :
Prof. dr. Dirk VALKENBORG

SUPERVISOR :
Dr. Liesbet PEETERS

2017
2018

Faculty of Sciences
School for Information Technology
Master of Statistics

Masterthesis

Mathematical model to predict the disease course of Multiple Sclerosis (MS)

Oluyomi Modupe Adesoji

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics, specialization Epidemiology & Public Health Methodology

SUPERVISOR :
Prof. dr. Dirk VALKENBORG

SUPERVISOR :
Dr. Liesbet PEETERS

Abstract

Multiple Sclerosis (MS) is an auto-immune disease of the brain and the spinal cord in which the immune system of an individual attack the protective sheath covering their neurons. MS is characterized by weakness, numbness, blurred vision, bladder dysfunction and lack of muscle coordination. Various forms of the disease exist and are categorized by how they progress from the onset. However, various test and examinations are available to monitor the course of the MS. The most widely used score by the neurologist is the expanded disability status scale. Also, evoked potentials such as visual evoked potentials(VEPs), somatosensory evoked potentials(SEPs) and motor evoked potentials(MEPs) have been used. It is of interest here to predict the course of MS by studying the EDSS score as a function of MEPs and some other clinical variables such as age, gender, and type of multiple sclerosis. The MEPs are projected firstly into the wavelet domain to produce a sparse representation of the data. Then a penalized regression model, wavelet-based logistic LASSO regression is fitted to these variables to predict the course of MS. It was observed that the MEPs predicted the EDSS score with 72% accuracy. Adding other patient characteristics improved the accuracy of the prediction by about 6%. Moreover, the results also show that only early stages of the disease were well predicted by the MEPs.

Keywords: *EDSS score, Evoked potentials, Wavelet, LASSO* .

Contents

List of Tables

List of Figures

1 INTRODUCTION

1.1 Background

Multiple Sclerosis is an auto-immune disorder of the central nervous system (brain and spinal cord), which is preceded by the degeneration of the myelin sheath. The protective sheath that covers the nerve fibers in an individual is attacked by own immune system thereby causing disruption in communication between the brain and the rest of the body. Hence, lack of muscle coordination, muscle weakness, slurry speech, decreased memory, decreased spontaneity, paresthesia, bladder dysfunction and visual impairment (Goldenberg, 2012). Although the cause of Multiple Sclerosis is unknown, it is widely believed to occur as a result of the synergistic effect of non-genetic and genetic factors. Some triggers such as viruses, bacteria, stress, smoking and other environmental factors in combination with genetic susceptibility have been implicated in the cause and progression of MS (Koriem, 2016).

Young people between 20 to 50 years of age are more likely to suffer from MS. This present great economic challenges as most of the people affected loose ability to keep working after about 10 years of the attack thereby resulting in total economic dependency. According to Atlas of MS 2013 , the estimated number of people with MS has increased from 2.1 million in 2008 to 2.3 million in 2013 and the incidence of MS in females approximately double that of males. The prevalence of MS varies across regions of the world but the large population of Nothern European Countries, Canada, New Zealand, and some other countries in the temperate regions are believed to suffer more from the disease than others (Milo and Kahana, 2010). Although, this has not been fully established since the disease is caused by both hereditary and non-hereditary factors but lack of vitamin D has been implicated these areas (Koriem, 2017).

Multiple sclerosis has been categorized into types based on the disease course; Relapsing – Remitting MS, Secondary Progressive MS, Primary Progressive MS, and Progressive-Relapsing MS (Goldenberg, 2012; Hauser and Goodin, 2005; Loma and Heyman, 2011). Relapsing-Remitting MS (RRMS) is the most common form of the disease characterized by periods of sudden worsening of symptoms(relapse or exacerbation) and immediate improvement or total absence of symptoms. The progression of RRMS with or without periods of relapsing or remitting is known as Secondary Progressive MS(SPMS) whereas, in Primary Progressive MS (PPMS), the symptoms of the disease worsen from onset without periods of relapsing or remission. The Progressive-relapsing MS (PRMS) is the least common form of the disease with the progression of the disease from the onset and intermittent worsening of the disease over the years. Patients who only experience a single episode with clinical symptoms are reported to have clinically Isolated Syndrome (CIS).

However, it is quite difficult to diagnose MS and available methods are not specific for detecting the disease alone but other diseases that could present the same symptoms. Therefore, a number of tests are employed to rule out other types of diseases and confirm MS especially after two or more episodes of MS-related symptoms. Both clinical and para-clinical tests are used to diagnose MS. Blood tests, Lumbar puncture, magnetic resonance imaging (MRI) and Evoked potential tests are jointly used. MRI scans are basically used to detect lesions or scarring of the myelin sheath in the brain or spinal cords while Evoked potential tests capture how long it takes messages from the eyes, hands, and legs to reach the brain and vice versa through an electrode because scaring of the myelin sheath obstruct the rapid passage of signals. Although the use of evoked potential tests have been a subject for debates, it is believed to be more useful for monitoring the course of multiple sclerosis than the subjective Expanded disability status scale (EDSS) introduced by Kurtzke (1983)

The Expanded Disability Status Scale (EDSS) is a 20-point scale (ranging from $0 =$ normal to $10 =$ death due to MS, marked by 0.5 increments) and is currently the most widely used measurement scale by neurologist to quantify disability due to MS (Kurtzke, 1983; McKay et al., 2016). MS patients who are able to walk without aid are categorized within the range 1 to 4.5 while patients that fall within the scale of 5 to 9.5 are defined by the inability to walk properly (see section 6.1). The EDSS classification is made from scoring impairment on the scale of 0 to 5 or 6 in eight functional systems (FSS), which are pyramidal(weakness), cerebellar(tremor), brain-stem (speech problems), sensory(numbness), bowel and bladder function, visual function, mental function and others. Apart from the fact that EDSS is subjective, it is largely dependent on walking abilities and the fact that increment of one at a lower scale does not translate to the same effect at a higher scale. But, it is the most popularly used score for monitoring MS progression, especially in clinical trials. However, to further assess the cognitive abilities and abilities to carry out daily activities, new outcome measures such as MS functional composites (MSFC) and patients reported outcome measures (PROM) have been proposed. See Van Munster and Uitdehaag (2017) for details.

Various forms of evoked potential tests are widely in use. They are; visual evoked potentials (VEPs), and somatosensory evoked potentials (SEPs), motor evoked potential (MEP). MEP test which was employed in this study is carried out by single or repeated pulse stimulation of the brain which causes the spinal cord and the peripheral muscles to produce neuro-electrical signals. These signals are recorded by intramuscular needle electrodes from hands or legs. The amplitude, latency, threshold and time taken by the signal to reach the peripheral muscles are important in the prognosis of MS disease. A decrease in amplitude of the signal over time can be indicative of an injury or lesion on the signal path (see section 2.1 for details). Also, due to

the fact that the transcranial stimulation on the head is varied from spot to spot around a segment of the brain and these spots might vary from neurologist to neurologist thereby inducing a lot of variation.

Therefore, the objectives of this study are to develop a generic method to analyze raw signals (time series data) instead of the average of amplitude and latency that has been used in literature. The raw signals are represented in another domain using wavelets method and thresholding is applied before a functional lasso regression model is fitted to select a number of features both clinical and non-clinical (MEP) which predict the outcome measure of MS diseases, the EDSS of each patient. Section two of this reports described the methodology employed to achieve study objectives, the results obtained are given in section three and section four contains further discussion and conclusion from the study.

2 METHODOLOGY

2.1 Data Description

The data used in this study contains evoked potential measurements of 447 patients at their first visit to the clinic. A repeated transcranial stimulation of a preselected region of the brain is carried out and an electrode is attached to either the left or right limbs to record the amplitude of the signal as it travels from the brain to the limbs. The region of the brain to be used is divided into two; left and right part and stimulation of these parts of the brain are done repeatedly to find a spot that induces activation of evoked potential in the neuronal tract of the patients. From the series of these stimulations on each patient, the MEP signal with the highest peak to valley is selected as the most representative signal to be used in the prediction of MS.

The interest of this study lies in analyzing a single evoked potential for individual patients from the different repeated measurements recorded over time, a visual method was used to select the signal with the highest peak to valley amplitude over time(Milliseconds). Peak selection is done visually by plotting a graph of repeated signals of an individual and the signal with the highest peak to valley is selected. However, since a lot of patients are involved, an ad-hoc method was implemented for computational ease by selecting the signal with the maximum absolute amplitude (millivolt). Other clinical measures such as age, gender, EDSS score, amplitude, latency, FSS score were also recorded. The EDSS score which is the clinical outcome measure is categorized into two levels. Patients within the range of 1 to 4.5 are classified as fully ambulatory while patients within the range of 5 to 9.5 are categorized as impaired according to the EDSS scale by Kurtzke (1983).

2.2 Wavelet Methods

2.2.1 Wavelets Decomposition

One of the greatest challenges of using the direct time series MEP in a statistical model is the high dimensional nature of the data. Since the data are sampled at discrete time points, it results in a lot of variables which most times are far more than the number of observations. However, a method to compress the information contained in the signal so that a few parameters will be estimated while retaining the properties of the whole signal is desired. Various data compression and de-noising methods exist in literature such as Fourier transform, wavelet transform, adaptive filtering, Savitzky-Golay filtering, etcetera (AlMahamdy and Riley, 2014). But, the discrete wavelet transform would be used in this setting.
Wavelets are families of functions commonly used in signal processing, data compression, and noise reduction. They are basis functions usually employed to approximate other functions resulting in few non-zero coefficients. Wavelets have some very important properties which make them very useful in statistics. Properties such as sparsity, localization, and efficiency. Wavelet analysis results in a few often large coefficients which are representative of the original signal, a property known as sparsity. Also, wavelets do not exist forever but rapidly decay to zero. Hence, the ability to analyze signals that are with abrupt transient regions. Only coefficients that overlap the discontinuity region are influenced unlike the Fourier transform where all bases regardless of location will interact with the discontinuity (Nason, 2010a).

In statistics methods involving wavelet transform have been applied to solve dimensionality problems and give a useful prediction of a scalar outcome variable as a function of functional predictors. since such predictors most often contain parameters that are much larger than the number of observations ($p >> n$). Many of such methods abound in literature. For example, Amato et al. (2006) performed dimension reduction of functional data through minimum average variance estimation (MAVE) and functional sliced inverse regression (FSIR) methods. Further, Morris and Carroll (2006) described a wavelet-based functional mixed model using Bayesian approach. Zhao et al. (2012) discussed a wavelet-based LASSO regression to select important coefficients from a set of signals that are already projected into the wavelet domain. Zhao et al. (2015) extended the wavelet-based LASSO regression by applying different weighting and screening approaches to the coefficients obtained by wavelets transform method. In this analysis, we adopt the methodology of Zhao et al. (2012) to find the regions and characteristics of the evoked potential that are predictive of EDSS.

Two bases functions of wavelets, the scale(Φ) and mother wavelet(Ψ) are very important in the construction of wavelets coefficients. To fix ideas, for a set of curves sampled n times, and $n = 2^J$, the discrete wavelet transform produces a vector of coefficients ($d_{j,k}$) consisting differ-

ent levels of finesse or coarseness for $j = 0, \cdots, J-1$ and $k = 0, \cdots, 2^j - 1$. j and k are level and location parameters respectively. The wavelet coefficients and smoothed data are produced through a pyramidal algorithm proposed by Mallat (1989). The Mallat algorithm makes use of two filters; high pass and low pass filter for the multi-resolution analysis of the signal. Coarser levels of the scale(Φ) and mother wavelet(Ψ) are obtained from the data (the finest coefficients).

For a function $L^2(\mathbb{R})$, an orthonormal basis is obtained by dilating and translating the scale (Φ) and mother wavelet (Ψ) as:

$$\Phi_{jk}(t) = 2^{j/2}\Phi(2^j t - k)\ , \Psi_{jk}(t) = 2^{j/2}\Psi(2^j t - k). \tag{1}$$

These functions are referred to as orthonormal because they satisfy the Parseval's relation which shows that the transpose of the matrix of scale function multiplied by the matrix of mother wavelet function or the matrix of scale function multiplied by the transpose of the matrix of mother wavelet function is equal to an identity matrix. Also, the relation explains that the wavelet coefficients have the same energy as the input data.

Then a function $\Gamma(t)$ can be represented by wavelet series which is the sum of the products of coefficients and the basis.

$$\Gamma(t) = \sum_{j,k\in\mathbb{R}} \beta_{j,k}\Psi_{j,k}(t), \ \ \Gamma(t) = \sum_{j,k\in\mathbb{R}} \beta\prime_{j,k}\Phi_{j,k}(t). \tag{2}$$

In the pyramid algorithm, the scale and wavelet coefficients are dilated by dilation coefficients g_n and h_n respectively at different resolution in space (Figure 8). The scale and wavelet coefficients at level j_0 are obtained by :

$$\beta_{j_0,k} = \sum_n h_{n-2k}\beta j, n, \ \ \beta\prime_{j_0,k} = \sum_n g_{n-2k}\beta\prime j, n \tag{3}$$

The low pass filters produce the scaling coefficients while the high pass filters produce the mother wavelet coefficients. The output has N/2 coefficients at the next coarser level and N/4 at the subsequent one as indicated by ($\downarrow$ 2) on figure 8. However, given an initial decomposition level j_0, function $\mathbf{\Gamma}(\mathbf{t})$ in the wavelet domain $L^2[0,1]$ is :

$$\Gamma(t) = \sum_{k=0}^{2^{j_0}-1} \beta\prime_{j_0,k}\Phi_{j_0,k}(t) + \sum_{j=j_0}^{\infty}\sum_{k=0}^{2^j-1} \beta_{j_0,k}\Psi_{j_0,k}(t) \tag{4}$$

thus, the $\beta\prime_{j,k}$ and $\beta_{j,k}$ which are the scale and the mother wavelet coefficient for level j_0, are computed by :

$$\beta\prime_{j_0,k} = \int \Gamma(t)\Phi_{j_0,k}(t)dt, \ \ \beta_{j_0,k} = \int \Gamma(t)\Psi_{j_0,k}(t)dt \tag{5}$$

Different families of wavelet exist and appropriate family to use for specific analysis problem depends on how compactly the bases function are localized in space and how smooth they are (Graps). Further, wavelets are also classified within a family based on the number of vanishing moments which however have a direct relationship with the coefficients. The vanishing moments describe the degree of polynomials the scaling function can generate. In this setting, Daubechies compactly supported wavelets are used with the vanishing moment of 4 with a periodic boundary option which restricts the previously defined function, $L^2(\mathbb{R})$ to $L^2[0,1]$, that is, a finite interval of $\mathbb{R}$ is of interest. The coefficients of the finer levels of the transform represent the local structure of the data while those at coarser level describe global properties of the data.

2.2.2 Thresholding

Based on the properties of wavelet coefficients discussed in the previous section, such as sparsity of the resulting coefficients, the concentration of energy of the original sequence in a few coefficients through Parseval's relation and orthogonality of the wavelet function used, it is then expected that small values(close to zero) of the coefficients contain very little or no energy while the large coefficients contain the most of the energy in the data. Thresholding entails setting or shrinking some of the wavelet coefficients below a defined value (δ) to zero. Two major types of thresholding exist, namely hard and soft. The hard thresholding methods (Equation (6)) sets all the coefficient less or equal the threshold value to zero while the soft thresholding (Equation (7)) shrinks coefficients both positive and negative (Figure 9) .

$$D^H(d \setminus \delta) = \begin{cases} 0, & Otherwise \\ d, & for\ |d| \geq \delta \end{cases} \tag{6}$$

$$D^S(d \setminus \delta) = \begin{cases} 0, & for\ |d| < \delta \\ d - \delta, & for\ d \geq \delta \\ d + \delta, & for\ d \leq -\delta \end{cases} \tag{7}$$

In this study, a range of increasing threshold values are selected and shrinkage is manually carried out using the hard thresholding technique. The residuals, which are the difference between the reconstructed and the original data, are checked visually using quantile-quantile plots and through two formal tests for normality; Shapiro Wick test and Anderson-Darling normality test. Thresholding is done at each value and stopped at the highest value where the assumption of normality on the residual is no longer met. This highest value at which normality assumption is still valid is considered optimum and used in thresholding the data producing a very sparse matrix

with many Zeroes. At lower decomposition level, most or all the wavelet coefficients shrinks to zero because this level contain the finest detail in the data. But, at higher levels the coefficients are summarised into coarser details.

2.2.3 Signal Reconstruction

The inverse of the pyramid algorithm described in section 2.2.1 is applied to reconstruct the data. This algorithm is known as inverse discrete wavelet transform. Although the LASSO model will be fitted in the wavelet domain, reconstruction is carried out to assess the magnitude of the difference (residuals) in the data after wavelet methods are applied. The reconstructed data also give a clear picture of how wavelet decomposition and thresholding can systematically produce a sparse matrix without discarding data. The wavelet coefficients, as well as other information from the thresholding, are used to reconstruct the original signal. The inversion equation for the reconstruction is:

$$\beta\prime_{j,n} = \sum_k g_{n-2k}\beta\prime_{j-1,k} + \sum_k h_{n-2k}\beta_{j-1,k}, \tag{8}$$

as shown on Figure 8b, where ($\uparrow 2$) represent the scaling up of the coefficients by putting Zeros between two subsequent coefficient and apply in equation (8) at the different resolution until the data is obtained.

Figure 1: Wavelets transform steps

2.3 Regression Modeling

2.3.1 LASSO Regression

Lasso regression is a variable selection and shrinkage method used in solving regression problem, especially in high dimensional data. Lasso solution has the property that the coefficients can shrink to zero hence useful in variables selection (Friedman et al., 2001). The coefficients of lasso regression $\hat{\beta}$ are estimated such that its value minimizes the expression below:

$$\hat{\beta}_\lambda = argmin_\beta \frac{1}{2}(Y - Z\beta)^T(Y - Z\beta) + \lambda \sum_{j=1}^{N} |\beta_j| \tag{9}$$

where λ is the complexity parameter(tunning parameter) that controls the amount of shrinkage. Therefore, the value of λ that will yield the most accurate prediction must be appropriately chosen, since selecting a high lambda value shrinks almost all the coefficients to zero while a value of lambda close to zero results in an over-parameterized ordinary least square regression (Zhao et al., 2012).

2.4 Functional Regression Model

Wavelet-based lasso regression method can be used to describe the relationship between a scalar response variable (Y_i) and functional regressors (curves).

$$logit(y_i) = \alpha + \int X_i(t)\Gamma(t)dt, \quad i = 1, \cdots n, \tag{10}$$

where the intercept β_0 is a scalar parameter, $\Gamma(t)$ is a square integrable function described in equation 4 and it describes the relationship between X and Y. Regions with large $|\Gamma(t)|$ are indicative of better prediction of Y by changes in $X(t)$. For ease of representation α can be dropped from (10) as explained by Zhao et al. (2012), and estimated as $\hat{\alpha} = \bar{y} - \int x(t)\bar{\Gamma}(t)dt$ where $\bar{y}$ and $\bar{x(t)}$ are sample means of the response and predictors respectively. X_i sampled at N equal spaced points can be projected into wavelet domain and represented with N wavelet coefficients as :

$$X_i = \sum_{k=0}^{2^{j_0}-1} d\prime_{i,j_0,k}\Phi_{i,j_0,k}(t) + \sum_{j=j_0}^{\infty}\sum_{k=0}^{2^j-1} d_{i,j_0,k}\Psi_{i,j_0,k}(t) = W^T H_i, \tag{11}$$

where,

$$d\prime_{j_0,k} = \int x_i(t)\Phi_{j_0,k}(t)dt, \;\; d_{j_0,k} = \int x_i(t)\Psi_{j_0,k}(t)dt \tag{12}$$

where W is an $N \times N$ orthogonal matrix associated with the scale and mother wavelet, D_i is a $N \times 1$ vector of wavelets coefficients of DWT of X_i. Also the $\Gamma(t)$ are represented in the wavelet domain as shown in equations (4) and (5). Then, due to the orthonormality characteristics of wavelet as shown by Zhao et al. (2015), the model in (10) can then be re-written as,

$$logit(y_i) = \sum_{k=0}^{2^{j_0}-1} d\prime_{i,j_0,k}\beta\prime_{j_0,k}(t) + \sum_{j=j_0}^{j}\sum_{k=0}^{2^j-1} d_{i,j_0,k}\beta_{j_0,k}(t), \;\; i = 1\cdots, n, \tag{13}$$

simply put in matrix form as:

$$logit(Y) = Z\beta$$

where β is a $N \times 1$ vector of coefficients, Z is a $ntimesN$ design matrix with ith rows containing $d_{i,j,k}$ values in the same order as β and $Y = (y_1, y_2, \cdots, y_n)T$. The β coefficients are estimated as describes in (9) can be reconstructed to the data scale using inverse discrete wavelet transform. However other scalar variables can also be added to the regression model. A logit link is used to linearly relate the systematic part of the model to the outcome.

The fact that a variable is selected and has non-zero estimate does not imply significance. Therefore, a post-selection inference method as proposed by Taylor and Tibshirani (2016) is applied to the results from the LASSO regression to compute P-values of the selected variables at a fixed lambda value.

2.5 Software

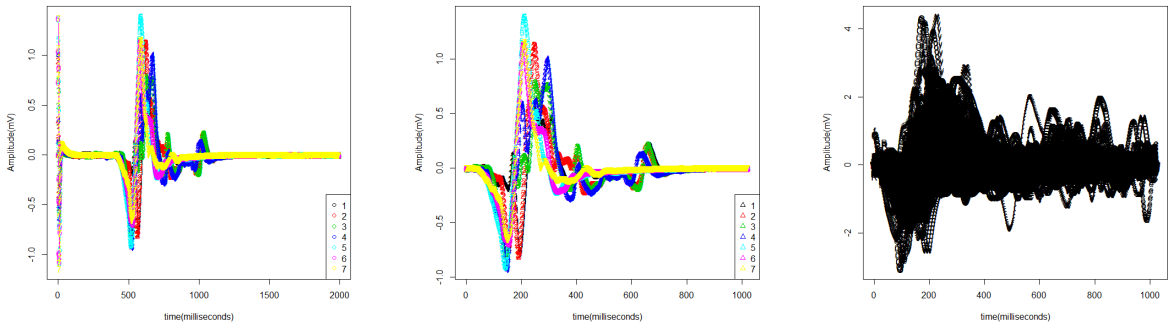
Package Wavetheresh (Nason, 2010b), glmnet (Friedman et al.) and selectInference were used in R statistical programming software version 1.0.153.

3 RESULTS

3.1 Exploratory

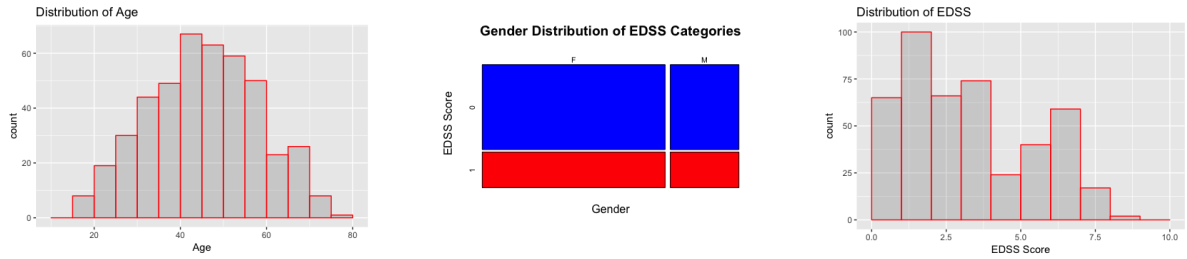
The data set contains both clinical and MEP measurements of 447 patients. The observed distribution of the clinical variables in the data is shown in figure 3. Patient's age varies between 20 and 80 years and it could be observed that the count of patients within the age bracket 30 - 60 years are higher compared to those below or above which might be due to the fact these are the active onset ages of the disease. Also, the gender distribution (figure 3b) shows that there are more female compared to male in the data set and more patients fall into the fully ambulatory category compared to the impaired category. There are 325 females and 122 males in the data set as reflected by the width of the boxes. Further, the distribution of the EDSS score (figure 3c), also indicates that many patients in the data set are below 4.5. 352 patients are in category 0, that is fully ambulatory while 132 are the category 1, impaired.

However, looking at the MEP curves for a patient (figure 2a), the initial fluctuation at the beginning (0ms) indicate the effect of the descending Transcranial stimulation (TMS) of the brain to activate the motor neurons on the spinal cord, followed by an initial period of latency before the onset of MEP which is in turn followed by a silent period. Since our interest lies in analyzing the MEP, the TMS effect and initial latency period are removed from the signal. The signal to be analyzed then begins from 377ms to 1400ms. Also, from the repeated test recorded for a single patient, a single MEP with the highest peak to valley amplitude is selected, which is test 5 on (figure 2b). This is done for all the patients and figure 2c shows the MEPs of all the 447 patients in the study.



(a) (b) (c)

Figure 2: *MEP signals:(a)-patient 700, (b)- patient 700 without latency period (c) - All patients curves*



(a) (b) (c)

Figure 3: *Distributions:(a)- Age, (b)- Gender , (c) - EDSS Score*

3.2 Wavelet Methods Result

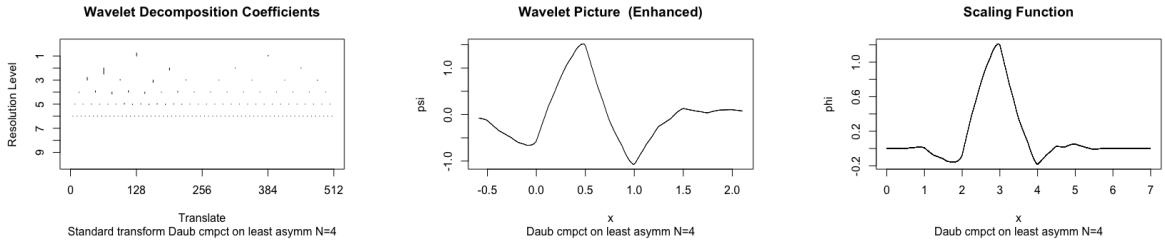
The decimated discrete wavelet transform of the Daubechies compactly supported wavelet family was applied to the MEPs. The least asymmetric type was used. It is so called because real-valued

compact orthonormal wavelets cannot be symmetric or antisymmetric according to Daubechies (1992). The MEPs are decomposed based on Mallat algorithm as discussed in section 2.2.1. Figure 4b and 4c displays the wavelet coefficient (h_n) of the wavelet family used for the fast wavelet transform. It is always twice the length of the vanishing moment of the wavelet family used, which in this project is 4. This implies that the scaling function can not generate polynomial greater than four. Since the length of the each patient MEP is 1024, then the number of levels (J) expected is $N = 2^J$. Coefficients $d_{j,k}$ are fewer and larger at coarser resolution levels and finer as the level increases. Applying the wavelet transform to the data yielded coefficients at 10 different resolution levels at different locations (translate k) as shown in figure 4a for one MEP. Ticker vertical marks show that the coefficients are large as opposed to dot at the higher revolution levels. Only 6 resolution level are displayed on the plot because the scale of the remaining levels are small. If the coefficient at each level are scaled up as in figure 10, then all the levels can be seen but the size of the coefficients cannot be compared anymore.

The coefficients at each level for a single MEP is arranged into a vector and these vectors are put together into a matrix (see Table 4 and 5). However, decomposition of the MEP resulted in many small coefficients that are approximately zero, hence, producing a matrix of dimension 447×1023. Since Parseval's relation proved that the most energy in the data is concentrated in the large coefficient in the wavelet domain, manual thresholding is applied to the decomposed data to shrink the small coefficients to Zero and further produce a more sparse matrix. To manually select the value at which thresholding is done(δ), a sweep of various value is done and $\delta = 0.25$ was selected based on normality checks on the residuals . The value of the residuals are obtained by subtracting the original data from the reconstructed data after thresholding. The assumption of normality seems good by mere looking at the quantile-quantile plot of the residuals of each MEP but, both Shapiro Wick test and Anderson-Darling test rejected the null hypothesis that the residuals are normally distributed. However at higher threshold value, both the visual and formal test perform woefully. Therefore a decision was taken to stick to the $\delta = 0.25$ for the thresholding.

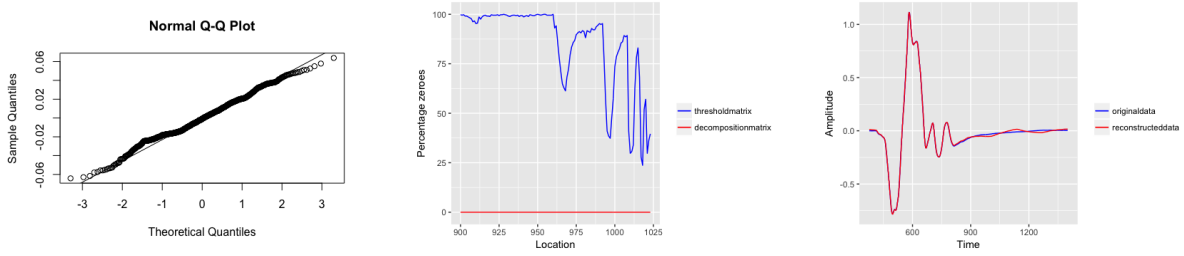
Further, reconstructing the data from the threshold data matrix confirms that the wavelet transform conserves the energy in the data as the reconstructed data is approximately the same as the original data (Figure 5c). Most of the wavelet coefficients become zeroes after thresholding and just about 45 columns out of 1023 columns of the thresholded data matrix have more than 10% of it data to be non-zero. Figure 5b shows a the percentage of zero in both threshold data matrix and decomposition data matrix from column 900 to 1023. It could be observed that the percentage of zeroes were almost 100% from column 900 to 962, that is, all the coefficients in these columns are already shrunk to zero while the decomposition matrix has 0% of zero values for all the matrix columns plotted but contains many coefficients close to zero (see Table 4 and 5). Therefore,

the LASSO regression model is fitted on the thresholded data dimension of 447×45 instead of 447×1024 of the original data.



(a) (b) (c)

Figure 4: *Wavelet decomposition: (a) - coefficients, (b) - wavelet function , (c) - scale function*



(a) (b) (c)

Figure 5: *(a) -qqplots of residuals, (b) - percentage zeroes in both decomposition and threshold matrix, (c) - Original and reconstructed data*

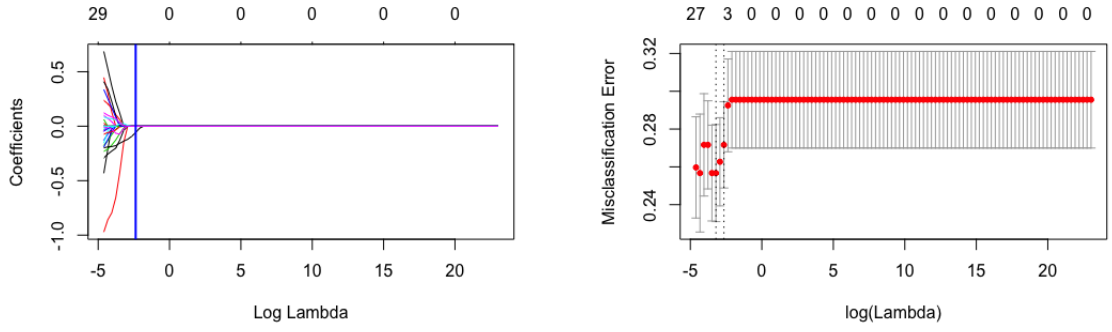
3.3 Wavelet-based Logistic LASSO Models Results

Lasso regression model in the wavelet domain is fitted on the thresholded coefficients. Three-quarter of the data is used to train the model and the complexity parameter lambda is selected over a grid of 100 values cross-validated 10 times. The optimal lambda that minimizes the miss-classification error with the smallest number of the coefficient is 0.123 but lambda value at one standard deviation from the minimum lambda value is often used in that the miss-classification error is comparable and it yields a less complex model a number of times. Based on this lambda value at one standard deviation from the minimum, 0.093, prediction of the response is done with the remaining one-quarter of the data set (the validation set). Figure 6a shows the coefficients of all the predictors selected and the vertical blue line indicated that at one standard deviation above the minimum, only one coefficient has a non-zero value. Also, figure 6b displays the various log of lambda values with their misclassification errors. The minimum log-lambda value and its one

standard error counterpart are indicated with vertical dotted lines.

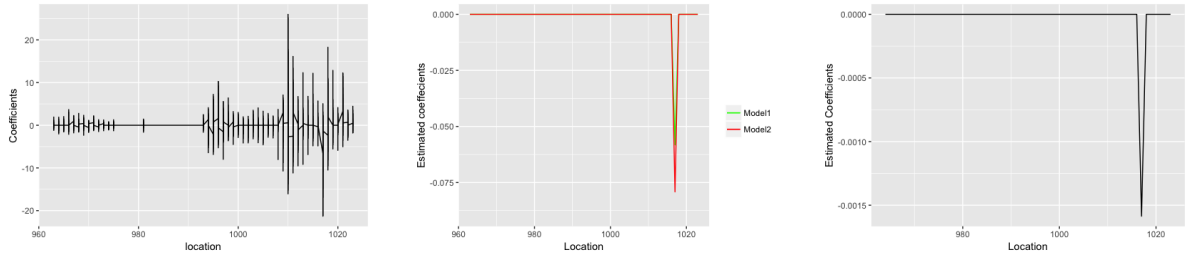
The number of variables selected is also displayed horizontally on top of the plot. Three forms of the logistic LASSO models were fitted through the process already described. Model 1 contains just the wavelet coefficients as predictors, model 2 contains the predictors in Model1 and patients characteristics such as age and gender while model 3 contains the predictors in Model 2 and one additional clinical variable, type of multiple sclerosis. About 25% data contained in the type of MS variable is missing, hence, a complete case analysis is carried out using only from data 243 patients. In model 1, 29 variables were selected but only one has a non zero coefficient at log-lambda value -2.6, which is the log-lambda value at one standard deviation from the minimum.

All the regression coefficient are shrunk to zero except coefficient at location 1017 in the wavelet domain (Figure 7b and 7c). This coefficient belongs to resolution level 2 with the longest tick mark as shown on figure 4a and represent the amplitude of patients at 441ms in the original sequence. All the functional coefficients selected belong to the coarser levels of decomposition 0 to 5 while all the finer levels are already shrunk to zeroes (Figure 7a and Table 6). The predictive accuracy of model 1 which contain only the wavelet coefficients on the EDSS score is 72%. The selected coefficient has a negative significant effect on the course of MS. Furthermore, Model 2 was also fitted by adding patients characteristics such as age and gender. Model 2 yielded an accuracy of 74% and age and gender has a negative significant effect on the response. A predictive accuracy of 78.6% was recorded for model 3 when the type of multiple sclerosis is added to model 2 and type of multiple sclerosis if RRMS is not significant. It is interesting to note that wavelet coefficient at location 1017, has a negative significant effect on the probability progression of MS.



(b) (c)

Figure 6: *Cross-validation (a) -Regression Coefficients , (b) -Complexity parameter selection*



(a) (b) (c)

Figure 7: *(a) -Thresholded data plot (b)-Estimated coefficients and position of Variables selected in model 1 and model 3 (c)- Estimated coefficients and position of Variables selected in model 3*

	Observed					
	Model 1		Model 2		Model 3	
Predicted	ambulatory	impaired	ambulatory	impaired	ambulatory	impaired
Ambulatory	81	31	81	29	45	11
Impared	0	0	0	2	2	3
Accuracy	72%		74%		78.6	

Table 1: Confusion matrices

	Model 1		Model 2		Model 3	
Variables	coefficient(se)	pvalue	coefficient(se)	p-value	coefficient(se)	p-value
intercept	0.689		2.13		1.79	.
age			-0.032(0.01)	$8.93 \times 10^{-}7$	-0.0166(0.0135)	0.0207
gender(M)				.		
1017	-0.065	$2.94 \times 10^{-}5$	-0.079(0.035)	$5.41 \times 10^{-}5$	-0.0016(0.403)	0.0003
type_MS(RRMS)					-1.004(0.044)	0.915

se - standard error

Table 2: Regression Coefficients

4 DISCUSSION AND CONCLUSION

Estimating MEPs directly in a regression model will yield an over-parameterized model with a lot of redundant parameters and very high variance. This is due to the fact that time series data of this format are high dimensional in nature. Wavelet transform has been employed in literature to tackle these problems by concentrating the important information of the data in a few coefficient thereby producing a sparse matrix. As it has been demonstrated in this study a 437×1024 dataset was represented by 437×45 coefficient in the wavelet domain. The reconstruction of these raw signals from the few wavelet coefficients indicated that the information loss is very small.

However, choice of the wavelet basis has a great influence on the outcome of the transform. Since the interest of the research is not to select the most optimal wavelet basis or investigate various forms of basis, a pragmatic choice of wavelet basis belonging to Daubechies family is made. This has been shown to perform well for localized data in time, frequency and has compact support. A supervised approach to select features that are important in predicting the EDSS score employed is logistic LASSO regression in the wavelet domain. logistic LASSO regression select important variables and shrink others to zero. However, this regression approach only performs well if the regularization parameter(λ) is carefully and accurately selected. Hence, the need for selecting it over a grid of values through cross-validation. In this study, the optimal lambda value for the three models lie between the range of -2.3 to -2.7.

In the three models fitted, variable 1017 was selected from the functional data and predicted EDSS with an accuracy of 72 to 78% in conjunction with other variables. This variable (1017) is the first element of level 2 of wavelet decomposition out of the 4 and approximately located at 441ms in the original time domain. All coefficient at finer scale within resolution level 6 to 9 turned to Zero after thresholding since they are small but reconstruction of the data confirmed the Parseval's relation that the total energy in the data are stored in large wavelet coefficients.

MEPs are quite largely related to the EDSS score in that both measures the functioning of the limbs. However, EDSS score is based on the physical ability of the patient to walk while the MEPs relies on the functioning of the motor neurons and how fast information sent from the brain could reach the hand or legs. It is not clear from the results if MEP measured from the hand neurons are better than those from the leg or if the side of the brain where the transcranial stimulation occurs has an effect on the predictive ability of the disease. But, age and the type of MS have an effect on the disease as shown in table 2. From the result of the predictions on table 1, It could be noted that fully ambulatory patients were much well predicted than the impaired ones, using the MEPs. This might be indicative of MEPs being of better diagnostic values at the initial stage of MS disease than at later stages. It might also be necessary to use

a less subjective scale of measuring the outcome of MS disease and build a better scale based on the characteristics of MEPs and other evoked potential tests in order to fully estimate the predictive abilities of MEPs on MS disease. Further, MEPs predictive abilities could also be improved if neurologist develops a better measure for part of the brain to be stimulated, rather than the random stimulations currently being carried out.

References

AlMahamdy, M. and Riley, H. B. (2014). Performance study of different denoising methods for ecg signals. *Procedia Computer Science*, 37:325–332.

Amato, U., Antoniadis, A., and De Feis, I. (2006). Dimension reduction in functional regression with applications. *Computational Statistics & Data Analysis*, 50(9):2422–2446.

Daubechies, I. (1992). *Ten lectures on wavelets*. SIAM.

Federation, M. S. I. (2013). Atlas of ms 2013: Mapping multiple sclerosis around the world.

Friedman, J., Hastie, T., and Tibshirani, R. (2001). *The elements of statistical learning*, volume 1. Springer series in statistics New York.

Friedman, J. H., Hastie, T., and Tibshirani, R. glmnet: lasso and elastic-net regularized generalized linear models, 2010b. *URL http://CRAN. R-project. org/package= glmnet. R package version*, pages 1–1.

Goldenberg, M. M. (2012). Multiple sclerosis review. *Pharmacy and Therapeutics*, 37(3):175.

Graps, A. An introduction to wavelets.

Hauser, S. L. and Goodin, D. (2005). Multiple sclerosis and other demyelinating diseases. *HARRISONS PRINCIPLES OF INTERNAL MEDICINE*, 16(2):2461.

Koriem, K. M. M. (2016). Multiple sclerosis: New insights and trends. *Asian Pacific Journal of Tropical Biomedicine*, 6(5):429–440.

Koriem, K. M. M. (2017). Corrigendum to 'multiple sclerosis: New insights and trends'. *Asian Pacific Journal of Tropical Biomedicine*, 7(5):493–504.

Kurtzke, J. F. (1983). Rating neurologic impairment in multiple sclerosis an expanded disability status scale (edss). *Neurology*, 33(11):1444–1444.

Loma, I. and Heyman, R. (2011). Multiple sclerosis: pathogenesis and treatment. *Current neuropharmacology*, 9(3):409–416.

Mallat, S. G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE transactions on pattern analysis and machine intelligence*, 11(7):674–693.

McKay, K. A., Jahanfar, S., Duggan, T., Tkachuk, S., and Tremlett, H. (2016). Factors associated with onset, relapses or progression in multiple sclerosis: a systematic review. *Neurotoxicology*.

Milo, R. and Kahana, E. (2010). Multiple sclerosis: geoepidemiology, genetics and the environment. *Autoimmunity reviews*, 9(5):A387–A394.

Morris, J. S. and Carroll, R. J. (2006). Wavelet-based functional mixed models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(2):179–199.

Nason, G. (2010a). *Wavelet methods in statistics with R*. Springer Science & Business Media.

Nason, G. (2010b). wavethresh: Wavelets statistics and transforms. *R package version*, 4(6.2013).

Souani, C., Abid, M., Torki, K., and Tourki, R. (2000). Vlsi design of 1-d dwt architecture with parallel filters. *INTEGRATION, the VLSI journal*, 29(2):181–207.

Taylor, J. and Tibshirani, R. (2016). Post-selection inference for l1-penalized likelihood models. *arXiv preprint arXiv:1602.07358*.

Van Munster, Caspar, E. and Uitdehaag, Bernard, M. (2017). Outcome measures in clinical trials for multiple sclerosis. *CNS drugs*, pages 1–20.

Zhao, Y., Chen, H., and Ogden, R. T. (2015). Wavelet-based weighted lasso and screening approaches in functional linear regression. *Journal of Computational and Graphical Statistics*, 24(3):655–675.

Zhao, Y., Ogden, R. T., and Reiss, P. T. (2012). Wavelet-based lasso in functional linear regression. *Journal of Computational and Graphical Statistics*, 21(3):600–617.

6 Appendices

6.1 Expanded Disability Status Scale(EDSS)

Score	Description
1.0	No disability, minimal signs in one FS
1.5	No disability, minimal signs in more than one FS
2.0	Minimal disability in one FS
2.5	Mild disability in one FS or minimal disability in two FS
3.0	Moderate disability in one FS, or mild disability in three or four FS. No impairment to walking
3.5	Moderate disability in one FS and more than minimal disability in several others. No impairment to walking
4.0	Significant disability but self-sufficient and up and about some 12 hours a day. Able to walk without aid or rest for 500m
4.5	Significant disability but up and about much of the day, able to work a full day, may otherwise have some limitation of full activity or require minimal assistance. Able to walk without aid or rest for 300m
5.0	Disability severe enough to impair full daily activities and ability to work a full day without special provisions. Able to walk without aid or rest for 200m
5.5	Disability severe enough to preclude full daily activities. Able to walk without aid or rest for 100m
6.0	Requires a walking aid – cane, crutch, etc. – to walk about 100m with or without resting
6.5	Requires two walking aids – pair of canes, crutches, etc. – to walk about 20m without resting
7.0	Unable to walk beyond approximately 5m even with aid. Essentially restricted to wheelchair; though wheels self in standard wheelchair and transfers alone. Up and about in wheelchair some 12 hours a day
7.5	Unable to take more than a few steps. Restricted to wheelchair and may need aid in transfering. Can wheel self but cannot carry on in standard wheelchair for a full day and may require a motorised wheelchair
8.0	Essentially restricted to bed or chair or pushed in wheelchair. May be out of bed itself much of the day. Retains many self-care functions. Generally has effective use of arms
8.5	Essentially restricted to bed much of day. Has some effective use of arms retains some self-care functions
9.0	Confined to bed. Can still communicate and eat
9.5	Confined to bed and totally dependent. Unable to communicate effectively or eat/swallow
10.0	Death due to MS

Table 3: EDSS Score

6.2 Further Results

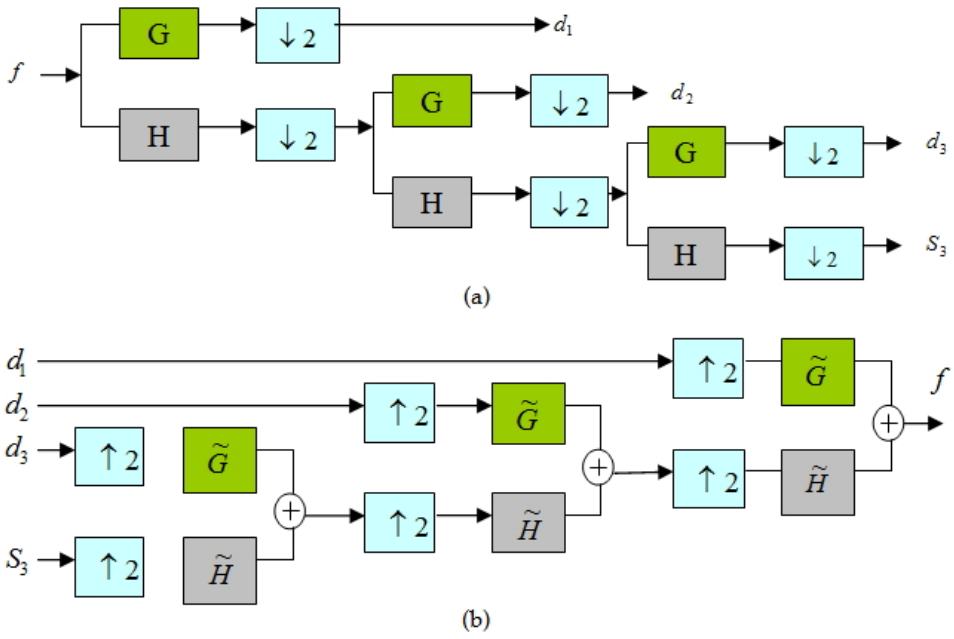
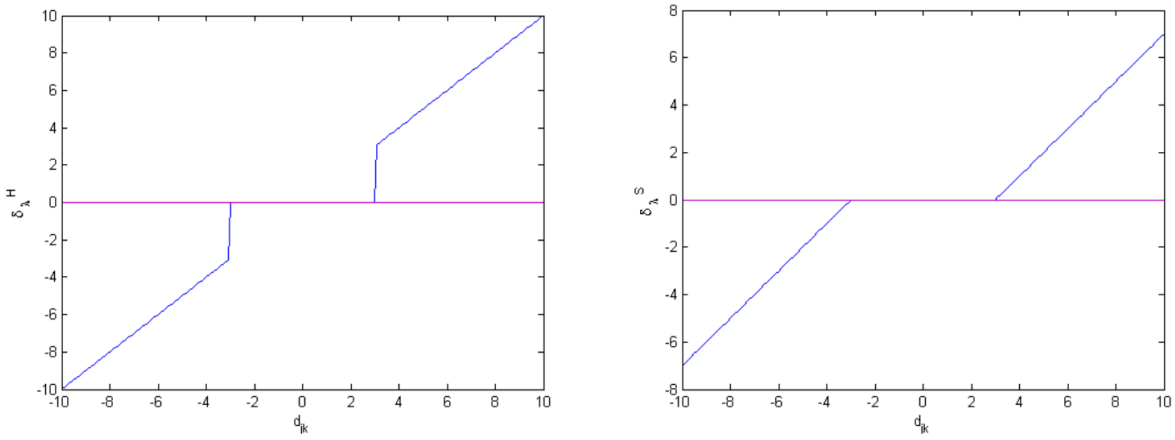


Figure 8: *Source: Souani et al. (2000), (a): DWT ,(b):IDWT*



(a) Hard-thresholding (b) Soft-thresholding

Figure 9: *Hard and soft threshold*

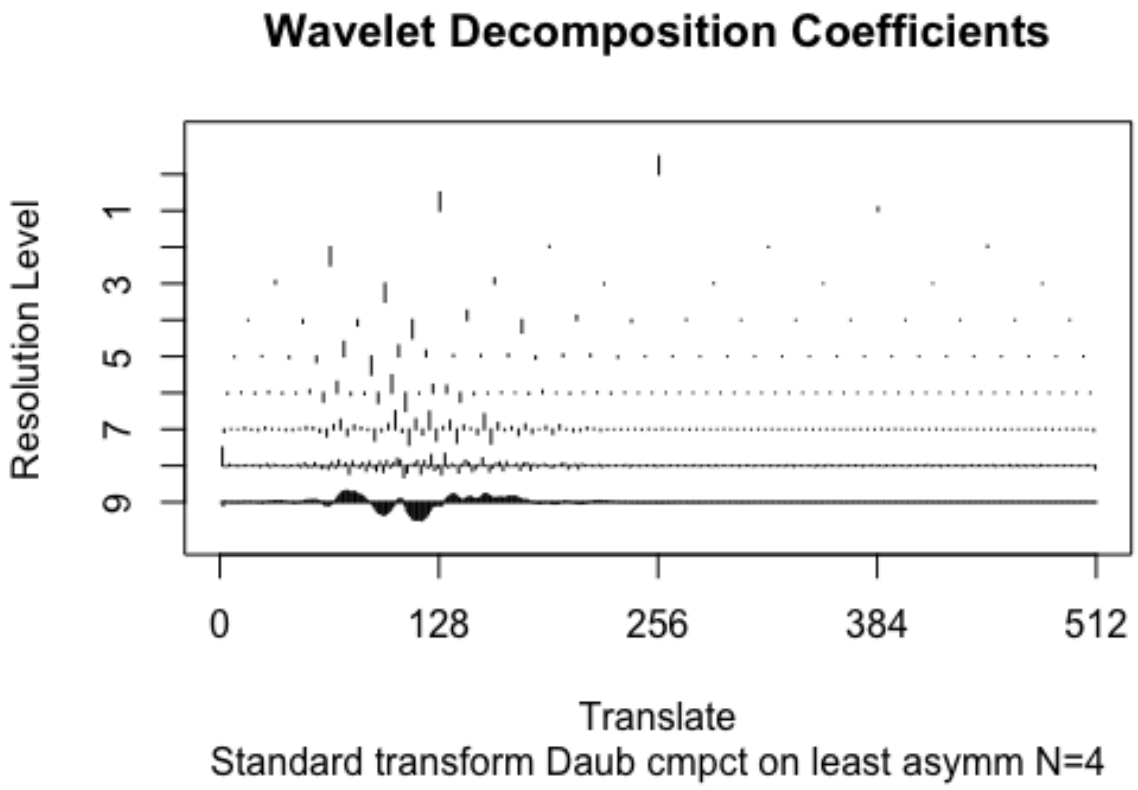


Figure 10: *Scaled decomposition coefficient*

6.3 Data in Wavelet Domain

	V1	V2	V3	V4	V5	…	V1010	V1011	V1012	V1013	V1014	V1015	V1016	V1017	V1018	V1019	V1020	V1021	V1022	V1023
1	-0.0002	-0.0013	0.0001	-0.0000	0.0001	…	0.5792	-2.5624	0.2913	0.0936	-0.0041	-0.0250	-0.3298	-6.6306	-2.3377	0.2380	0.2144	3.2699	0.6969	0.4804
2	-0.0005	-0.0038	0.0002	-0.0002	-0.0002	…	-2.6985	2.1928	0.0204	0.1497	0.6022	-0.0095	-0.1836	-4.9529	0.6108	-0.0070	0.0682	2.1136	0.3558	0.6940
3	-0.0002	-0.0016	0.0002	-0.0001	0.0001	…	-6.8913	-0.4155	-0.0670	0.0646	-0.0228	-0.0070	-0.0065	-2.8368	-1.0363	0.0863	0.1321	1.2428	0.3472	0.2223
4	-0.0001	-0.0004	-0.0000	-0.0000	-0.0000	…	0.1456	0.5182	-0.0319	0.0011	0.0009	0.0006	0.0004	0.0498	0.2881	-0.0352	0.0031	-0.0444	-0.0494	0.0437
5	0.0013	0.0113	-0.0015	-0.0001	-0.0001	…	0.0565	-4.2314	1.3570	-0.0640	0.1005	0.0024	-0.0872	-1.0776	1.1400	-0.5532	0.4810	0.1010	-0.3439	0.6393
6	0.0071	0.0739	-0.0092	0.0000	0.0001	…	-0.0132	-0.0180	0.0051	-0.0061	0.0048	-0.1691	0.2716	0.1387	-0.0195	-0.0100	0.3092	0.0724	0.1170	0.1636
7	0.0096	0.0295	-0.0067	-0.0028	-0.0040	…	0.0918	-0.2506	-1.0886	0.4453	0.6073	-0.7449	0.3289	0.2801	0.2712	-0.1133	0.3454	0.1204	0.2362	0.1297
8	-0.0000	-0.0003	0.0001	-0.0000	-0.0001	…	0.9125	-2.1252	0.0450	-0.0159	-0.0087	0.0023	-0.2059	-3.7004	-1.6852	0.1101	0.1334	1.8506	0.4353	0.3440
9	0.0002	0.0022	-0.0003	-0.0001	-0.0000	…	-0.0146	-0.0066	0.0738	-0.0347	0.0903	-1.0897	-0.6578	0.0416	0.0407	0.1925	0.6022	-0.0120	-0.0244	-0.0722
10	-0.0006	-0.0052	0.0008	0.0001	-0.0002	…	-8.3877	2.3866	-0.2557	0.2372	-0.0006	-0.0050	-0.0672	-5.5938	0.2795	0.0924	0.3416	2.5472	0.4423	0.8249
⋮																				

Table 4: Decomposition coefficient matrix for the first 10 patients

	V1	V2	V3	V4	V5	…	V1010	V1011	V1012	V1013	V1014	V1015	V1016	V1017	V1018	V1019	V1020	V1021	V1022	V1023
1	0.00	0.00	0.00	0.00	0.00	…	0.58	-2.56	0.29	0.00	0.00	0.00	-0.33	-6.63	-2.34	0.00	0.00	3.27	0.70	0.48
2	0.00	0.00	0.00	0.00	0.00	…	-2.70	2.19	0.00	0.00	0.60	0.00	0.00	-4.95	0.61	0.00	0.00	2.11	0.36	0.69
3	0.00	0.00	0.00	0.00	0.00	…	-6.89	-0.42	0.00	0.00	0.00	0.00	0.00	-2.84	-1.04	0.00	0.00	1.24	0.35	0.00
4	0.00	0.00	0.00	0.00	0.00	…	0.00	0.52	0.00	0.00	0.00	0.00	0.00	0.00	0.29	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00	…	0.00	-4.23	1.36	0.00	0.00	0.00	0.00	-1.08	1.14	-0.55	0.48	0.00	-0.34	0.64
6	0.00	0.00	0.00	0.00	0.00	…	0.00	0.00	0.00	0.00	0.00	0.00	0.27	0.00	0.00	0.00	0.31	0.00	0.00	0.00
7	0.00	0.00	0.00	0.00	0.00	…	0.00	-0.25	-1.09	0.45	0.61	-0.74	0.33	0.28	0.27	0.00	0.35	0.00	0.00	0.00
8	0.00	0.00	0.00	0.00	0.00	…	0.91	-2.13	0.00	0.00	0.00	0.00	0.00	-3.70	-1.69	0.00	0.00	1.85	0.44	0.34
9	0.00	0.00	0.00	0.00	0.00	…	0.00	0.00	0.00	0.00	0.00	-1.09	-0.66	0.00	0.00	0.00	0.60	0.00	0.00	0.00
10	0.00	0.00	0.00	0.00	0.00	…	-8.39	2.39	-0.26	0.00	0.00	0.00	0.00	-5.59	0.28	0.00	0.34	2.55	0.44	0.82
⋮																				

Table 5: Threshold coefficient matrix for the first 10 patients

Coefficients	Model1	Model2	model3
Intercept	0.68816417	2.13	1.7986
age	-	-0.0317	-0.0166
anatomy(APB)	-	-	0
gender(Male)	-	0	0
type_ms(PPMS)	-	-	0
type_ms(RRMS)	-	-	-1.0004
963	0	0	-
964	0	0	0
965	0	0	0
966	0	0	0
967	0	0	0
968	0	0	0
969	0	0	0
970	0	0	0
971	0	0	0
972	0	0	-
973	0	0	-
974	0	0	-
975	0	0	-
981	0	0	-
993	0	0	0
994	0	0	0
995	0	0	0
996	0	0	0
997	0	0	0
998	0	0	0
999	0	0	0
1000	0	0	0
1001	0	0	0
1002	0	0	0
1003	0	0	0
1004	0	0	0
1005	0	0	0
1006	0	0	0
1007	0	0	0
1008	0	0	0
1009	0	0	0
1010	0	0	0
1011	0	0	0
1012	0	0	0
1013	0	0	0
1014	0	0	0
1015	0	0	-
1016	0	0	0
1017	-0.0580	-0.079	-0.00158
1018	0	0	0
1019	0	0	0
1020	0	0	0
1021	0	0	0
1022	0	0	0
1023	0	0	0

- means not selected

Table 6: Selected coefficients in the models and their values

6.4 R Codes

```
#write a function for decomposing
decomp1d<-function(x,family="DaubLeAsymm",
filter.number=4, bc="periodic",
min.scale=0,type='wavelet'){
wds<-apply(x,1,wd,
filter.number=filter.number,
min.scale=min.scale,
  family=family, bc=bc,type=type)
}

#decompose the data
mm<-decomp1d(seriesmat)
#extract C matrix
mmc<-sapply(mm, "[",1,simplify = 'array')
#extract D matrix
mmd<-sapply(mm, "[",2,simplify = 'array')
thresh1<-function(x, type = "hard",
levels=0:9,policy = "manual",
boundary = FALSE, verbose = T,
 return.threshold = FALSE,
 value=0.25,dev=madmad){


  tds<-lapply(x,threshold,
  type=type, policy = policy,
  levels=levels,
  boundary =boundary,
    verbose = verbose,
    return.threshold = return.threshold,
    value=value,
    dev=dev)

}

#threshold the coefficients
cc<-thresh1(mm)
```

```
#extract the D matrix
ccc<-sapply(cc, "[",2,simplify = 'array')
ccd<-sapply(cc, "[",1,simplify = 'array')

#reconstruct the coefficicient
recomp1d<-function(x, verbose = FALSE,
bc = 'periodic', return.object = F,
filter.number=4,family = 'DaubLeAsymm'){
  wrs<-lapply(x,wr,
  filter.number=filter.number,
 family=family, bc=bc,
 return.object=return.object)
}
rr<-recomp1d(ccd)

#lasso
grid <-10^seq(10,-2,length=100)

#selecting lambda
set.seed(123456)
cv.outlog <- cv.glmnet(x1log,ytrainlog,
alpha=1,lambda=grid,
type.measure = 'class',family='binomial')
plot(cv.outlog)
#one standard deviation
bestlambdalog=cv.outlog$lambda.1se
#-2.37236
log(bestlambdalog)
```

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
Mathematical model to predict the disease course of Multiple Sclerosis (MS)

Richting: **Master of Statistics-Epidemiology & Public Health Methodology**
Jaar: **2018**

in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Voor akkoord,

Adesoji, Oluyomi Modupe

Datum: **12/03/2018**