Applied Statistical Modelling and Inference in Ophthalmology:

Analysis of visual field and video data for glaucoma patients.

A thesis presented in total fulfilment of
the requirements for the degree of
Doctor of Philosophy in Statistics
at Massey University, Manawatu, New Zealand.

Brigid Betz-Stablein

March 13, 2014

Abstract

Eyesight is arguably the most important of our senses with the eye absorbing 80% of external information from our surroundings. The field of ophthalmology studying the anatomy, physiology and diseases of the eye, is of extreme importance. Many methods exist to measure vision and the eye, creating a large range of interesting datasets. We developed methods to analyse three datasets from subjects with glaucoma, the second leading cause of blindness worldwide.

Visual field testing using standard automated perimetry, is the most common method for monitoring glaucoma progression. A numerical matrix representing the dimmest intensity seen by a particular locus on the eye is outputted. This can be thought of as a map, and disease mapping techniques applied. We employed conditional autoregressive priors to account for the spatial correlation structure in the visual field results, in a way that respects the physiological and optical properties of the eye. Model diagnostics showed our model superior to the currently used point-wise linear regression methods.

Visual field mean deviation, the mean light intensity across all loci adjusted for age matched controls, provides a global estimate of glaucoma progression. We investigated the shape of the relationship between mean deviation and time over long series of visual fields using splines. We considered imposing a monotonic non-increasing constraint. When a curve deviated from being linear or monotonic non-increasing, this was an indication of physiological or treatment change in the eye.

We developed methods to extract and analyse data from video sequences of retinal venous pulsation, observed as change in blood flow, varying with the cardiac cycle. Video sequences were divided into individual frames, and the mean pixel intensity was calculated separately for three vessel segments representing the artery, lower vein and upper vein. Simple harmonic

terms modelled the periodic component of the trend. The non-periodic trend, caused by patient movement, was modelled by linear splines. An autoregressive process modelled error correlation. Retinal blood flow has been linked to many diseases, so the characteristics of these curves have clinical importance.

Publications arising from this thesis

Betz-Stablein, B., Morgan, W.H., House, P.H., Hazelton, M.L. (2013). Spatial modeling of visual field data for assessing glaucoma progression. *Investigative Ophthalmology & Visual Science* **54(2)**, 1544-1553.

Acknowledgements

Foremost, I would like to thank my primary supervisor Prof. Martin Hazelton. Martin has gone above and beyond the duties of a supervisor. I am very thankful for his unending patience, enthusiasm and broad knowledge. Martin has supported me through every stage in my thesis, including those unexpected, and provided me with guidance I could not have done without. Thanks Martin, I could not have hoped for a better supervisor.

In addition I would like to thank my secondary supervisor Prof. William Morgan. Without Bill we would have had no data and no interesting statistical problems to solve. Bill has had enthusiasm for this thesis from the start. I am thankful to him for all his expertise in ophthalmology and the support he has provided me throughout my journey. I would also like to acknowledge the support of Dr. Philip House who also was involved with data collection and provided feedback on the visual field work.

I would like to thank my family for their lifelong support. In particular I am thankful to my parents Prof. Ralph Stablein and Dr. Joy Panoho for the love and time they have invested in me. They have provided an example of achieving excellence which motivated and led me to pursue my PhD. Through the ups and downs of my thesis, and life, they have always been there with encouragement, advice and if required tissues and mallow puffs. I love you both very much.

My thanks also go to the other postgraduate students, especially Kate Richards, with whom I have been able to talk when I needed to talk and share silence when silence was required. Thanks for making my various offices a place I wanted to come to everyday and joining me on the required post-graduate pursuit of free food. I am thankful to my friends at VCC who have walked this journey with me, in particular my flatmates who have helped me through the final push of these last few months.

My thanks goes to Massey University and the Institute of Fundamental Sciences for their financial support. I would also like to acknowledge the support of the secretarial and computing staff, particularly Colleen Blair and Peter Lewis.

Finally I would like to thank God who had bought me immense strength, joy, hope and wisdom. I am thankful for the challenges He has used through this period of my life to draw me closer to Him, and show me that 'I can do all things through Christ who strengthens me' (Phillipians 4:13).

List of Abbreviations

AGIS	Advanced Glaucoma Intervention Study
AIC	Akaike information criterion
asb	Apostilbs
AUC	Area under the curve
BLUP	Best linear unbiased predictor
BYM	Besag, York and Mollie model (disease mapping)
CAR	Conditional autoregressive
CIGTS	Collaborative Initial Glaucoma Treatment Study
dB	Decibels
DIC	Deviance information criterion
DLS	Differential light sensitivity
EMGT	Early Manifest Glaucoma Trial
GPA	Glaucoma progression analysis
IOP	Intraocular pressure
ICP	Intracranial pressure
LTF	Long term fluctuation
MCMC	Markov chain Monte Carlo
MD	Mean Deviation
OAG	Open-angle glaucoma
ODF	Ophthalmodynamic force
PLR	Point-wise linear regression
POAG	Primary open-angle glaucoma
POBF	Pulsatile ocular blood flow
REML	Residual maximum likelihood estimation
RGB	Red-green-blue colour model
ROC	Receiver operating characteristic
SAP	Standard automated perimetry
SITA	Swedish interactive threshold algorithm (SAP method)
SMR	Standard mortality ratio
SPROG	Our model in Chapter 4 (Spatial Progression)
STF	Short term fluctuation
SVP	Spontaneous vein pulsation
VPSG	Vein Pulsation Study Trial in Glaucoma
VF	Visual Field

List of Figures

List of Tables

Contents

Chapter 1

Introduction

Statistics is critical to the field of medicine. Statistics are used to calculate prevalence of disease as well as monitor disease epidemics. The application of statistics in medicine has led to the identification of major risk factors for diseases such as cardiovascular disease, the leading cause of death in the United States. New drugs and treatments will not be approved unless they have undergone the rigorous statistical process of a randomized clinical trial. In a randomized clinical trial statistical principles are employed in designing the trial, collecting and analysing the data, and interpreting the results. Just as statistics are useful in medicine, medical datasets are also valuable for the development of statistical methods. Many interesting statistical problems arise from the analysis of medical data, and ophthalmic datasets are no exception.

The eye provides 80% of external information absorbed through our senses, hence this complex organ is paramount. With the cost of blindness being huge to both the individual and society, preventing blindness is extremely important. Many tests and measures are employed to monitor vision and the eye, thus resulting in a wide range of ophthalmic datasets. These datasets give rise to many interesting and challenging statistical problems. Often ophthalmic datasets contain repeated measures and are longitudinal. Many contain data from both eyes, which requires correlation to be taken into account. Imaging techniques used for diagnostics require their own methods of analysis. Currently many of the datasets are analysed with-

out gold standard on an event basis, which gives plenty of scope for removing subjectivity and creating analytic statistical models. Therefore this thesis focuses on novel approaches to analysing ophthalmology data.

In this thesis our data comes from subjects with glaucoma, the second leading cause of blindness in the world. Glaucoma is caused by a build up of fluid in the eye, which increases the eye's intraocular pressure, resulting in cell death. Once glaucomatous damage had occurred, vision loss is irreparable. About 3% of the general population has glaucoma, with the prevalence increasing with age. While glaucoma cannot be cured, treatment can prevent further vision loss. Due to the irreversible nature of glaucoma diagnosis and treatment of this disease (aimed at preventing progression), is extremely important. Artes & Chauhan (2005) state that one of the most challenging aspects of managing glaucoma is determining whether or not progression has occurred. This is due to the subjectivity and variability associated with the testing methods. We provide background information on glaucoma in Chapter 2.

In Chapter 2 we also introduce Bayesian statistics. Bayesian methods are being increasingly used in statistical applications. They can be more flexible and do not require the rigid assumptions of many frequentist methods. In addition Bayesian analysis allows for the incorporation of prior information into the model. Bayesian methods are generally implemented through the iterative process Markov Chain Monte Carlo (MCMC). MCMC methods sample model parameters and hence provide the means to approximate any desired characteristics of the underlying Bayesian distribution. These are also introduced in Chapter 2.

The most common method for testing progression in glaucoma is Standard Automated Perimetry (SAP). SAP tests areas across the visual field to see if there has been any functional vision loss. A threshold score of the dimmest intensity that can be seen at each locus on the visual field is recorded. We discuss this in detail in Chapter 3. While SAP is standard practice in both clinical practice and research trials, there is no gold standard for assessing this data. The current analysis methods can be divided into 3 categories; subjective analysis, event-based

analysis and trend based analysis. Subjective analysis generally involves classifying the field as either progressing or not. Event-based analysis compares the current field to a baseline field, while trend-based analysis uses a sequence of fields. From a statistical point of view trend-based analysis is the most appealing as the decision of progression is based on all the information. The standard analytical software PROGRESSOR uses pointwise linear regression to analyse serial visual fields. This method does not take into account the correlation structure of loci in the eye, and struggles to deal with the large amounts of variation in SAP data. In addition problems of multiple testing are introduced when performing multiple (52) regressions on one eye.

Visual field threshold data can be likened to a map, as the 52 loci create a matrix, with coordinates corresponding to their location on the visual field. Applying disease mapping techniques to the visual field matrix allows us to correctly account for the correlation between loci in the eye. However, the physiology of the eye introduces more complexity to this statistical problem. Whilst points sit next to each other on the visual field, they do not directly map to the optic disc. In addition each eye has a blind spot, where there are no cells to absorb light. The blindspot on each eye therefore acts independently to the other loci of the visual field. Within the disease mapping framework, we must also consider how to model visual field trend and incorporate measurement error into the model. We explain the necessary theoretical background and our novel approach to modelling visual field data in Chapter 4.

Longer series of visual fields are investigated in Chapter 5. Of interest is whether longer series of visual field summaries are adequately modelled by linear trends. We apply these to the mean deviation, which is the mean across all test loci of the subject's deviation from their age matched controls. As mentioned, once glaucomatous damage has occurred, vision loss is irreversible. We would therefore expect the trend to be non-increasing. Investigating the shape relationships between mean deviation and time is important from a clinical basis, as not only can progression be monitored, but appearance of an unusual curve (i.e. sigmoidal or

convex) would indicate something unusual happing to the subjects eye. In the case of non-linearity splines provide a useful and flexible tool to investigate time series data. However these splines must be constrained to meet the condition of being non-increasing. We compare linear and both constrained and uncontrained spline regressions on 18 serial visual field datasets. A post-hoc analysis is carried out to see if the behaviour of the curves is indicative of any physiological or treatment changes in the eye.

Retinal blood flow is the only vascular blood flow that can be viewed in vivo, as the retina can be viewed through the pupil. Modelling blood flow has important medical applications as changes in retinal blood flow characteristics have been linked to many diseases such as diabetes and coronary heart disease. The change in blood volume in the retina is observed through retinal venous pulsation. Video data of this phenomenon was obtained over several cardiac cycles. Initially image processing techniques must be considered to extract numerical information from the video images. Our aim is to extract a representation of the periodic variation in retinal venous pulsation, due to the change in blood flow varying with the cardiac cycle. Additional problems such as subject movement during filming must also be considered. These challenges are discussed in Chapter 6 where we explain our method to extract and analyse periodic video data.

Finally our conclusions and suggestions for future research are presented in Chapter 7.

Chapter 2

Background

2.1 Ophthalmology

Eyesight is taken for granted by many, however it is the most used and most relied on of senses with approximately 80% of the information we take in from our surroundings acquired through sight (Klein 2010). The field of ophthalmology is devoted to studying the anatomy, physiology and diseases of the eye. An ophthalmologist specialises in medical and surgical treatments for the eye. In this section we provide some background information on how the eye functions to create sight. The data from which we build our statistical models comes from glaucoma patients. Therefore we introduce the disease, and discuss its testing, progression and treatment.

2.1.1 The Eye

The act of seeing begins when light enters the eye. The photo-receptors (rods and cones) in the retina absorb light, passing this visual information to the ganglion cells. The ganglion cells transfer this information to the brain via electrical nerve pulses through the approximately one million nerve fibres connecting the ganglion cells to the brain (See Figure 2.1). These nerve fibres spread out in the inner layer of the retina, converging at the optic disc. The optic disk contains axons which carry electrical impulses, but no photo-receptors. This results in a small part of the retina which cannot see. This location is known as the blind-spot. While

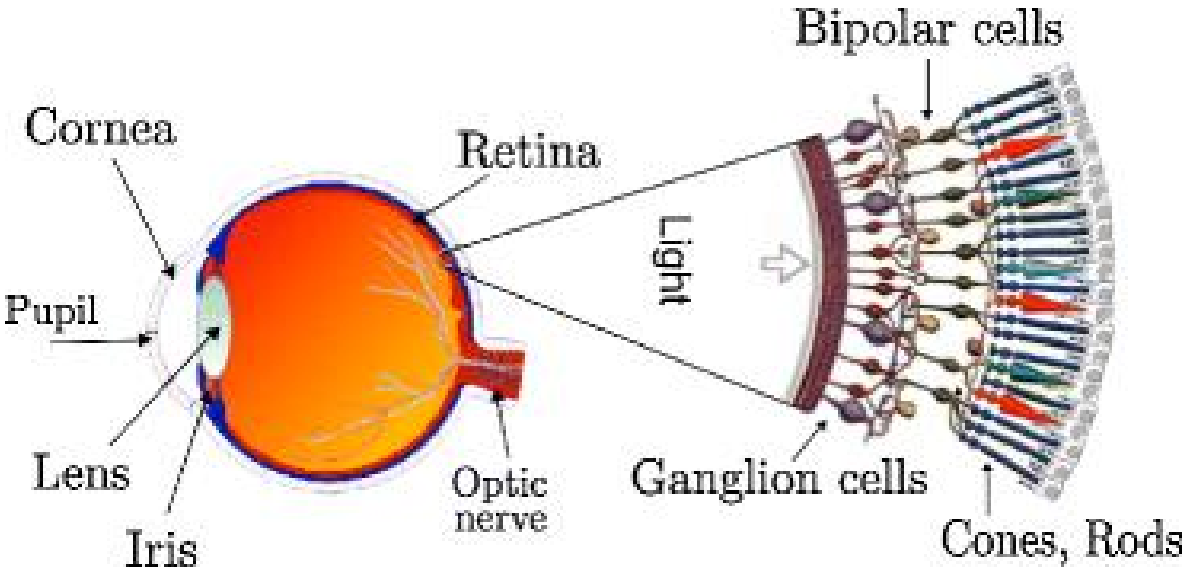


Figure 2.1: Diagram of the eye showing how light is absorbed through the rods and cones and transferred to the optic nerve via the ganglion cells. Reprinted with permission from Springer (Benosman 2010). Licence number: 3235670334509.

the brain compensates for this missing information, the blind spot is identified in visual field testing (an important diagnostic test for glaucoma, discussed below) (Flammer 2003).

2.2 Glaucoma

Glaucoma is the second leading cause of blindness, behind age-related macular degeneration in the developed world and cataracts in the developing world. Glaucoma accounts for 12.3% of blindness world wide (Resnikoff et al. 2004). It is estimated that around 3% of the general population have glaucoma, however only about half of these people are aware of the disease, and even less are receiving adequate treatment (Flammer 2003). Since the risk of glaucoma increases with age, the prevalence of glaucoma also increases when examining older populations.

2.2.1 Pathophysiology

As a person ages some retinal cells die, however in glaucoma the death occurs at an advanced rate (Flammer 2003). Glaucoma is caused by a build up of fluid in the eye, which in turn

causes cupping of the optic disk. Cupping occurs due to the death of retinal ganglion cells leading to permanent, irreparable vision loss (Artes & Chauhan 2005). Figure 2.2 compares a healthy optic disc with one that has cupping damage. The light area indicated by the arrow in Figure 2.2 B shows an optical cup. While the light is still absorbed by the photo-receptors, transmission of the visual information does not occur as the connection between the eye and brain is effectively severed (Flammer 2003).

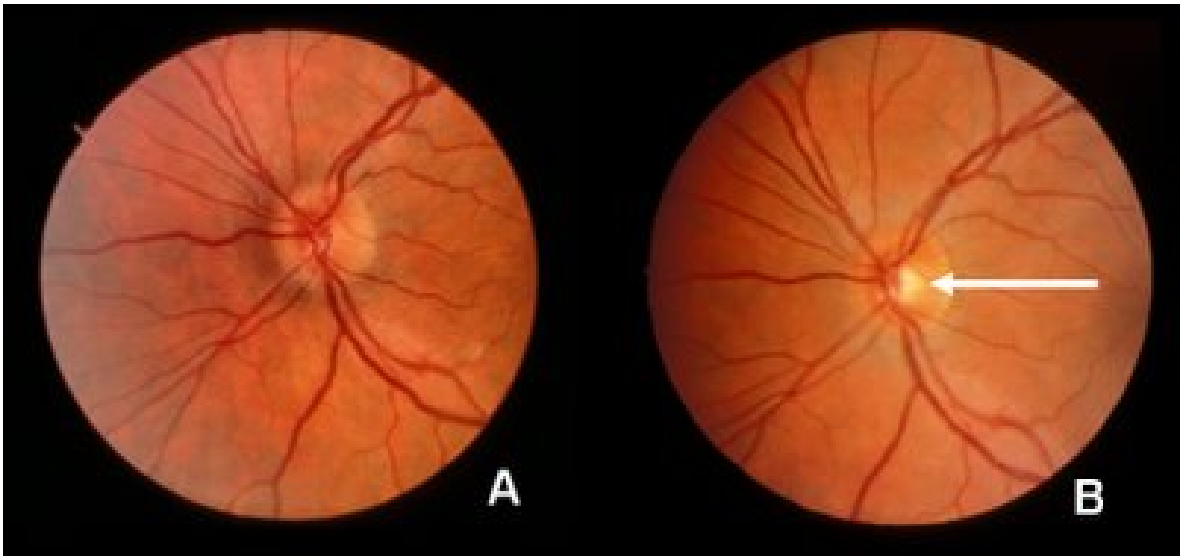


Figure 2.2: A shows a normal optic disc. In B cupping can be seen by the lighter region indicated by the arrow. Reprint permission requested from FEP. (Kwon et al. 2008).

2.2.2 Diagnosis

Generally glaucoma is diagnosed when a person has elevated intraocular pressure (IOP), visual field loss and optic disk and/or nerve damage. A person with ocular hypertension (high IOP), but no functional or structural damage is referred to as a glaucoma suspect. Several factors are taken into account when diagnosing glaucoma. These include assessment of the optic disk, visual field testing, measurement of IOP and demographic factors (Artes & Chauhan 2005). Glaucoma is difficult to diagnose, particularly in the early stages of the disease as patients can have severe visual field loss, but still have 20/20 vision. Generally by the time a patient is aware of the disease, it is quite advanced, as the brain compensates for holes in the visual field.

One of the most challenging aspects of glaucoma management is determining whether an individual's eye has progressed (Artes & Chauhan 2005). This is made more difficult by the lack of a gold standard for both testing and analysing glaucoma data, as well as lack of consensus on the clinical definition of a progressing eye (Chauhan et al. 2008, Vesti et al. 2003). Important decisions such as modifying the subject's therapy are made based on determining if an eye is stable or progressing (Artes & Chauhan 2005).

Progressive cell loss can be observed in both structural and functional changes. Structural damage is observed in the cupping and excavation of the optic disc. A common method to detect this structural loss is to examine photographs of the optic disc, known as fundus photographs (see Figure 2.3). The ophthalmologist must decide whether there had been cupping and if this cupping is normal or pathological (Flammer 2003). Functional loss is observed through vision loss transversing the visual field. The most common method to detect visual field loss is through Standard Automated Perimetry (SAP) which is described in detail in the following chapter (Section 3.2).

2.2.3 Risk factors

Risk factors for glaucoma include elevated IOP, increased age, family history, being of African descent, myopia and low diastolic perfusion pressure (Kwon et al. 2009). Specific genetic mutations are also being linked to glaucoma (Rivera et al. 2008). IOP is the only treatable risk factor, therefore treatments focus on lowering IOP.

Intraocular Pressure (IOP)

Movements such as blinking or changing eye direction, put pressure on the eye. IOP helps maintain the eye's shape. It prevents eye tissue from swelling by forcing the fluid containing cell 'garbage' back into the blood stream. Both the lens and the cornea have no blood vessels, as this would impair vision. Therefore they need to be constantly irrigated. This is achieved through aqueous humor. IOP is maintained by balancing the production of fluid by ciliary

Figure 2.3: Fundus photograph of normal optic disc.

body and the drainage of that fluid via trabecular meshwork within the anterior chamber angle. Most glaucoma is caused by blockage of structures within this 'angle'. See Figure 2.4.

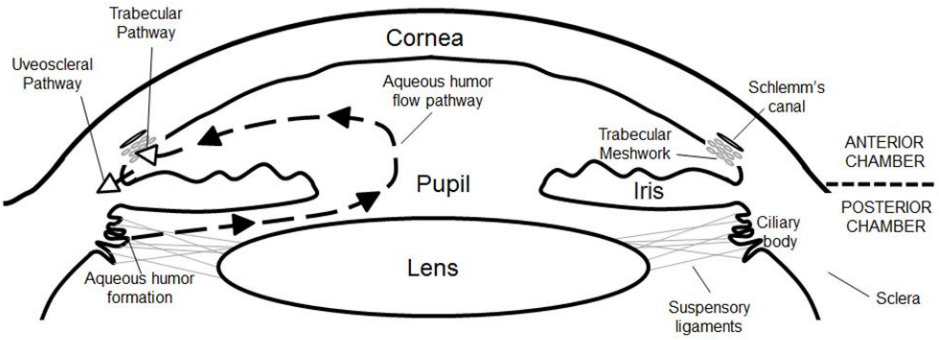


Figure 2.4: Diagram of aqueous humor pathway in the anterior and posterior chambers of the eye. A disruption in this pathway leads to changes in IOP, the only treatable risk factor in glaucoma. Reprinted from Yoko & Walter (2013), open source.

Most people have a normal IOP of about 15mmHg. This ranges from 9-21mmHg. Patients with glaucoma have a higher mean IOP (see Figure 2.5). An increase in IOP can be caused by an increase in production of aqueous humor, or a resistance to outflow. Without treatment, elevated IOP leads to cellular stress, structural change including cupping, and damage/loss of the retinal ganglion cells through apoptosis (Kwon et al. 2009). Whether a certain level of IOP leads to glaucoma damage usually depends on other risk factors. Note that in addition to an increased IOP, glaucoma patients also have increased fluctuations of IOP (refer to Figure 2.6), so in order to measure the peak IOP multiple measures of IOP at different times of the day are required (Flammer 2003, Walsh 2011).

2.2.4 Treatment

Once vision loss has occurred there is no treatment to correct this loss. Because eye cells are part of the central nervous system, when they die they are not replaced, unlike other cells in the body such as muscle, or skin. Therefore once the damage occurs it is permanent and cannot be corrected (Flammer 2003, Kwon et al. 2009). However early detection and treatment of glaucoma can minimize and prevent irreversible field loss. Clinical trials have shown that by reducing a person's IOP, the onset and progression of glaucoma can be delayed. Therefore all

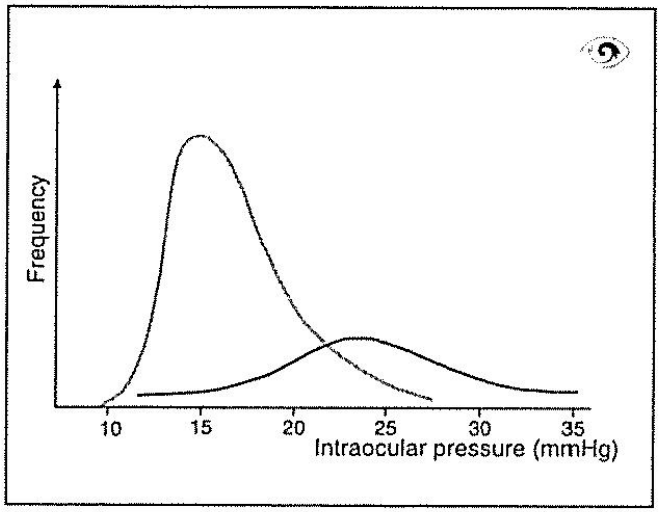


Figure 2.5: IOP distribution curves for patients with and without glaucoma. The flatter curve with a higher range is represents glaucoma patients, while the curve centered on 15 reflects the normal population. Reprinted with permission from Glaucoma, 2nd edition, by Josef Flammer, ISBN 0-88937-269-1, p.81.

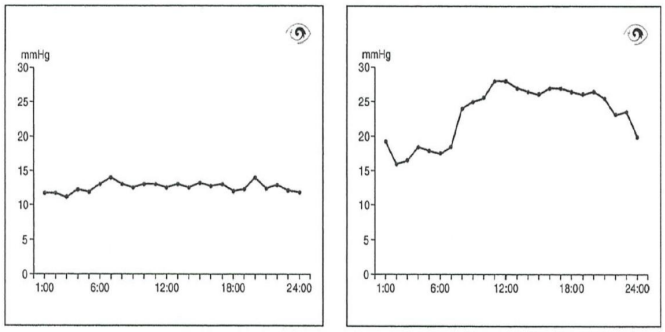


Figure 2.6: Fluctuations in IOP over a 24 hour period for a normal patient (left) and a patient with glaucoma (right). Reprinted with permission from Glaucoma, 2nd edition, by Josef Flammer, ISBN 0-88937-269-1, p.34.

current treatments focus on reducing the IOP through medical or surgical procedures (Kwon et al. 2009). The first line of treatment is lowering IOP through eye drops. If progression continues laser treatment or surgery may be pursued.

2.2.5 Classification of Glaucoma

There are three types of glaucoma. These are open-angle glaucoma, closed-angle glaucoma and developmental types (Kwon et al. 2009). These categories are divided further into primary and secondary categories. Glaucomas are categorized by the cause of elevated IOP. Primary open-angle glaucoma (POAG) is the most common in the western world. Congenital glaucoma onsets at birth and is due to the chamber angle not being fully developed. Infantile or juvenile glaucoma onsets during infancy or adolescence, and is caused by an abnormal angle of the chamber. In POAG the angle of the chamber is normal, but there is an increased IOP due to strong resistance in fluid outflow (throughout the trabecular meshwork, see Figure 2.4). In Primary Angle-Closure Glaucoma a complete or partial barrier exists, created by the iris, which prevents fluid outflow. The onset of Primary Angle-Closure Glaucoma is often fast and symptoms quite severe. There are also many diseases, injuries, operations and treatments that lead to an increase in IOP and ultimately glaucoma. These are termed secondary glaucomas (Flammer 2003). The majority of subjects from whom we have data have been diagnosed with POAG.

2.2.6 Retinal Venous Pulsation

Spontaneous Vein Pulsation (SVP) occurs due to the collapse and refilling of the retinal veins (Donnelly & Subramanian 2009). It is suggested that SVP occurs due to the pressure gradient between cerebrospinal fluid and IOP, which varies in response to the cardiac cycle (Morgan et al. 2008, 2004). Donnelly & Subramanian (2009) suggests that SVPs occur when intraocular pulse pressure is greater than cerebrospinal fluid pulse pressure. Morgan et al. (2004) show that people with glaucoma are less likely to have SVP than glaucoma suspects or normal eyes. They showed that SVP was seen in 54% of glaucoma subjects, 75% of glaucoma

suspects and 98% of those with normal eyes. These differences were statistically significant (P<0.001). In subjects without SVP, venous pulsation can be induced by a method known as ophthalmodynamometry, where a measured force is applied to the eye. This measured force is called ophthalmodynamic force (ODF). Retinal venous pulsation, spontaneous or induced, is discussed further in Chapter 6.

2.3 Modelling Background

The majority of the methods we develop within this thesis lend themselves to a Bayesian framework. Bayesian models are becoming more and more commonplace in biostatistics as they can offer advantages over the traditional frequentist methods. In this section we introduce Bayesian statistics and briefly discus Markov Chain Monte Carlo (MCMC) algorithms. Specific details of the methods will be discussed as they are used within the thesis.

2.3.1 A Brief History of Bayesian Statistics

While Bayes' Theorem was initially conceived in 1763 (Bayes 1763), Bayes' Theorem, as we know it today, was first published by Laplace in 1774 (Fienberg 2006). Bayes' Theorem supposes that we have a prior belief about an event. We then update this belief with information from our data to form a posterior belief. We assume $\{B_i\}$ to be a mutually exclusive and exhaustive set of events and formally define Bayes' Theorem as:

$$Pr(B_i|A) = \frac{Pr(A|B_i)P(B_i)}{\sum_j Pr(A|B_j)Pr(B_j)}, \tag{2.1}$$

where $Pr(B_i|A)$ is the conditional probability of event B_i, given A.

Bayes' Theorem can be interpreted in terms of beliefs where A is our evidence based on data, and $\{B_i\}$ is our set of beliefs. However Bayes' Theorem is applicable regardless of how the probabilities are defined. Instead of being interpreted as a set of beliefs, $\{B_i\}$ can also be based on long run of relative frequencies. This gives rise to ideas of objective and subjec-

tive probability, which were developed around the 1840s. It was then that the distinction arose between the probability of something, and one's personal belief about the probability of something. In 1907 Pearson (Pearson 1907) explained the importance of using *a priori* probabilities in applied situations, where past experience can contribute to the current analysis.

Between 1912 and 1922 Fisher developed his own approach to statistical inference - the likelihood. The likelihood is the probability of the data viewed as a function of model parameters. This was a move away from Bayesian probability and towards 'frequentist' statistics - a word which arose when the term 'Bayesian' came into common use (Fienberg 2006). Bayesian statistics combines the prior and the likelihood through Bayes' Theorem to form a posterior distribution from which inference is made.

World War II and the post-WWII effort saw the development of a Bayesian approach to sequential data analysis, as well as Empirical Bayesian ideas (an approximation to the full Bayesian method) and the Monte Carlo method. However at this time frequentist statistical methods were still dominant. During 1950s and 1960s the methodology continued to develop and the adjective 'Bayesian' began to appear in publications. Nonetheless, progress in conducting Bayesian analyses was handicapped by the computational difficulty of implementing Bayesian methods.

The introduction of Monte Carlo Markov chain method made it possible to undertake analysis of previously uncomputable problems (Fienberg 2006). Metropolis et al. (1953) first introduced the idea of MCMC sampling for numerical simulations in chemistry. Hastings (1970) generalised Metropolis's method to the statistical problem of sampling from a posterior distribution. Since then the improvement in computing power has increased the use of both MCMC and Bayesian methods, which are now commonplace in many areas of applied science.

2.3.2 Basics of Bayesian Statistics

Bayesian statistics arises from the extension of the likelihood paradigm. In Bayesian statistics we allow the parameters within the likelihood to have their own marginal distributions. These distributions are known as priors. Unlike frequentist statistics where a parameter is fixed and the data is random, Bayesian statistics treats the parameter as stochastic (Congdon 2001). This allows for the priors to have parameters with distributions of their own (hyper-priors) creating a natural hierarchical modelling framework. By specifying our prior belief as a probability distribution we create a joint probability model from the product of our prior and the likelihood distribution:

$$p(\theta, y) = f(y|\theta)\pi(\theta), \tag{2.2}$$

where $f(y|\theta) = L(\theta)$ is the model likelihood based on data y. From Equation 2.2 we can obtain the posterior distribution for θ given the data. That is

$$p(\theta|y) = \frac{f(y|\theta)\pi(\theta)}{h(y)}, \tag{2.3}$$

where $h(y)$ is the marginal density of the data. This posterior can be regarded as updating our prior beliefs about θ (as specified by π) in light of the data. See Figure 2.7 for an illustration. In principle $f(y|\theta)$ is the target in Bayesian inference, although in practice it is often summarised, e.g., by the posterior mean of θ, and the posterior quantiles, etc.

Choice of Priors

Prior distributions can be thought of as the current idea of how much variation is in the parameter set (Lawson 2009). The choice of prior depends on how much knowledge we have about the parameters. In the absence of useful prior knowledge a vague prior is chosen. For a general case a uniform prior or normal prior with a large variation can be applied. Increasing the spread of the prior results in a flatter prior, which is less informative. In medical datasets, such as ophthalmic data, we often have prior knowledge. For example, clinical expert opinion can provide helpful bounds on plausible effect sizes, such as the rate of vision loss per year. In such cases an informative prior can be used. In this case a normal prior can be used, centred

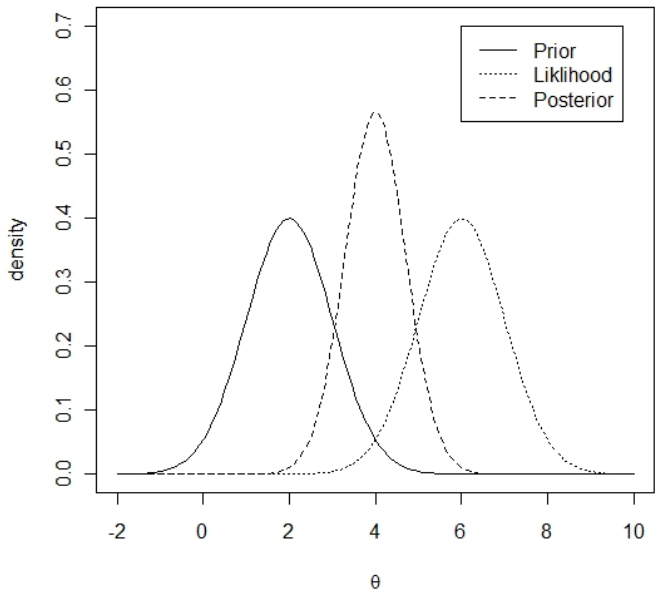


Figure 2.7: Prior, likelihood and posterior distribution from a normal/normal model with one observation.

on the *a priori* most probable value. In principle the variance of an informative normal prior perfectly reflects the precision of our knowledge about the parameter in question. However, in practice the variance is often inflated to reduce perceived subjectivity.

Prior choice also depends on model type. For logistic regression coefficients Gelman et al. (2008) recommends the use of the Cauchy prior as a default prior, with mean 0, and a scale of 10 for the constant and 2.5 for the other coefficients. This distribution is applied to variables which have been scaled to have a mean zero and a standard deviation of 0.5. Gelman et al. (2008) also recommends these priors for linear regression, as do Liang et al. (2008). However for standard linear regression the response variable must also be scaled. This prior is stable and only weakly informative, thus providing minimal prior knowledge. While Gelman et al. (2008) found the scale 0.75 to be optimal, they recommend using a scale of 2.5 as this is a more conservative approach. The scale of 2.5 weakens the prior, therefore bringing Bayesian posterior mode estimates closer to the traditional maximum likelihood.

A common choice for a prior distribution for a variance parameter is the inverse gamma prior, which is related to the scaled inverse chi-squared distribution (Gelman et al. 2003). This is the recommended prior for use in conditional autoregressive priors which will be introduced in Section 4.2.1 (Besag et al. 1991). Alternative diffuse priors, such as applying a uniform distribution on the standard deviation, or a half-t distribution are discussed by Gelman (2006).

2.4 Markov Chain Monte Carlo Methods

Many statistical problems involve complicated and high dimensional probability distributions, from which we wish to make inference. In some situations the posterior distribution is not available in closed form, typically because the necessary normalising terms are integrals that are very difficult to evaluate.

Recall from Equation 2.3 that the posterior distribution for θ can be written:

$$p(\theta|y) = \frac{f(y|\theta)\pi(\theta)}{h(y)}, \tag{2.4}$$

where $h(y) = \int f(y|\theta)\pi(\theta)d(\theta)$ is the normalising constant. Markov Chain Monte Carlo (MCMC) is a method that uses an iterative process to simulate random variables from complex distributions. MCMC allows sampling from the posterior distribution without having to explicitly evaluate $h(y)$. This facilitates evaluation of complex hierarchical models where it is not possible to use traditional 'numerical integration or analytic approximation techniques' (Brooks 1998).

2.4.1 Markov Chains

A Markov Chain is a stochastic process or sequence of random variables, where the transition between values depends only on the previous variable. Formally, $\theta_n : n = 0, 1, 2, \ldots$ is a

Markov chain if

$$Pr\left(\theta_{n+1} \in A|\theta_n, \theta_{n-1}, \ldots, \theta_0\right) = Pr\left(\theta_{n+1} \in A|\theta_n\right), \tag{2.5}$$

for all sets A. With each iteration (n) a new value is added to the chain. The probability of moving from the current value to the new value is given by the transition probability. At any point, n, we can compute the unconditional probability $Pr(\theta_n \in A)$ for any A. Initially, as the chain evolves, its marginal probabilities will usually change. However, given certain conditions on the Markov chain, as $n \to \infty$, $Pr(\theta_n \in A)$ will no longer be dependent on n, so that $Pr(\theta_{n+1} \in A) = Pr(\theta_n \in A)$, and hence the Markov chain becomes stationary. Convergence is attained when a chain is stationary, therefore post convergence the chain consists of a sample of dependent values from a stationary distribution (Gamerman & Lopes 2006).

2.4.2 MCMC algorithms

Two common MCMC sampling schemes are the Gibbs sampler (Geman & Geman 1984) and the Metropolis-Hastings (MH) algorithm (Metropolis et al. 1953, Hastings 1970). Given an initial value, the MCMC algorithm samples a candidate value from a proposal distribution, q. The candidate value is then accepted or rejected with an acceptance probability, α. The probability of moving from the current value θ^t to the candidate value θ^* is:

$$\alpha(\theta^t, \theta^*) = \min\left[1, \frac{L(\theta^*)\pi(\theta^*)q(\theta^t|\theta^*)}{L(\theta^t)\pi(\theta^t)q(\theta^*|\theta^t)}\right], \tag{2.6}$$

where L is the likelihood and π the prior distribution. This is equivalent to

$$\min\left[1, \frac{p(\theta^*|data)q(\theta^t|\theta^*)}{p(\theta_t|data)q(\theta^*|\theta^t)}\right], \tag{2.7}$$

as the normalising constant h from Equation 2.4 cancels out on the top and bottom. The Gibbs sampler is a special case of the MH algorithm where every candidate value generated is accepted i.e. $\alpha = 1$ (Gamerman & Lopes 2006). The critical idea is that θ^t is a Markov chain whose stationary distribution in the required posterior distribution. It follows that sampling

iteratively using MCMC will eventually produce samples from the posterior.

A simple MCMC algorithm is described below. To sample the vector $(\theta, \xi)^T$, the algorithm is as follows:

1. Initialize θ^0 and ξ^{t-1}. Set $t = 1$.
2. Generate candidate $\theta^* \sim q_\theta$, conditional on ξ^{t-1}.
3. Calculate the acceptance probability α as in Equation 2.6.
   Set $\theta^t \leftarrow \theta^*$ with probability α; otherwise $\theta^t \leftarrow \theta^{t-1}$.
4. Generate candidate $\xi^* \sim q_\xi$, conditional on θ^t.
5. Calculate the acceptance probability π.
   Set $\xi^t \leftarrow \xi^*$ with probability π; otherwise $\xi^t \leftarrow \xi^{t-1}$.
6. Set $t \leftarrow t + 1$.
7. Repeat steps 2-6 until required number of iterations complete.

The above algorithm describes component-wise updating, where each component is accepted or rejected with its own probability conditional, on the other elements. Block updating can be carried out and can be more computationally effective. In the block case the candidate is a vector $(\theta, \xi)^T$ which is accepted or rejected as a whole.

The choice of proposal distribution depends on the type of sampler we are using. For an independence sampler each new value is selected from the proposal distribution independent of the previous value. The problem with an independence sampler is the choice of proposal distribution. Ideally the proposal distribution should approximate the target, i.e. the posterior, but in practice we do not know the posterior. A common approach is to use the prior as the proposal distribution. Often, in the case of diffuse or dispersed priors, independence sampling results in low acceptance rates.

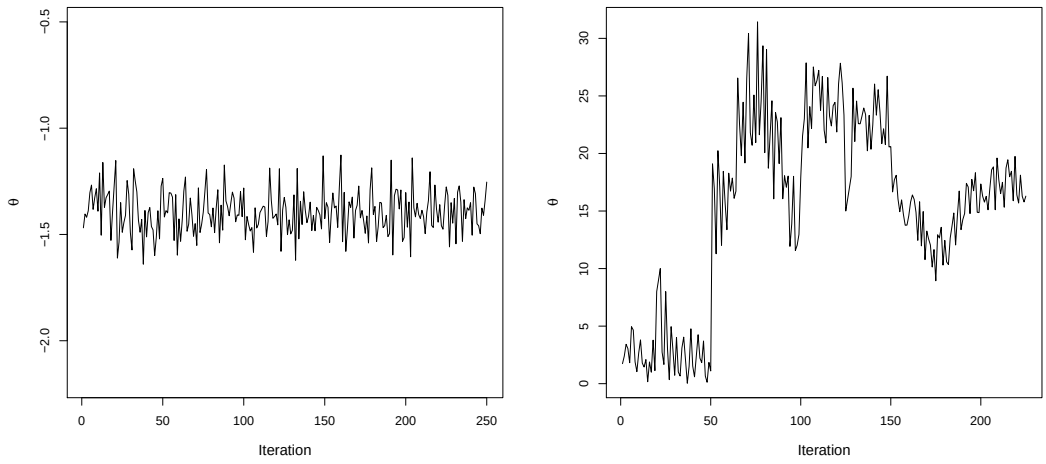
An alternative to independence sampling is a random walk algorithm, in which the new value chosen is a step away from the previous value. Formally, $q(\theta^t|\theta^*) = q_1(\theta^* - \theta^t)$. Here the candidate value is $\theta^* = \theta^t + z$, where z is a random increment drawn from proposal distribution distribution q_1. As the proposal distribution is symmetric, the acceptance probability reduces to:

$$\alpha(\theta^t, \theta^*) = \min\left[1, \frac{L(\theta^*)\pi(\theta^*)}{L(\theta^t)\pi(\theta^t)}\right]. \tag{2.8}$$

2.4.3 Assessing Convergence

Recall that the MCMC algorithm produces a sequence of samples $\theta^1, \theta^2, \ldots, \theta^t$ which form a Markov chain. For sufficiently large t the marginal distribution of θ^t will converge to the posterior. However, we need to judge when convergence has occurred. Figure 2.8 shows trace plots for a single converging and non-converging Markov chain. To test whether a single chain has converged the Geweke diagnostic can be used (Geweke 1992). The idea is that if the chain has converged (i.e. is stationary) the values at the beginning of the chain will have the same properties as those at the end. If the chain is stationary the means of samples taken from each tail of the of the chain are calculated from the same normal distributions. A z-score can then be used to interpret the difference in means, thus an absolute score <2 indicates convergence. For example the difference in means score between the first 10% and final 50% for the converging chain in Figure 2.8 (a) is 1.08. For the non-converging chain the absolute difference in means score is 5.8. The Geweke diagnostic can be calculated using the `coda` package in R (Plummer et al. 2006, R Core Team 2013).

When multiple chains are used to calculate a posterior the Gelman and Rubin convergence diagnostic can be used. This diagnostic test uses analysis of variance to test whether dispersion between chains is bigger than the dispersion within chains (Gelman & Rubin 1992). For computational efficiency, only single chains are used in this thesis, therefore this method is not detailed.



(a) Convergent Chain

(b) Non-convergent chain

Figure 2.8: Trace plots for converging and non-converging chains.

2.4.4 Model Selection

The Deviance Information Criteria (DIC) is a method to measure overall quality of a Bayesian model (Spiegelhalter et al. 2002), and is a generalization of the AIC (Akaike Information Criterion, see Section 6.7). DIC calculates the goodness of fit of a model penalised by the model's complexity.

DIC is defined as:

$$DIC = \overline{D(\theta)} + pD, \tag{2.9}$$

where $\overline{D(\theta)}$ is a measure of the goodness of fit and pD accounts for model complexity. Deviance ($D(\theta)$) is -2 times the log-likelihood:

$$D(\theta) = -2\mathsf{log}L\left(\theta\right). \tag{2.10}$$

Some authors include an additive constant in the definition of $D(\theta)$, however this is irrelevant for model comparison as the constants cancels out. Therefore we choose to omit it in the definition as is common in current literature.

$\overline{D(\theta)}$ is defined as the posterior mean of the deviance:

$$\overline{D(\theta)} = E_{\theta|y}\left[D(\theta)\right]. \tag{2.11}$$

pD adjusts for the effective number of parameters and is defined as:

$$pD = \overline{D(\theta)} - D(\hat{\theta}) = E_{\theta|y}\left[D(\theta)\right] - D(E_{\theta|y}[\theta]), \tag{2.12}$$

where $D(\hat{\theta})$ is the deviance evaluated at an estimate of θ, typically at the posterior mean (Gamerman & Lopes 2006, Lawson et al. 2003). The lower the value of DIC, the better the model.

2.5 Statistics in Ophthalmology

Ophthalmic data provides many challenges and points of interest from a statistical point of view. Many longitudinal datasets exist in both clinical practice and epidemiological studies. Often data is collected from both eyes and the correlation must be accounted for. Correct statistical modelling must be applied in order to both draw out the full potential of a dataset and to make accurate statistical inference.

Statistics is crucial when judging the effect of new treatments or diagnostic tools (Coleman 2009). The use of statistics in ophthalmology has been reviewed by Juzych (1992) and more recently Fan et al. (2011). Basic statistical techniques such as descriptive statistics, t-tests and contingency tables were most commonly used between 1970 and 1990 (Juzych 1992). With the complexity of ophthalmic data and efficiency of computers both increasing, the application of more advanced statistical methods has greatly increased since 1992. Methods such as ordinary and logistic regression, linear mixed models, survival analysis and models accounting for the correlation between eyes, are becoming more common (Fan et al. 2011).

Bayesian methods have been successfully applied to ophthalmic data sets with more accurate

results than traditional frequentist methods. Medeiros et al. (2012) compare ordinary least squares with a joint Bayesian regression model combining structural confocal scanning laser ophthalmoscopy measurements with functional visual field mean deviation. They found that the Bayesian model gave more accurate results for predicting mean deviation and rim area. Bayesian analyses combining structural and functional measurements have been carried out by others including Russell et al. (2012) and Mederios, Zangwill, Mansouri, Lisboa, Tafreshi & Weinreb (2012).

Bayesian meta-analyses have also been carried out in order to correctly account for variation at all levels. Rudnicka et al. (2006) used a Bayesian logistic meta-regression model with random effects to investigate the prevalence of glaucoma by ethnicity, gender and age. They found a higher prevalence of glaucoma in those of black ancestry, however the white populations showed a steeper increase of prevalence with age. Men were also more likely to have glaucoma. Rudnicka et al. (2012) carried out a similar Bayesian meta-analysis investigating the prevalence of age-related macular degeneration.

There are many examples of Bayesian methods being used to apply machine learning classifiers to ophthalmic data. Zhu et al. (2010) apply linear regression and a radial basis function neural network within a Bayesian framework to predict individual visual field sensitivities from retinal nerve fibre thickness. Zhu et al. (2011) apply this method to look at the agreement between structural and functional measurements. Other examples of Bayesian machine learning classifiers include Bowd et al. (2008, 2012) who combine structural and functional measurements to better predict glaucoma progression.

We provide a comprehensive review of specific statistical methods in ophthalmology within the relevant chapters of this thesis. We discuss statistical methods for analysing visual field data in Chapter 3. In Chapter 5 we review the use of splines in ophthalmic literature. The statistical modelling of video data is reviewed in Chapter 6.

2.6 Chapter Summary

In this chapter we have given a brief background of the eye and discussed the details of glaucoma, relevant to understanding and motivating the methods in this thesis. We have provided an overview of Bayesian statistics and its implementation through MCMC methods. More detail is provided throughout the thesis as specific techniques are introduced.

Chapter 3

Visual Field Analysis

3.1 Introduction

The visual field is the total area the eye can see while it is focused on a central point of fixation. As glaucomatous damage leads to vision loss, testing the visual field is a useful way to determine if a subject has glaucoma and if progression of the disease has occurred. The amount of functional loss in any eye can be quantified by a single visual field test. By analysing a series of visual fields, change in vision loss can be observed, and the rate of progression can be estimated. This information is crucial for making decisions about treating glaucomatous damage, and preventing further progression loss from occurring (Artes & Chauhan 2005).

Visual field analysis has been used as the sole end point to predict progression for many clinical trials, as well as being the main method used in clinical practice when making therapeutic decisions (Chauhan et al. 2008). This chapter describes standard automated perimetry, the main method for collecting data on visual field function. We then review the current methods for analysing visual field data. Following this we present our data set from the Vein Pulsation Study Trial in Glaucoma (VPSG) and data from the Lion's Eye Institute trial registry in Perth, Western Australia. We present some exploratory analysis and investigate modelling the large amounts of measurement error in visual field data.

3.2 Testing a Single Visual Field

The most common method of measuring visual field function is standard automated perimetry (SAP) (Artes & Chauhan 2005). SAP has been used as an endpoint in many recent clinical trials, and it is recommended that it be a standard method for analysing glaucoma progression in clinical practice (Chauhan et al. 2008). Since its introduction standard automated perimetry has replaced manual perimetry as the preferred method for assessing the visual field. SAP removes the variability of the examiner and provides standardized results across clinics and trials using this methodology (Walsh 2011).

Standard automated perimeters tests the visual field of one eye at a time. A subject is asked to focus on a central fixation point. They are then presented a series of lights, in a random order at various intensities across the visual field. When a subject sees a light they press a button. The machine then continues to test various loci on the eye at different intensities and records the lowest intensity which can be seen at each locus on the eye. The standard 24-2 test measures 54 (52 non-blind) loci across the visual field. Another test is the 30-2 test which measures 76 loci across the visual field. The 24-2 test is more routinely used as the measurements in the additional outer ring of the 30-2 field are more variable.

Standard automated perimeters measure the differential light sensitivity of the visual field. While the eye finds it hard to estimate an absolute value of light it is good at detecting contrast. Static perimetry is the most common, where a stimulus of known size, intensity, and location is projected against an illuminated white background, which is also of known intensity. As the stimulus intensity increases the probability of seeing the stimulus increases. Differential light sensitivity (DLS) is measured in decibels (dB), where a brighter light corresponds to a lower dB reading. Decibels are measured on the negative log-scale of brightness in Apostilbs. An increase in 10dB corresponds to a 10 fold reduction in Apostilbs (see Table 3.1). This means that high dB values are relatively compressed, while low values are elongated. This results in functional damage at higher sensitivities appearing smaller, while damage at lower

sensitivities will be emphasized (Reus & Lemij 2005). It is more important to detect changes at lower sensitivities because this corresponds to more advanced glaucoma.

Table 3.1: Relationship between apostilbs (asb) and decibels (dB).

Apostilbs (asb)	Decibels (dB)
10,000	0
1000	10
100	20
10	30
1	40

The Humphrey Field Analyser (Carl Zeiss Meditec Inc., Dublin, CA) is the most commonly used SAP in clinical practice. It uses a maximum light intensity of 10,000 apostilbs (asb) bulb which is equivalent to 0dB. The full threshold technique uses a double crossing method, in which the intensity of a stimulus at a certain location is adjusted until it is not seen. When this occurs the stimulus is increased by 4dB until it is identified again. The stimulus is then reduced by 2dB until it is not seen again, thus the visual field is crossed twice. A visual field threshold at a certain location is defined as the point at which a patient sees 50% of the stimuli presented to them. Therefore, the recorded visual field threshold is the dimmest light observed 50% of the time it is presented (see Figure 3.1). Using this strategy it is possible to get an accurate reading only presenting 5 stimuli per location. This strategy is known as the 4-2 strategy and starts by testing an initial 4 loci on the visual field at 25dB. The thresholds are measured twice, and these values are used as an initial for adjacent areas. An additional four loci are also tested twice and the 8 loci are used to measure short term fluctuation. If a locus differs substantially from the age-matched expected value it will also be tested twice (see Figure 3.2).

Stimuli are presented to locations randomly across the field, rather than sequentially, so the patient does not know where to expect the next stimulus, thus discouraging cheating. For participants with higher retinal sensitivity, the variation is lower, and thus the reproducibility is higher. Therefore participants with better vision tend to have steeper frequency of seeing curves. Subjects with lower sensitivity tend to have flatter curves indicating more variability

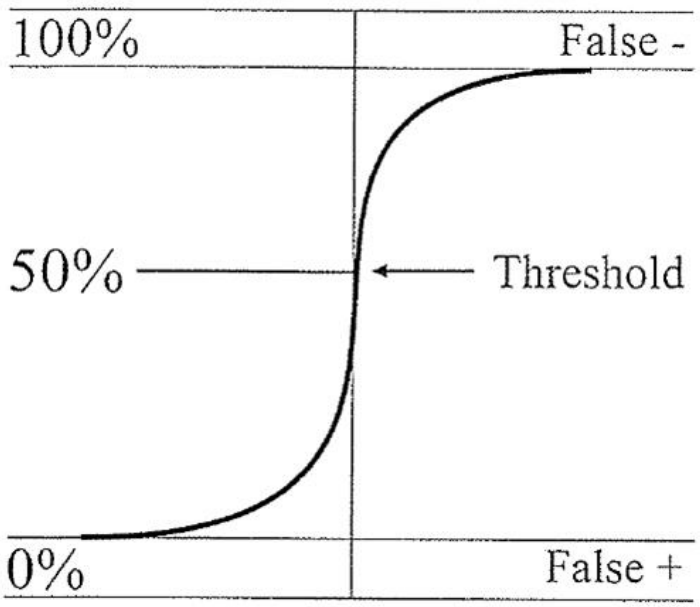


Figure 3.1: Frequency of seeing curve. As the intensity of the stimulus increases, the probability of seeing it also increases. Therefore the threshold is defined as the lowest intensity seen 50 % of the times presented. False + and false − refer to the rates of viewing stimuli. Reprinted with permission from BMJ Publishing Group Ltd. Licensee number:3235680130943. Walsh (2011), pg 93.

in whether they view the stimuli or not. (See Figure 3.3).

An alternative to the full threshold 4-2 strategy described above is the Swedish Interactive Threshold Algorithm (SITA). The main advantage of SITA over the full threshold techniques is that SITA is faster than full threshold techniques, resulting in shorter test times and less patient fatigue. This is achieved without reducing the quality of the data. SITA maintains similar variability properties, however threshold estimates are approximately 1 dB higher than when using the full threshold technique (Artes et al. 2002). A full threshold takes about 15 minutes per eye while SITA can be completed within 4 minutes. Therefore this algorithm has widespread clinical use.

3.2.1 Visual Field Output

As can be seen in Figure 3.2, visual field thresholds are outputted in a matrix format. This can also be likened to a map. Each locus tested on the visual field corresponds to a point on

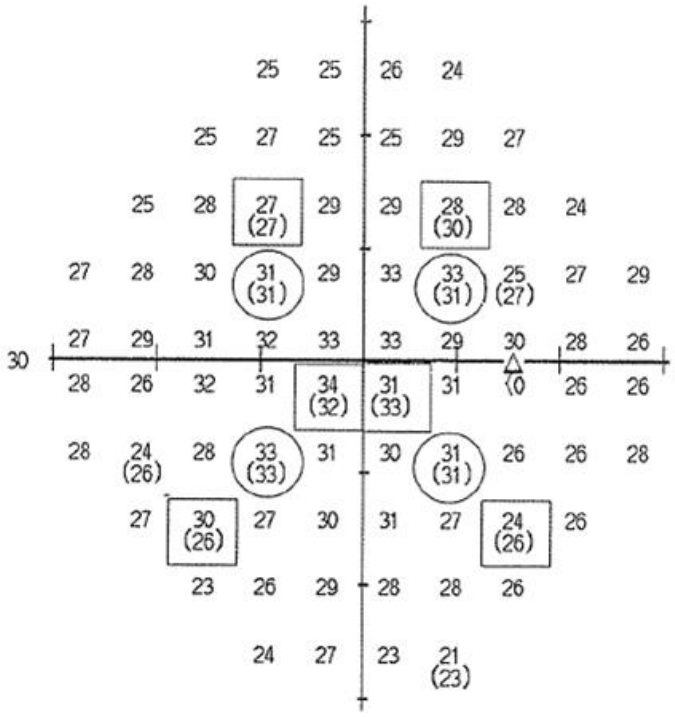


Figure 3.2: Initial loci (circled) and loci used to test STF (circled and squared). Loci retested due to a substantial difference from the age matched values are in brackets. Reprinted with permission from BMJ Publishing Group Ltd. Licence number:3235680130943. Walsh (2011), pg 96.

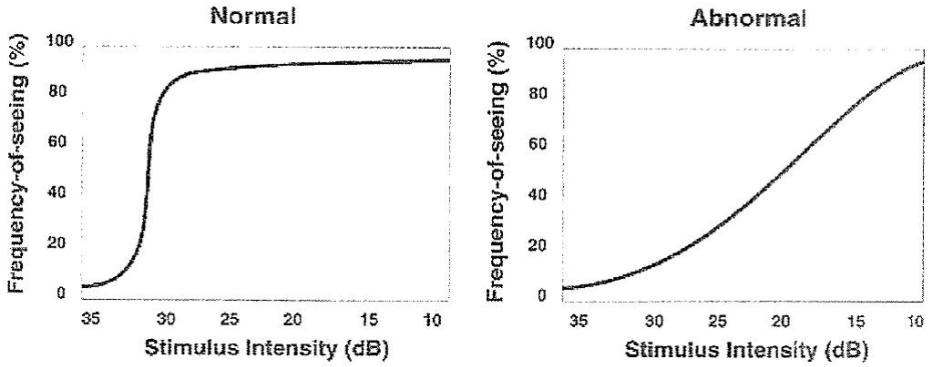


Figure 3.3: 'Normal central regions with high threshold intensity tend to have steep FOS curves. Abnormal or peripheral regions with reduced sensitivity demonstrate a broadened curve with greater threshold uncertainty.' Reprinted with permission from BMJ Publishing Group Ltd. Licence number:3235680130943. Walsh (2011), pg 93.

the optic disc where the ganglion cells receive light. Garway-Heath et al. (2000) divided the visual field into six sectors, based on the areas they correspond to on the optic disc. Figure 3.4 (left) represents the 52 loci tested in visual field testing. The optic disc sector each VF section maps to is shown in Figure 3.4 (right). However, the VF does not map directly to the optic disc. Specifically VF locations along the temporal horizontal mid-line do not map in adjoining locations on the optic disc. This is represented in Figure 3.4, where pairs of adjacent cells either side of the (horizontal) mid-line, specifically, at the temporal side of the eye, correspond to highly separated pathways across the retina and into the optic disc. (See dotted line in Figure 3.4 separating the superior-temporal and inferior-temporal sections on the visual field).

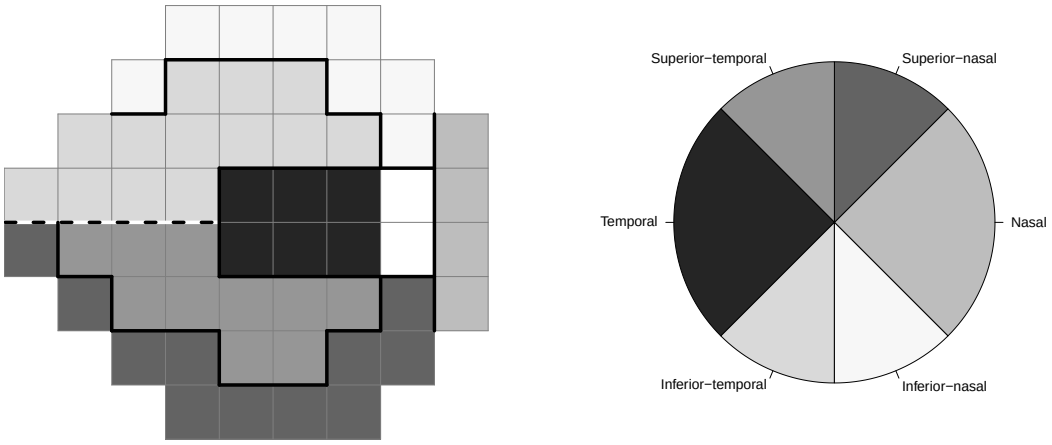


Figure 3.4: Visual field locations (a) mapped to the sectors of the optic disc (b) as by Garway-Heath et al. (2000). Dotted line shows non-adjacent cells, white cells show the 2 loci closet to the blind spot. Thick black lines separate the sectors in (a).

The patterns of visual loss seen in patients correspond 'to the anatomy of the nerve fibre layer of the retina and in its projections to the optic nerve' (Walsh 2011). The visual field test provides a link between the structural damage and functional VF loss. Damage to the nerves seen in Figure 3.5, will be observed as vision loss in the corresponding region of the visual field.

The functional loss observed through the visual field is observed as cell death known as cup-

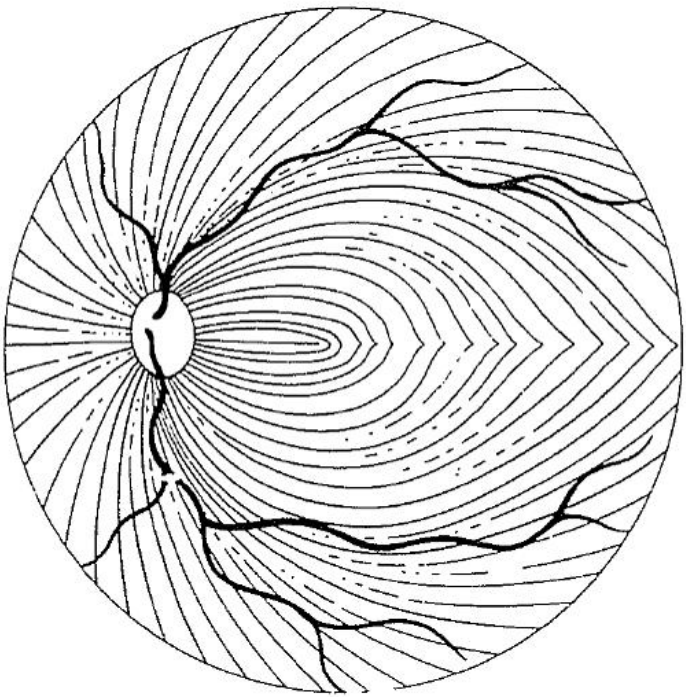


Figure 3.5: Layers of nerve fibres across the optic disc. Reprinted with permission from BMJ Publishing Group Ltd. License number:3235680130943. Walsh (2011), pg 128.

ping on the optic disc (see Chapter 2). A large deal of variation exists in the patterns of glaucomatous cupping. Some patients experience 'fairly diffuse, concentric loss of the neuroretinal rim, while others have extremely localized loss' (Walsh 2011). Thus glaucomatous visual field loss shows similar variation. Some subjects' visual fields show fairly localised defects, while others show scattered patterns of loss.

As well as outputting the threshold values, STATPAC, the built in software of the Humphrey Visual Field Analyser, also outputs measures such as total deviation, pattern deviation and the Glaucoma hemifield test. Total deviation compares the thresholds being analysed to normal age matched thresholds. Pattern deviation is similar to total deviation, but also adjusts for generalized depression which uniformly affects the whole field. This is useful because localized loss is generally more diagnostic. This is important as it can minimize the effects of cataracts on a VF test. P-values are calculated for both total and pattern deviations and indicate whether the defect observed is of clinical significance or due to chance. As total and

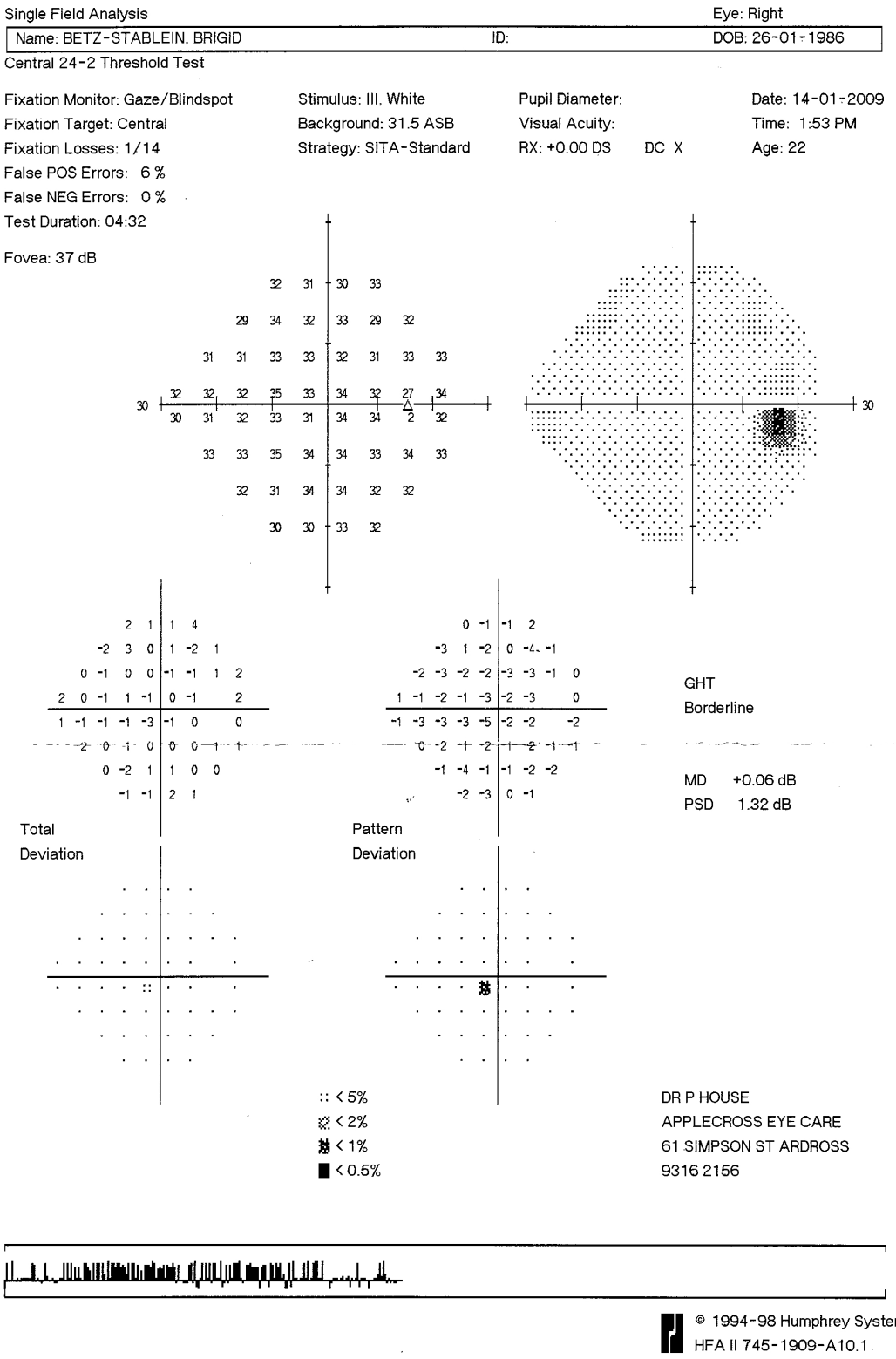


Figure 3.6: Visual field output for a normal eye.

pattern deviation are calculated based on point-wise differences to the age-matched controls, P-values take into account change in variation across the visual field.

Global indices of VF loss are also calculated via STATPAC. Mean deviation (MD) is calculated as the average across the eye of the point-wise differences between a test sensitivity and a normal age matched reference value at each loci. Therefore Mean Deviation does not pick up localized loss, but is a better indicator of generalized sensitivity loss. Chauhan et al. (2008) suggests that mean deviation be used to measure rates of change. Mean deviation generally differs between 0 and -2.5dB/yr in glaucoma subjects (Chauhan et al. 2008). The ability to detect this change depends on the variability of the visual fields over time. In a longitudinal study of glaucoma patients Artes & Chauhan (2005) defined low, moderate and high variability in MD as having standard deviations of 0.5, 1, and 2dB respectively. Pattern Standard Deviation provides a measure of the unevenness of the surface, i.e. how the shape of the field differs from a normal field. Short term fluctuation (STF) is the standard deviation of the 10 repeatedly tested locations. Corrected pattern standard deviation attempts to adjust for STF, but neither STF or corrected pattern standard deviation are currently used.

3.2.2 Diagnosing Glaucoma and monitoring progression with SAP

Glaucoma can be diagnosed based on the functional loss from one visual field, however other clinical factors such as IOP and family history are generally taken into consideration. Tables 3.2 and 3.3 show two suggested sets of criteria for diagnosing glaucoma.

3.3 Variation in the Visual Field

Visual field data contains a notable amount of variation. In order to detect statistically significant progression in glaucoma, the true change in a visual field must be larger than the noise present in the data (Chauhan & Johnson 1999, Heijl et al. 1989). One of the greatest challenges remains to differentiate between random fluctuation and true progression (Walsh

Table 3.2: Minimal Criteria for Grading Abnormality (Central 30°). Reproduced from Caprioli (1991).

Strict
$\geq$ 4 adjacent points $\geq$ 5 dB loss* each
$\geq$ 3 adjacent points $\geq$ 10 dB loss* each
Difference of $\geq$ 10 dB across nasal horizontal meridian at $\geq$ 3 adjacent points
Exclusions: physiologic blind spot; superior and inferior rows
Moderate
$\geq$ 3 adjacent points $\geq$ 5 dB loss* each
$\geq$ 2 adjacent points $\geq$ 10 dB loss* each
Difference of $\geq$ 10 dB across nasal horizontal meridian at $\geq$ 2 adjacent points
Exclusions: physiologic blind spot; superior and inferior rows
Liberal
$\geq$ 2 adjacent points $\geq$ 5 dB loss* each
$\geq$ 1 adjacent points $\geq$ 10 dB loss* each
Difference of $\geq$ 5 dB across nasal horizontal meridian at $\geq$ 2 adjacent points
Exclusions: physiologic blind spot; superior and inferior rows

**Loss is relative to normal values or to values of surrounding points. For probability maps that compare measured thresholds to normal values substitute P <0.05 for 5 dB loss and P <0.01 for 10 dB loss*

Table 3.3: Minimal Criteria for Diagnosing Acquired Glaucomatous Damage. Reproduced from Walsh (2011).

A glaucoma hemifield test (GHT) outside normal limits on at least 2 fields
OR
A cluster of 3 or more non-edge points in a location typical for glaucoma, all of which are depressed on the pattern deviation plot at a
P <5% level and 1 of which is depressed at a
P <1% level on 2 consecutive fields
OR
A corrected pattern standard deviation that occurs in less than 5% of normal fields on 2 consecutive fields

2011), due to the large amounts of variation in visual field data. In the Ocular Hypertension study, Keltner et al. (2000) showed that 86% of 703 retests did not pick up on defects found in earlier examinations. This section reviews what is known about the variation in visual field data, particularly with relation to ocular hypertension and glaucoma patients.

Additional effects that should be considered include patient fatigue. This can lead to an over-estimation of progression or initial glaucomatous damage. Other eye conditions can effect the VF responses such as miosis and mydriasis, media opacities, eyelid and nasal effects (ptosis). Refractive errors and corrective lens/frame artefacts can also affect the results.

3.3.1 Types of VF variation

Visual field variation can be divided into two categories, short-term fluctuation (STF) and long-term fluctuation (LTF). Short term fluctuation is variation within a single test, and due to instability of the threshold being tested as well as the attentiveness of the patient. STF is calculated by measuring 10 locations in the VF twice within the same test; however this is not measured with the SITA algorithm. STF can be caused by an inconsistent patient or by true VF loss. Long term fluctuation is fluctuation between tests. The causes of LTF are not really known. Hypotheses include fluctuations in IOP, and the age of the patient. LTF in stable patients is also correlated with initial sensitivity and distance from the fixation point (Walsh 2011).

3.3.2 Causes of variation

The earliest investigation into the variation of visual field analysis was carried out by Heijl et al. (1987). Heijl et al. (1987) evaluated 74 subjects with normal eyes to assess variability of SAP across the visual field. Table 3.4 shows the average inter-test variation from the 74 subjects, calculated from 3 VFs taken over a period of 4 months. They found that variability of the thresholds for subjects without glaucoma is dependent on eccentricity. That is variation

in visual fields is dependent on the distance from the fixation locus to the locus in the VF being measured. Analysis of a 30-2 field showed that for intra-test variability (STF) the variation in 4 peripheral loci was 27% higher than 6 of the central loci. Inter-test variation (LTF) was also dependent on eccentricity. Heijl et al. (1987) also showed that within a test, variation was dependent on location, however if a patient had higher variation in one locus they were likely to have higher variation in another locus.

Table 3.4: Inter-test variation of normal eyes by location (Heijl et al. 1987).

Inter-test Variation of normal eyes (Heijl et al. 1987)								
			4.70	4.20	3.70	4.30		
		3.40	3.10	2.80	3.00	2.90	3.00	
	3.40	3.30	2.40	3.30	2.10	2.20	3.50	3.40
4.20	2.70	2.30	2.30	2.00	2.20	2.40		3.70
5.80	2.20	2.30	1.90	1.60	1.80	2.10		3.30
	3.00	2.40	1.80	2.40	2.20	2.40	3.90	3.50
		2.70	3.40	2.40	2.40	2.30	2.50	
			3.70	2.40	2.40	3.00		

It was also found that the distribution of variation was not normal, but skewed. The distributions of variation became less Gaussian the further away from the fixation point. In a 24-2 field the outer circle of the 30-2 visual field is not tested, due to the additional variation. Physiologically this dependency could be explained by the decrease in neural channel density towards the eccentricities of the eye, thus decreased signal-to-noise ratio, and hence more variation. Another hypothesis is that variability of the visual field is related to the density of functional ganglion cells (Henson et al. 2000). Variation is also seen to increase in visual fields with more damage (Chauhan & Johnson 1999). Therefore a lower initial sensitivity is associated with greater variation (Spry et al. 2003) (see Figure 3.7 from Walsh, 2011). As Heijl et al. (1987) pointed out, if greater variation is seen in one locus in the eye, greater variation will be seen across the eye. It follows that variation in glaucoma patients may not be as location dependent as patients with normal VFs.

Chauhan & Johnson (1999) also note that due to the finite light sensitivities being tested the

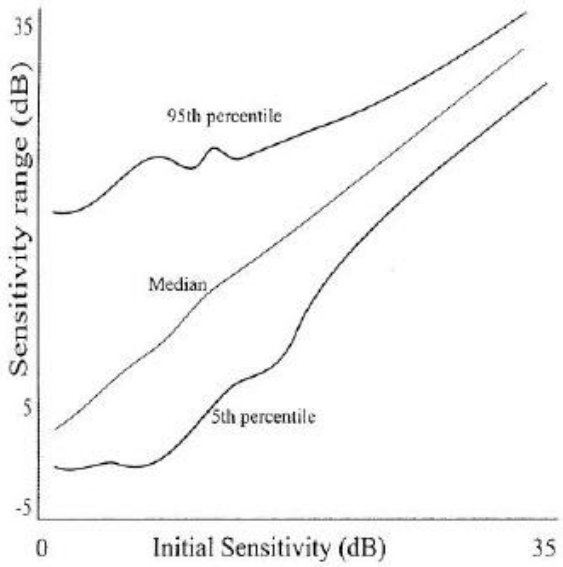


Figure 3.7: 95% Confidence intervals of subsequent threshold sensitivity against initial threshold sensitivities for single test locations. Reprinted with permission from BMJ Publishing Group Ltd. Liscenece number:3235680130943. Walsh (2011), pg 94.

measured variability in locations with advanced damage is artificially truncated. This is also extenuated as once a patient is truly blind in a locus on the visual field, bar fixation error, there should be no variation in results.

While SAP reduces the variability of the examiner (compared with manual perimetry), the accuracy still depends on the patients' responses, which can be affected by 'optic, neural and psychological factors' (Walsh 2011). It has been shown that instructions can have a significant effect on the variation in visual fields (Kutzko et al. 2000). Learning effects and patient fatigue can also have an effect on the amount of variation seen in a visual field (Wild et al. 1991). However many of these effects can be controlled. Establishing an accurate baseline field helps reduce the variation due to the learning effect. Patients often perform poorly on their first attempt, but improve over the next 3-4 fields. Therefore patients generally repeat the test several times in the weeks or months following initial test, as the learning effect will eventually plateau. It is also important to confirm a newly recognized defect with a second

test.

SAP has some built in techniques to measure field test variability. In order for a VF to be considered reliable they should have a fixation loss of $< 20\%$ and a false positive rate $< 30\%$ (Walsh 2011). Clear instructions from the perimeterist can reduce the number of false positives as they are generally due to subject error rather than physiological factors. Interestingly false positive rates did not differ between glaucomatous and normal eyes (Bengtsson & Heijl 2000). This is seen in previous research where it has been shown that variation increases in eyes with more visual field loss. It is also noted that glaucomatous eyes have more variation than normal eyes, as well as higher visual fatigue. Fixation is also monitored throughout the test, by illuminating stimuli in the blind-spot. If a patient responds it is assumed they have poor fixation.

Originally false negatives were included in visual field output to indicate the reliability of a patient, however Bengtsson & Heijl (2000) suggest that this is not the case for glaucoma patients. Instead of being an indicator of reliability, false negatives increased with loss of visual function of the patient, therefore more false negatives occurred in eyes with higher field loss.

3.3.3 Reducing variation by changing standard SAP

Wall et al. (2009) tested the hypothesis that larger stimuli would result in a more uniform variation across the visual field. They analysed test re-test variability using 4 techniques, Standard Automated Perimetry size III using SITA, SAP size 4 with full threshold (SAP V), matrix (FDT II) and motion perimetry. As noted previously, Wall et al. (2009) found that variability increased as sensitivity decreased for SAP III ($r^2 = 0.22$), however this increase was not as profound and in some cases almost absent with the other methods ($r^2 = 0.12, 0.02, 0.02$ respectively). This lower variability is likely due to the larger stimulus size, and increased sampling used by the other three methods.

How much variation reduction is necessary to adopt new VF testing method?

Turpin & McKendrick (2011) looked at how much the variability in visual field testing would need to be reduced in order predict progression earlier than the current SAP methods. Progression was assessed using a Point-wise Linear Regression (PLR) model on simulated datasets, in which the amount of variation could be controlled. Depending on the rate of progression, and the frequency of visual fields, the variability would need to be reduced by 30%-60% to detect progression one year earlier. The mean variance reduction required to make a clinical difference in detecting progression was 20%, which would result in detecting progression approximately one VF earlier. This means that in order for a method to replace SAP it would have to result in a decrease in variability of at least 20% from the current methods.

Medeiros et al. (2011) notes that in clinical practice visual field analysis is only performed on those who are suspected of having glaucoma and that healthy eyes have different characteristics and variability in visual fields than diseased eyes. Most often research focuses on reducing the noise of visual fields and taking more accurate measurements rather than improving the analytical methods (Artes et al. 2002, Wild et al. 1991, Gardiner et al. 2004). The following section presents a review on analytical methods that have been developed to improve the detection of progression from visual field examinations.

3.4 Serial Visual Fields and Testing Progression

In order to diagnose glaucoma progression, series of visual fields must be examined. It is important to consider the VF guidelines with all the other clinical information used to diagnose glaucoma to gain the entire clinical picture. Other important variables include the patient's age, IOP control, nerve fibre layer status, optic disk appearance, and the patient's assessment of vision. Guidelines for recognizing glaucoma progression are summarized in Tables 3.5 and 3.6.

Table 3.5: Guidelines for Recognizing Progression. Reproduced from Anderson (1992).

New defect in previously normal region: cluster of 3 points worsening by 5 dB each, 1 of which has worsened by 10 dB
Previously abnormal region has deepened if: 3 or more points have deteriorated by 10 dB each
Previously abnormal region has widened if: 2 or more new or contiguous points are involved

Note: This scheme requires reliable fields: In general, progression should be confirmed on subsequent field or fields. Adapted with permission from Anderson, DR:

Table 3.6: Guidelines for Evaluating Change From One Test to Another in Single Test Location and in Adjacent Pair of Test Locations. Reproduced from Zulauf & Caprioli (1992).

	Consider as Progression if Change (dB) Exceeds	
Initial Sensitivity (dB)	*In Single Test Location*	*In Pair of Test Locations*
25-30	8	3.5
20	12	5
15	15*	6.5
10	10†	8

* *Values may fluctuate between 0 and normal*

† *A single test location drop to 9db may be due to fluctuation*

Chauhan et al. (2008) concludes that detecting the progression of glaucomatous eyes depends on the frequency with which the visual fields are measured, and upon the variability of these fields. Also affecting this is the magnitude of change that is considered to be clinically significant. While it is reasonable to use the criteria in Tables 3.5 and 3.6 when dealing with a few fields, this becomes difficult as the patient accumulates more fields. Additionally it is important to quantify the amount of measurement error in a given test. A review of the current methods of analysis to formally diagnose progression follows.

3.5 Current Analysis of Visual Field Data

As mentioned in Chapter 2, one of the most important challenges in managing glaucoma is detecting whether progression has occurred. This is routinely carried out by examining a patients' visual field, however there is currently no consensus on the gold standard for analysing visual field data. Ernest et al. (2012*a*) reviewed literature prior to April 2009 and found that

47% of studies using quantitative methods calculated rates of progression, while the remainder dichotomized the outcome. Artes & Chauhan (2005) mention 3 types of methods for determining progression; subjective analysis, event-based analysis and trend-based analysis. Subjective analysis generally involves binary classification into categories such as glaucomatous or not, progressing or not. There does not seem to be a standardized version of these methods, and there is poor agreement between methods. Event-based analysis compares the baseline visual field to the current visual field and compares the differences to test-retest variability of a independent group of subjects. Trend-based analysis uses linear regression. A problem with all these methods is that they use arbitrarily chosen cut-offs to define progression. An advantage of event-based analysis is that less tests are required compared to trend-based analysis. Because rate of progression is a continuous variable, a binary variable will not take advantage of/provide all relevant information.

Progression methods use threshold sensitivities, as well as other SAP outputs such as pattern deviation and mean deviation. Our focus is on methods inputting threshold sensitivities, however for a more complete review we mention some of the more common and novel techniques using other SAP measures. The remainder of this section reviews current analysis methods for visual field data. We discuss the more traditional approaches including scoring, glaucoma probability analysis and pointwise linear regression. We also discuss some more novel techniques including neural networks (machine learning classifiers), and spatial analyses.

3.5.1 Event based analysis

The Advanced Glaucoma Intervention Study (AGIS) and Collaborative Initial Glaucoma Treatment Study (CIGTS) are two methods used previously in clinical trials that are based on scoring systems. For the AGIS method, a score between 0-20 was calculated and progression is defined as an increase in score by four in three or more consecutive fields (Vesti et al. 2003). This score is based on the number of adjacent depressed VF test loci on the total deviation plot in the nasal, upper and lower hemifields (Sample et al. 2005). The CIGTS method bases their score on the probability levels of the total deviations. Again a final score

between 0 and 20 was calculated, however under CIGTS progression is defined by a score increase of three in three or more consecutive fields.

Glaucoma progression analysis (GPA) is a software provided on the Humphrey visual field analyser It is an updated version of glaucoma change probability analysis or STATPAC2. GPA uses two baseline examines and compares these to the current exam. Progression is defined by the Early Manifest Glaucoma Trial (EMGT) using GPA. GPA uses pattern deviation plot values, rather than total deviation plots used by STATPAC2. EMGT defines statistically significant progression when deterioration is seen in the same location in 3 or more consecutive tests (Sample et al. 2005). Heijl et al. (2008) showed that EMGT identified progression earlier than AGIS and CIGTS.

3.5.2 Point-Wise Linear Regression

There are many point-wise linear regression (PLR) criteria used to evaluate the 52 non-blind loci on the eye. A commonly used software is PROGRESSOR which uses 52 'independent' PLR across, one per non-blind locus, sequential visual field thresholds (Fitzke et al. 1996). PROGRESSOR uses progression criteria of 1dB per year which is approx 10 times greater than normal age related decline, and 2dB per year for edge points as these are assumed to be more variable. PROGRESSOR defines edge points as anything 15 degrees or more from the fixation point. A statistically significant slope is a slope that is different from zero, with $p < 0.1$. Two to three loci must meet these slope criteria each with a significant p-value for an eye to be classified as progressing. Many other progression criteria have been investigated and are outlined in Table 3.7. Their relative diagnosing power is discussed in Section 3.5.6.

De Moraes et al. (2011) used a simple modification to standard PLR. They hypothesized that if a large number of tests is used, outliers would affect the ability to predict progression by increasing the variability in the test. Instead of using all available visual fields they selected the first 3,4 or 5, and the most recent 3,4 or 5 fields (6,8 or 10 fields total). Progression was

Table 3.7: Summary of PLR methods progression criteria where p is the p-value associated with the slope and locus refers to each tested loci (i) on the visual field.

Method	Progression criteria
$P1$	p ≤ 0.001 for ≥ 1 point (Smith et al. 1996)
$P2$	p ≤ 0.001 for ≥ 2 points (Smith et al. 1996)
$P3$	p ≤ 0.001 for ≥ 3 points (Smith et al. 1996)
$P4$	p ≤ 0.001 for ≥ 4 points (Smith et al. 1996)
$V1$	p ≤ 0.01 and slope $\leq -1dB$ per year for ≥ 2 points over 3 consecutive tests (Vesti et al. 2003)
$V2$	p ≤ 0.01 and slope $\leq -1dB$ per year for ≥ 3 points over 3 consecutive tests (Vesti et al. 2003)
$F1$	p ≤ 0.1 and slope $\leq -1dB$ for inner points, $\leq -2dB$ for points beyond 15° for ≥ 2 points (Fitzke et al. 1996)
$F2$	p ≤ 0.1 and slope $\leq -1dB$ for inner points, $\leq -2dB$ for points beyond 15° for ≥ 3 points (Fitzke et al. 1996)
VS	p ≤ 0.05 and slope $\leq -1dB$ per year for ≥ 1 non-edge† point, $\leq -2dB$ for edge points, for ≥ 1 point(Viswanathan et al. 1997)
$B1$	p ≤ 0.05 and slope $\leq -1dB$ per year for ≥ 2 non-edge† points (Birch et al. 1995)

†Edge points defined for a 30-2 so this refers to only 2 points on the 24-2 field

defined as a slope > -1.0dB per year ($P < 0.01$), and due to increased variability in the outer regions, edge points required a slope > -2.0dB per year. While this method showed similar sensitivity to traditional PLR, it showed higher specificity. Excluding visual fields means those which are included are given more weight. While this may falsely reduce the variation in some instances, it will also give more weight to outliers in the fields remaining in the analysis.

3.5.3 Machine Learning Classifiers

Both supervised and unsupervised data mining techniques have been used to identify progression or classify VF as being glaucomatous or normal. Jampel et al. (2011) reviewed several studies using machine learning classifiers and concluded that their diagnosing ability was as good, but not better than other assessment techniques. Some machine learning classifier methods are described below.

Sample et al. (2002) compared five types of machine learning classifiers with a more traditional approach from the software STATPAC2. The analysis was carried out on 114 eyes, with a minimum of 4 visual fields per eye. The machine learning classifiers used had been previously trained in a study by the same authors (Goldbaum et al. 2002). Each eye was analysed at each time point and classified by each method as glaucomatous or normal. No gold standard substitute was used in this study, so the results of the various methods were compared against each other. The traditional method saw thirty-six eyes switching classification during the period. The same eyes were identified by at least one machine learning classifier as becoming glaucomatous, and this classification preceded that of the STATPAC approach.

Boden et al. (2007) investigated the use of quadratic discriminant analysis and support vector machines with Gaussian kernels on 4 different VF clustering schemes. Loci on the visual field were clustered by sector (Garway-Heath et al. 2000), by glaucoma hemifield, by statistical clustering (Goldbaum et al. 2005) or randomized into clusters. Boden et al. (2007) found that while clustering the VF loci may help with data reduction, it did not improve the performance of either machine learning classifiers.

Goldbaum et al. (2005) apply the unsupervised data mining technique of clustering to a data set consisting of 189 normal eyes and 156 with glaucomatous optic neuropathy (glaucomatous damage to the optic nerve). The aim was to see if this technique could provide meaningful clusters and distinguish between patterns of glaucomatous visual field damage. They expanded on a basic clustering model by using a variational Bayesian independent component analysis mixture model. This model divides the SAPs into the optimal number of clusters, and then further separates the fields by finding the optimal number of maximally independent axes. Goldbaum et al. (2005) found that the model divided the data into two clusters, placing 68.6% of the glaucomatous eyes in one cluster (G), and 98.4 % of the normal eyes in the second cluster (N). Cluster G, was optimally described by six independent axes, while cluster N was adequately represented by one axes. They found that the axis patterns created by their

model were consistent with VF patterns identified by experts. In summary the variation Bayesian independent component analysis mixture model is able to create clusters, largely differentiating between normal and glaucomatous eyes and within these clusters identify and grade patterns of VF loss.

The machine learning classifiers trained above by Goldbaum et al. (2005) were then used by the same authors to identify progression in a longitudinal series of visual fields from 191 eyes (Sample et al. 2005). Sample et al. (2005) found that their method identified a higher proportion of eyes progressing compared to the AGIS or EMGT criteria. Goldbaum et al. (2009) applied the same variational Bayesian independent component analysis mixture model to a much larger database of SITA VFs, consisting of 1146 normal eyes and 939 glaucomatous eyes. The model identified three clusters within the dataset. The first contained a majority of normal fields, the second consisted of mildly abnormal fields, and the third mostly moderate to severely abnormal fields. The clusters were divided into 2, 2 and 5 axes. As in the previous study a post-hoc analysis of the axes showed different patterns of visual field loss, and the distance along from the centroid of the cluster represented the severity of the VF loss pattern.

3.5.4 Spatial Filters

To attempt to reduce variability in visual field analysis, spatial filters have been applied to SAP threshold data. As mentioned earlier the visual field data can be viewed as numerical matrices, where it is expected that loci neighbouring each other are closer in value than loci that do not share boundaries (see Figure 3.4). Therefore information from the neighbouring regions can be used to weight the analysis, in an attempt to reduce noise due to measurement error. Fitzke et al. (1995) introduce this idea by applying a Gaussian filter to visual field data. A 3 by 3 grid is applied to a central point, taking the weighted average of the points within the grid. This was repeated for each point resulting in spatial smoothing. Point-wise linear regression is then used on the smoothed values. The authors concluded that test-re-test variability was greatly reduced (Fitzke et al. 1995). However Spry et al. (2002) showed that this method also reduces the sensitivity in detecting true change, especially in areas where

the progression was seen in clusters of loci ≤ 9, the size of the grid, where the method did not perform better than point-wise linear regression on the raw data.

Another problem with this analysis is that Fitzke et al. (1995) did not take into account the physiology of the eye. This anatomic relationship between the visual field matrix and the optic disk is discussed earlier in Section 3.2.1 and can be seen in Figure 3.4 (Garway-Heath et al. 2000). Gardiner et al. (2004) developed a spatial filter to account for the physiological differences, while still taking advantage of the neighbouring information of the points situated beside each other on both the optic disk, and the visual field matrix. The spatial filter was derived from a database consisting of 98,821 visual fields.

Initially covariances were calculated between all points on the field. If we assume that the sensitivity,(S_A), at point A can be predicted by a linear combination of all other points, then the covariance between point A and point B is $\mathsf{Cov}\,(S_A, S_B) = \sum_i k_i \mathsf{Cov}\,(S_i, S_B)$ (Gardiner et al. 2004). Regressions were performed on each point producing the constants k_i. Since k_i is known the sensitivity at each point can be calculated, given the sensitivities at the other points. The equation for the filtered point, S_A^f, is a combination of the predicted sensitivity $(\hat{S}_A)$ and the raw sensitivity (S_A):

$$S_A^f = c\hat{S}_A + (1-c)S_A, \tag{3.1}$$

where $\hat{S_A} = \sum k_i S_i$. If the prediction was completely accurate then the minimum of a function occurs when the deviation of the variance equals zero. Therefore it can be shown that the correlation $\hat{c} = \left(\sum_i k_i^2 + 1\right)^{-1}$. However given that it is known that the predictions are not completely accurate, and that some points vary more than others, the correlation between the predicted and raw sensitivities were taken into account so that c is multiplied by the correlation between the predicted and raw values of A, thus $c = \hat{c} \times \mathsf{Corr}_A$. When a prediction is closer the raw value, c will be smaller. As S_A^f is calculated for each point on the eye, the filter takes the form of a matrix of values varying for each point, but correlated based on their

position in the eye. More details including the restrictions placed on the $k_i's$ can be found in Gardiner et al. (2004).

The authors applied this filter to a simulated data set and found that the specificity and sensitivity improved using the spatially filtered data. It should be noted however that the simulated data was limited in that it was only simulated over a period of 5 years, with a 2 dB loss per year which means that the noise free values only varied between 20-30 dB. Gardiner et al. (2004) note that the spatial filter is designed for glaucoma suspects (patients with ocular hypertension). Therefore how the filter behaves with eyes with glaucoma or eyes with a larger with-in eye variation is untested. Strouthidis et al. (2007) apply the novel spatial filter created by Gardiner et al. (2004) to longitudinal data. They conclude that it shows similar specificity to point-wise linear regression (PLR), but an increased rate of detected progression. The visual fields Strouthidis et al. (2007) tested were from ocular hypertensive subjects with normal visual fields at baseline. So again it is uncertain as to how the filter would behave with more advanced subjects.

3.5.5 Methods combining SAP thresholds and Structural data

Garway-Heath et al. (2002) looked at the relationship between the number of ganglion cells in the eye and the visual field sensitivity. As previously mentioned, visual field sensitivity is measured on the log scale in decibels. The decibel can also be expressed as $10 \times \log(1/\text{Lambert})$, where Lambert measures the test spot intensity. Linear and quadratic regressions were used to investigate the relationships between differential light sensitivity (DLS) (as decibels and 1/Lambert) and both pattern electroretinogram (related to the number of functioning ganglion cells), and neuroretinal rim area. Curvilinear relationships were found when variables were regressed with DLS, whilst linear relationships were found when pattern electroretinogram and neuroretinal rim area were regressed against DLS as 1/Lambert. In the past it had been supposed that there was a set value of ganglion cells, or a functional reserve, at which one starts to lose visual field functionality. Largely this is due to structural damage being identified before functional damage. This is due to the nature of the logarithmic scale, as well

as the high test-retest variability in visual field testing. This work, however, supports the hypothesis of a continuous linear structure-function relationship between number of ganglion cells and 1/Lambert DLS. If a linear structure-function relationship with the log scale DLS is assumed, underestimation of early functional loss and over estimation of more severe functional loss will result, as the true relationship is curvilinear. Reus & Lemij (2005) also looked at the relationship between DLS and structural measurements. They concluded that due to the curvilinear relationship between DLS and the structural measurements, the structural tests were able to identify glaucomatous loss earlier than SAP. Because dB are measured on the logarithmic scale, a clinically relevant structural change measured on the linear scale may be associated with only small changes in DLS. Reus & Lemij (2005) conclude that confocal scanning laser ophthalmology and scanning laser polarimetry might be better at estimating early glaucomatous damage.

More recent studies have combined SAP data with structural datasets within a machine learning environment. Bowd et al. (2008) found that combining optical coherence tomography data with SAP did not significantly improve the performance of the model compared with the structural and functional models on their own. Bowd et al. (2012) found that relevance support vector machine classifiers, combining SAP total deviation and confocal laser scanning ophthalmoscope images, more accurately identified progression than either the SAP or onfocal laser scanning ophthalmoscope global indices. Li et al. (2013) combine a temporal bootstrap model with Hidden Markov Models in order to create a pseudo time-series model from cross-sectional data. Average values from the VF sectors as defined by Garway-Heath et al. (2000) from 160 eyes were combined with Heidelberg Retinal Tomography data. The authors found that their model was able to separate stable states showing rim narrowing and abnormal VF sensitivities, and transitory states with fields showing subtle rim narrowing and VF sensitivity loss. This agreed with the current understanding of progression where rim loss may precede VF sensitivity loss and vice versa.

Bayesian Analysis

Medeiros et al. (2011) used Bayesian hierarchical models to combine structural and functional data in order to better detect the progression of glaucoma. It is suggested that an improved assessment of progression may be achieved by combining the structural measurements of the retinal nerve fibre layer and the functional measurements of the SAP visual field index. This model requires an additional dataset containing retinal nerve fibre layer measurements which are obtained using scanning laser polarimetry.

A joint multivariate mixed effects model was used within a Bayesian framework in order to combine these longitudinal measures. Random effects allowed for both within and between patient variation. Eyes were nested within patient in order to account for the correlation between a subject's two eyes. A multivariate skew t-distribution was used as a prior for the random effects as non-normality was suspected. MCMC sampling was carried out in WinBUGS and a specific eye was judged as progressing if the upper bound of the 95% credible interval was less than 0. These results were compared to the more standard approach of ordinary least squared linear regression, which was carried out individually on each eye.

Medeiros et al. (2011) ran the model on 434 eyes from 257 participants. At the baseline visit 38% were glaucomatous, 55% were glaucoma suspects and 7% were normal eyes. Compared with linear regression, the Bayesian method identified a significantly higher proportion of progressing eyes. Of the 405 glaucomatous and glaucoma suspect eyes the Bayesian method identified 22.7% as progressing compared with 12.8% by the regression method. When using optic disk stereographs as a reference endpoint for progression, the Bayesian method again outperformed the regression method identifying 74% as progressing compared with 37%. Of note is that the Bayesian method was able to detect eyes with faster rates of change than the regression method which struggled to classify with slopes with larger standard errors. From a clinical point of view these cases are more important to identify as they are likely to require intervention. However combining structural and functional information requires the collection of two datasets. In addition, because the model uses visual field index rather than

the individual threshold values the model is limited. Information on spatial correlation cannot be incorporated into the model, and while the model can give an overall eye progression rate, it cannot provide rates by sector or hemifield.

3.5.6 Comparison of analysis methods

The lack of gold standard for analysing visual fields often leads to investigators developing their own methods to suit their particular study (Ernest et al. 2012*b*). This makes comparing studies difficult or impossible and can prevent distinguishing whether an observed improvement is due to a new treatment or a different analytic method.

Viswanathan et al. (1997) compared STATPAC2 and PROGRESSOR on a data set containing 19 eyes and concluded that PROGRESSOR preformed better than STATPAC2. Vesti et al. (2003) compared several techniques including two scoring methods, a method based on Glaucoma Change Probability and, methods based on point-wise linear regression. They concluded that AGIS and CIGTS, had high specificity, but did not predict as many cases of progression as the other methods. While the Glaucoma Change Probability methods detected progression earliest they were not as specific. The pointwise-linear regression method used by Vesti et al. (2003) was specific, but took the longest to confirm progression. Vesti et al. (2003) states that

> The ideal method for analysing visual field change should be sensitive, detect progression with few examinations, maintain high specificity, and be resistant to fluctuation.

Unfortunately none of the methods they looked at achieved all the attributes listed above. Thus, they concluded that further research be undertaken to develop and improve methods of visual field analysis.

Meta Analysis of VF methods

Ernest et al. (2012*b*) used a network meta-analysis to directly and indirectly compare 30 different visual field progression detection methods. Direct comparisons were made when methods were used on the same population, and indirect comparisons were computed using a common reference. Articles were chosen systematically based on criteria reported in Ernest et al. (2012*b*). The meta-analysis was carried out on data from 12 articles containing a total of 30 methods. Methods were grouped into six subsections.

The first contained methods based on Glaucoma Progression Analysis (GPA). The second group consisted of two methods based on the Advanced Glaucoma Intervention Study (AGIS). The third group contains methods based on point-wise linear regression (PLR), including PROGRESSOR. Group four also is based on linear regression, but did not utilize loci threshold values, instead using indices such as mean defect or the visual field index. Another group included two methods which combined PLR methods with linear regression of the mean defect and utilized spatial filters. In the final group progression was defined by clinicians observing the visual fields, but blinded to any other clinical information.

The methods were ranked based on progression rate. The mean estimate of progression incidence over 6 years was 21%, ranging from 0.06% to 62%. The large difference highlights the importance of method chosen for analysing progression, and also the importance of taking into account methodology when comparing studies. A longer follow up period and worse baseline visual field also contributed to progression incidence. The incidence in progression in this study is lower than some other studies, this is most likely due to the other studies requiring a minimum of 3 VFs each year while, the average in this study was 1.7 VFs per year. Thus these results are more likely to resemble those seen in current clinical practice. Methods that predicted the highest rates of progression incidence included PROGRESSOR, the TNT program and one of the clinical methods. Those with lower estimates of progression were the majority of the index methods and the PLR with tighter criteria as to what was considered significant.

While this study shows the variability in the prediction of incidence of progression, the question of 'true' progression still remains. The question of whether it is better for a method to be more sensitive or more specific is raised. Ernest et al. (2012*b*) suggests that a highly specific method is better for epidemiological studies, and also for clinical practice when treatments have considerable side-effects. However because there is no gold standard, it is hard to quantify how sensitive or specific a method is. The authors also note that PLR regression methods not only have the capacity to classify someone as progressing or not, but the slope also provides an estimate of rate of progression. Ernest et al. (2012*b*) concludes that 'An ideal method combines statistically and clinically relevant criteria with tailor-made cut-off values'.

3.6 Data

In this section we introduce our dataset. This consists of 194 series of visual fields (1448 fields in total) from 98 subjects (2 subjects only had one glaucomatous eye). The overall mean VF threshold score (excluding the blind spot) is 23 with a standard deviation of 10. The median overall VF threshold is 27 dB, which better reflects the mode of the data due the large variation in scores, and the weighting towards zero (range 0-49). This can be seen in Figure 3.8. This section describes the participants, and the SAP technique used to obtain the visual fields. We then describe the spatial nature of the data and carry out exploratory data analysis, including modelling VF thresholds over time.

3.6.1 Participants

Our data was obtained from participants enrolled in the Vein Pulsation Study Trial in Glaucoma (VPSG) and on the Lion's Eye Institute trial registry, in Perth, Western Australia. All 98 participants had a form of open angle glaucoma (OAG). One subject had OAG following successful laser iridotomy for angle closure in one eye, the other eye being blind. Of the remaining 97 subjects, 8 eyes had pseudo-exfoliation syndrome and 5 had pigment dispersion syndrome, with the rest (180 eyes) having primary OAG. The mean age was 66.9 years (SD

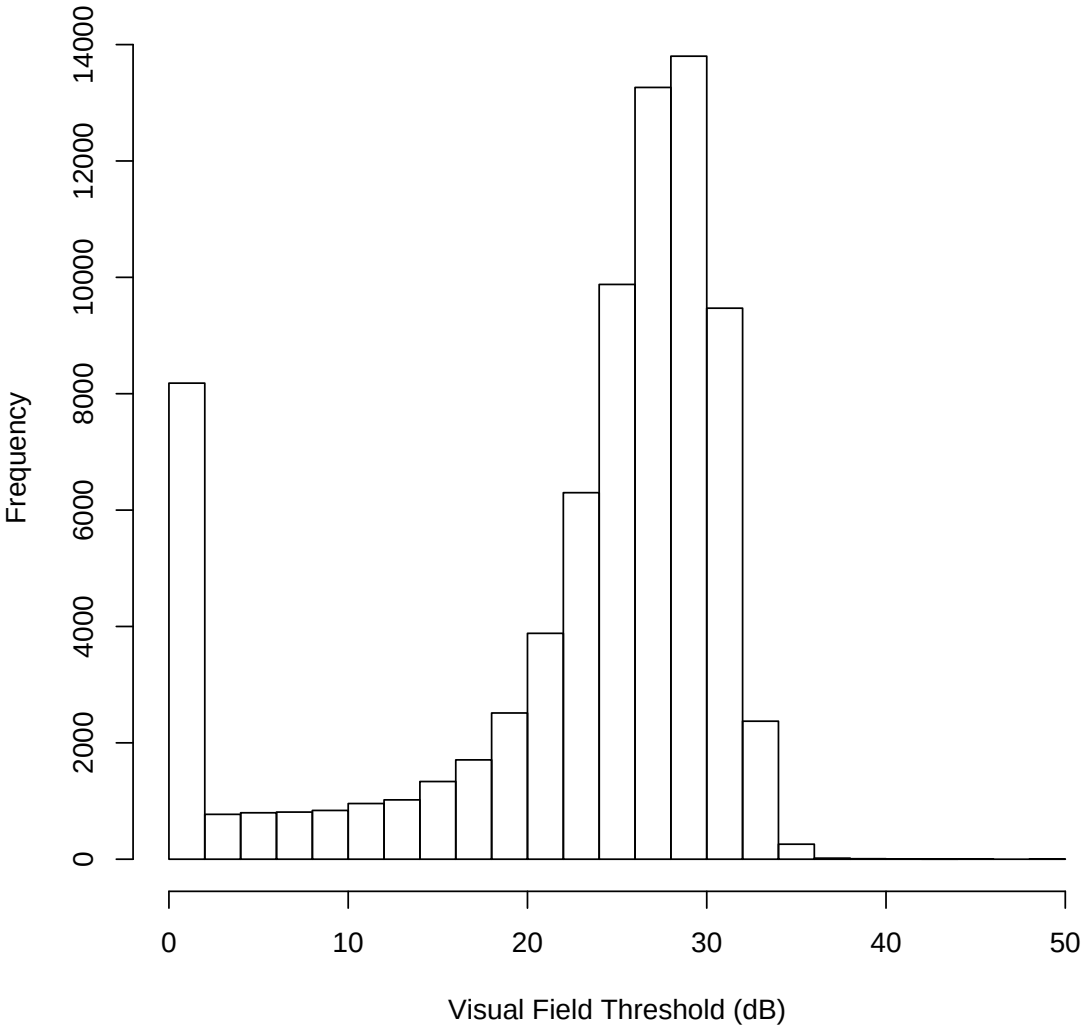


Figure 3.8: Histogram of the distribution of VF threshold intensities for all subjects (excluding the blind-spot).

9.7), with 34% males and 66% females. Informed consent was obtained for all participants. The trial adheres to the tenets of The Declaration of Helsinki and was approved by the University of Western Australia's ethics committee. The mean follow up time for participants was 2.5 years (range 0.2-9.4 years), with an average of 7.5 visual fields (range 2-21) taken during this period.

3.6.2 Standard Automated Perimetry

The Humphrey Field analyser 24-2 program was used with white on white stimulus size III. Fifteen participants from the registry used full threshold SAP, while the 81 participants from the VPSG trial used the Swedish Interactive Threshold Algorithm (SITA). As discussed previously SITA estimates are approximatly 1dB higher than the full threshold technique (Artes et al. 2002), therefore appropriate adjustments were made to the observed field thresholds of participants with fields from both techniques. All subjects had undergone at least 2 visual field tests before participating in the trial, thus minimizing the learning effect. VF output from left eyes were inverted allowing the same modelling as right eyes.

3.6.3 Spatial Correlation

Our data is similar to disease data collected over a geographical region. While disease data has population and case counts, which are aggregated over a set of disjoint geographical areas, we have scores for each area, which can be likened to counts, aggregated over a grid representing points on the eye. Therefore we would expect to see correlation in the spatial pattern of vision loss, reflecting the underlying damage to the optic disc, particularly within the sectors of each eye. However the correlation structure of the VFs is not as simple as expecting more correlation in regions that are neighbouring than regions that are not. This is largely due to that fact that the VF does not map directly to the optic disk. As discussed earlier loci may appear beside each other on the VF although they may not necessarily project along an adjacent pathway to the retina. In addition it is known that adjacent points within a sector (Figure 3.4) are more similar than adjacent points between sectors. Another complication in the VF is the blind spot of the optic disk, which if measured correctly will record values of

0 dB, and therefore does not undergo vision loss in the same manner as the rest of the eye. Figure 3.9 shows the median sensitivity (dB) for each point on the visual field, and illustrates the spatial correlation across the visual field.

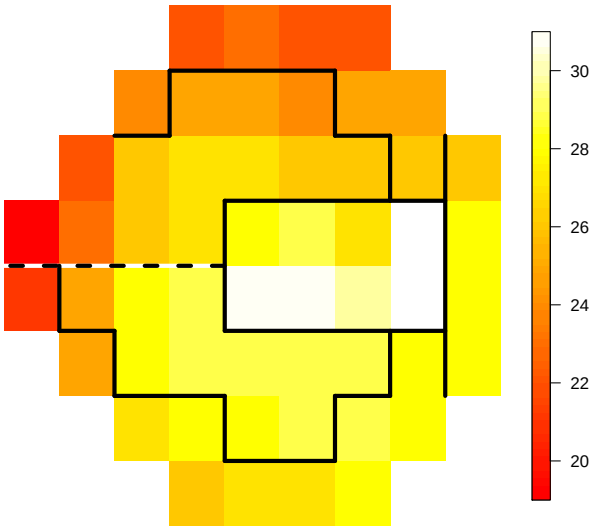


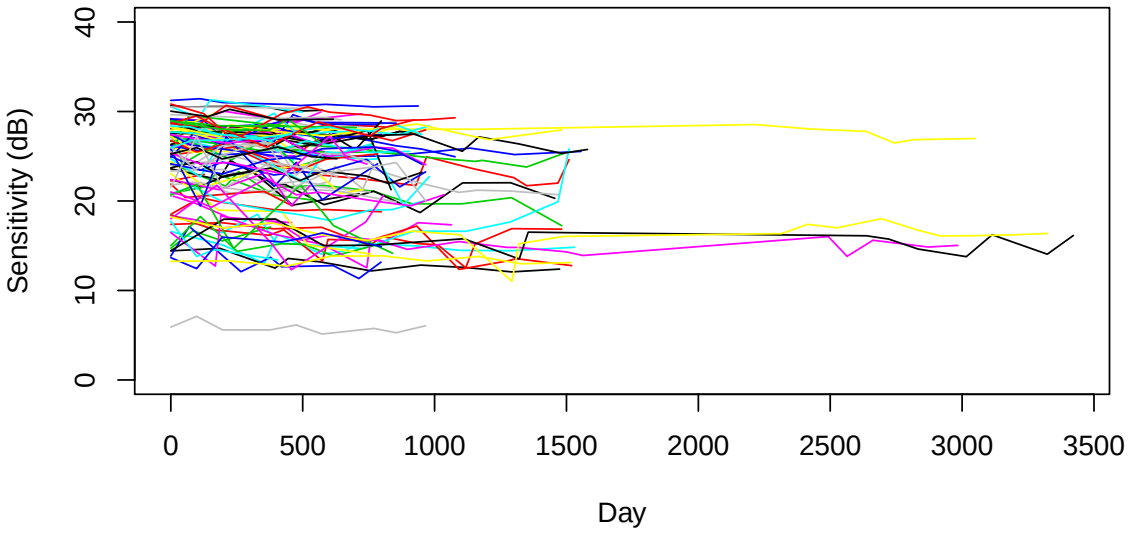
Figure 3.9: Mean VF sensitivity (dB) for all 194 eyes by VF location

3.6.4 VF Trend

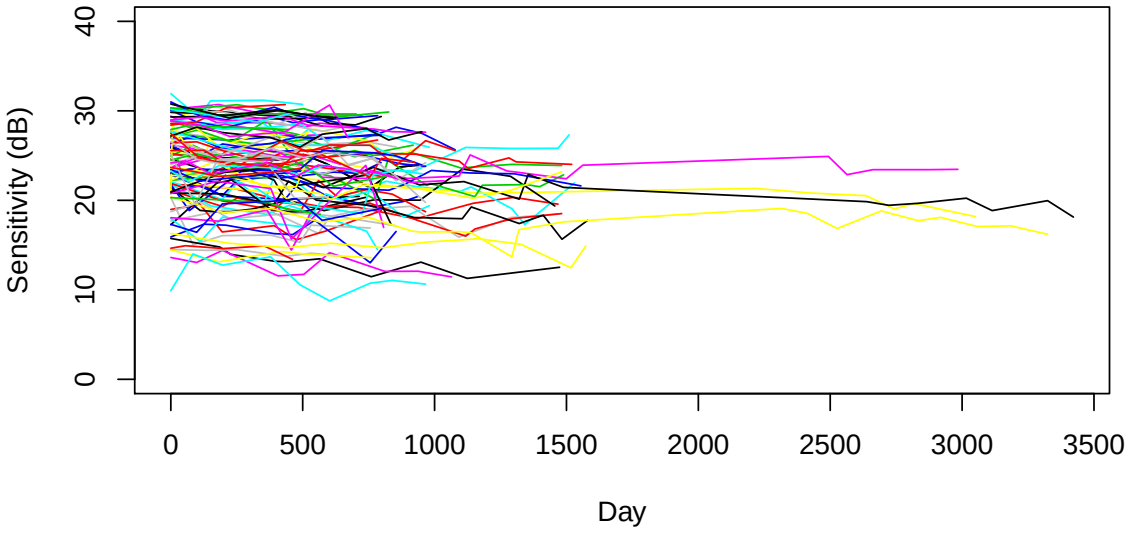
As discussed in Chapter 2 once cell death has occurred, vision loss is permanent. Therefore if the VF could be measured without error, measurements would either be stable or decreasing over time. Figure 3.10 shows the average trend of each eye over time, while Figure 3.11 shows individual trend for each VF location for two eyes.

3.7 Modelling Measurement Error

To investigate measurement error we examined a representative subset of 411 full threshold visual fields from 34 eyes (from 17 subjects). The difference in re-tested loci within each field were used to estimate variation (see STF in Section 3.3.1). A mean variance of 9.4 dB^2 was

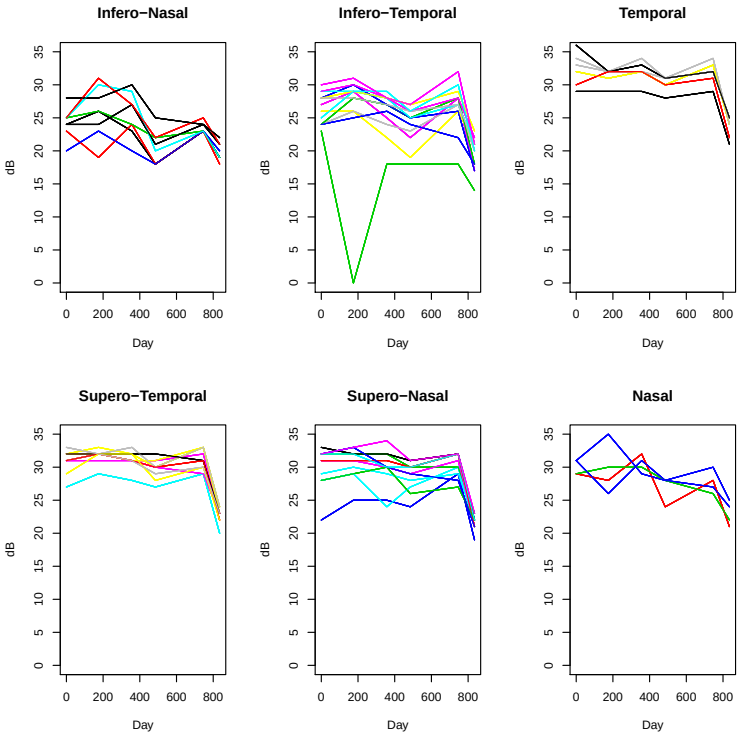


(a) Right Eye

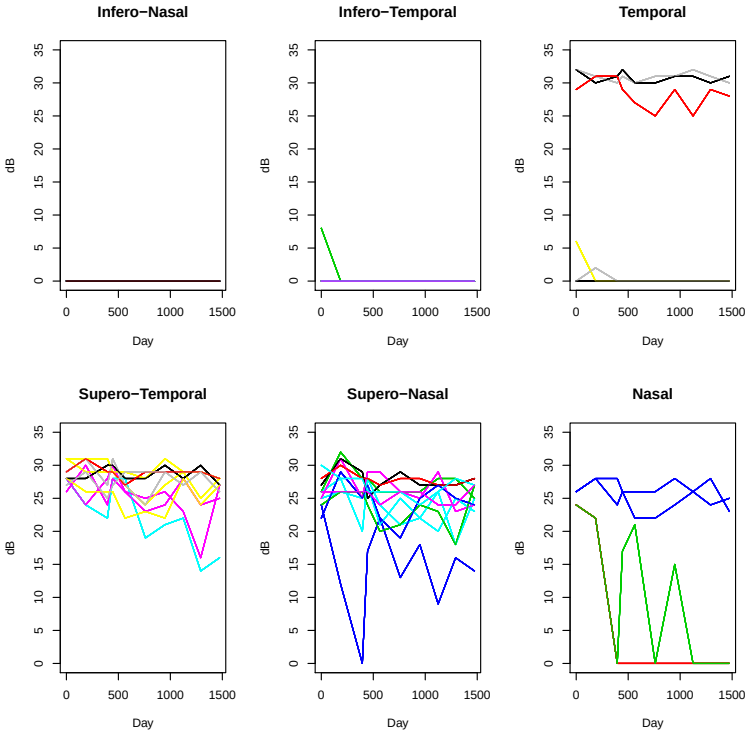


(b) Left Eye

Figure 3.10: Average trend of VF thresholds (sensitivity) plotted for each eye.



(a) Eye One



(b) Eye Two

Figure 3.11: Line graphs showing observed VF sensitivities in dB over time separated by sector. Each line represents one of the 52 non-blind spots on the visual field.

calculated for visual field loci across the eye. Figure 3.12 shows the standard deviation at each locus within the eye. Table 3.7 presents our standard deviations by location. Comparing these to those of Heijl et al. (1987) (Section 3.3, Table 3.4) there are no notable differences. Heijl et al. (1987) reported slightly more variation in the outer loci. However it should be noted that standard deviations from Heijl et al. (1987) are inter-test variations, while ours are intra-test variations. It should also be noted that Heijl et al. (1987) looked at variation in normal subjects, while our dataset consisted of glaucomatous eyes.

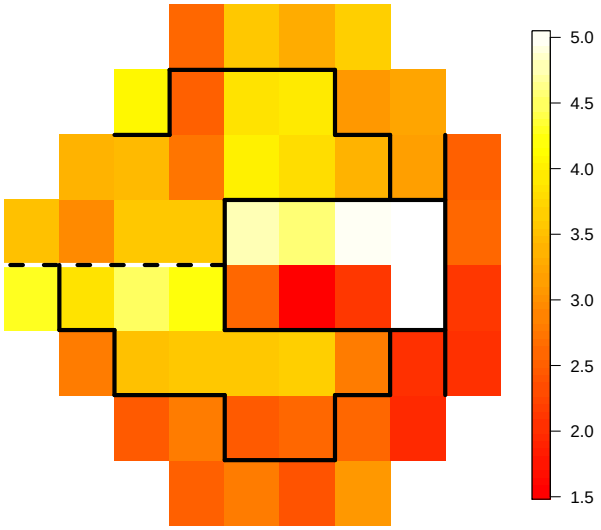


Figure 3.12: Heat map of mean standard deviations for individual loci for glaucomatous eyes.

Table 3.8: Mean standard deviations for individual loci for intra-test glaucomatous.

			2.58	3.55	3.29	3.63		
		4.11	2.50	3.88	3.91	3.07	3.21	
	3.35	3.47	2.75	4.01	3.82	3.39	3.16	2.54
3.51	2.96	3.62	3.55	4.82	4.60	5.05		2.62
4.30	3.85	4.50	4.24	2.61	1.48	2.06		2.08
	2.83	3.53	3.62	3.55	3.63	2.77	2.04	2.00
		2.48	2.77	2.45	2.56	2.61	1.93	
			2.53	2.82	2.37	3.10		

In Section 3.3.2 we discussed variation due to eccentricity and severity of disease. We therefore attempt to model the magnitude of measurement error at each locus in terms of the following variables:

- **Severity of disease.** Severity of disease was represented by VF loci means. Means for each point on each eye at each time point were calculated from the intra-test repeated measures data. A loci mean was then calculated from these values.

- **Eccentricity.** Eccentricity was modelled by distance from fixation point. The centre of fixation is considered 0,0. Every location is then considered 6 degrees apart, 3 degrees in both directions from the fixation point.

- **Sector.** Sector was treated as a factor as as defined by Garway-Heath et al. (2000) (see Figure 3.4).

- **Locus.** Locus was treated as a factor, including the 52 non-blind loci.

- **Edge Point.** Edge point was defined as whether or not the point is on the edge of the 24-2 field. It was treated as a binary factor.

- **Age.** Age was calculated in years at the time of each visual field test.

- **Person.** Person was incorporated as a random effect and included in a mixed model.

Figure 3.13 plots the visual field attributes listed above against the magnitude of measurement error as represented by log(variance+0.5). A small constant, 0.5, was added to avoid taking the log of 0. R^2 values from univariate modelling are included in the caption. We transformed measurement error due to its skewed distribution. Significant associations were found in univariate modelling for all attributes except age which is as expected due to the data being from only 17 people. While significant associations were noted it can be seen, from the plots and the R^2 values, that little of the variation seen in measurement error is due to any one visual field attribute.

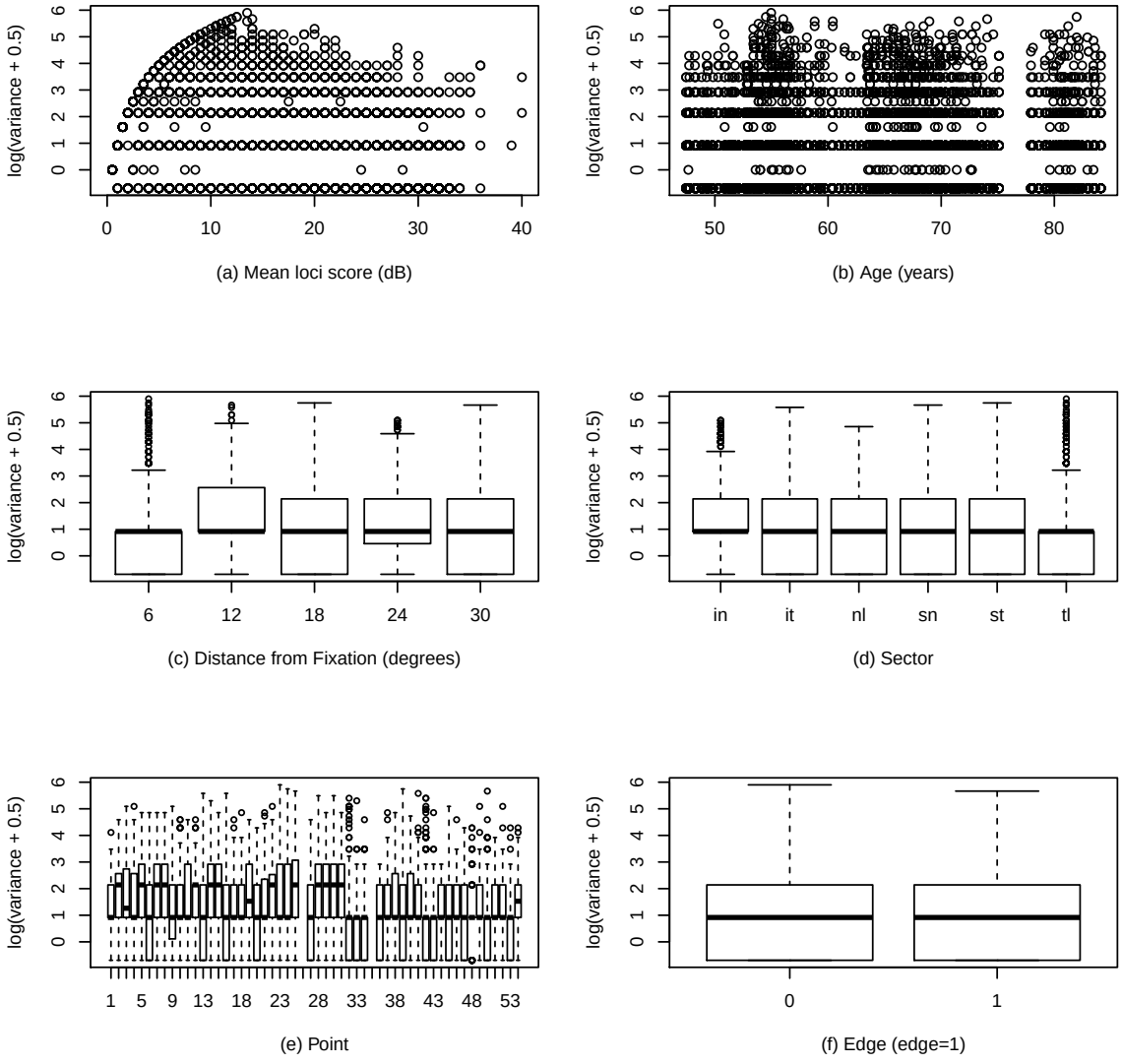


Figure 3.13: Plots show correlation between the magnitude measurement error (log(variance + 0.5)) and visual field attributes. R^2 values for univariate models are as follows; (a) 0.1754, (b) <0.0001, (c) 0.0040, (d) 0.0152, (e) 0.0924, (f) 0.0007.

Multivariate modelling was carried out using backward selection with a removal criterion p > 0.05. The two significant variables were severity of disease (modelled by mean loci score) and locus modelled as a factor. The R^2 value for this model, 0.189, shows little of the variation being explained by these two variables. As the magnitude of measurement error can vary from person to person we used mixed models to incorporate person as a random effect. While the random effect for person is significant it only explains 10.1% of the total variance of measurement error.

The measurement error itself is subject to a lot of variation and modelling the variation in error seems extremely difficult. Given the models showed low power, even after accounting for person to person variation we adopted the principle of parsimony, and chose to model measurement error as a constant. While we have chosen to model measurement error as a constant the actual value will be allowed to vary for each individual based on the prior distribution explained in the following chapter (Section 4.3.2). This allows for diversity which clearly exists between people. Because each parameter will be updated in context of the other parameters this will also allow the measurement error to be calculated with reference to the eye mean.

3.8 Chapter Summary

In this chapter we have introduced visual field data and described standard automated perimetry. We have reviewed the current methods for analysing VF data, and highlighted some of the problems with these methods. In Section 3.6 we describe and carry out exploratory data analysis. We discussed the spatial nature of the data and indicated the importance of taking this into account when modelling VF data. We showed that a linear trend is sensible to model VF loss, and that while many variables affect measurement error, the relationships are weak and patterns are unclear so that modelling measurement error as a constant is a reasonable parsimonious approach. The next chapter discusses the methodology required to implement these modelling requirements and applies this approach to our dataset.

Chapter 4

Spatial-Temporal analysis of Visual Field Data

4.1 Introduction

Due to the spatial nature of the visual field data described in Section 3.6.3, we can apply disease mapping techniques to the visual field datasets. These methods cannot be applied indiscriminately to our data since the visual field has spatial properties that are not encountered in geographical disease mapping. In particular, the spatial relationships on the visual field arise from a somewhat complicated projection from the retina. As mentioned above, loci may appear beside each other on the visual field though they do not necessarily sit next together on retina, and the visual field includes a blind spot corresponding to the point where the optic nerve enters the eye (Strouthidis et al. 2010).

The following section introduces disease mapping and explains common statistical methodology. Section 4.3 describes the adaptions necessary to apply these methods to visual field data. In Section 4.4 we present out results, and include a comparison to current PLR methods reviewed in 3.5.2. This is followed by our discussion and conclusions.

4.2 Disease Mapping Background

Disease mapping traditionally investigates the geographical distribution of a disease within a spatially defined region. 'Disease mapping studies aim to summarize spatial variation in disease risk, in order to assess and quantify the amount of true spatial heterogeneity and the associated patterns, to highlight areas of elevated or lowered risk and to obtain clues as to the disease aetiology' (Best et al. 2005). Cases of disease are plotted either individually if their exact location is known, or as counts within smaller regions of the disease map. Cases are then compared to the expected number of cases, given the population characteristics which also have a geographical distribution (Lawson & Williams 2001). Cases are compared to either the standard rate of disease being mapped, or using a control rate from an independent disease within the same region. This is crucial to test whether there is any difference in actual disease prevalence from the expected background level of disease. Without this standardization maps can be misleading as a high rate of a disease in a certain area may be due to the difference in population of urban and rural regions.

While the importance of disease mapping has been recognized since Snow's mapping of the London cholera outbreak in 1854 (Snow 1936), it has only become a common practice more recently due to the availability of data and the advancement of computers. The development of statistical methods for disease mapping has been even more recent beginning with the introduction of Markov Chain Monte Carlo (MCMC) methods to disease mapping by Besag et al. in 1991. Since then a variety of methods have been proposed particularly within the Bayesian framework which provides a natural paradigm for the often hierarchical structure of spatial data (Best et al. 2005).

This section summarizes the common statistical methods used in disease mapping, more specifically the BYM model and conditional autoregressive prior which will be employed in our analysis.

4.2.1 Disease Mapping Methods for Count Data

As mentioned above, there are two basic forms of spatial data. These are point-reference data, and areal data. Point-reference data exists when the exact location of the individual disease occurrence is known, for example the coordinates of an address. When the exact location of the disease occurrence is unknown, areal or count data arises. This type of data is more common in health data for reasons such as confidentiality. Therefore areal data consists of counts of data within small administrative regions such as zip codes or counties. Because visual field threshold data can be likened to counts this review focuses on disease mapping methodology for count data (Lawson & Williams 2001).

Wakefield (2007) reviewed the methods for analysing count data in the context of disease mapping. Given that counts in areas that are geographically close will be correlated with one another, disease mapping uses this to its advantage to provide estimates of risk which are smoothed across the neighbouring regions. The simplest approach to disease mapping is to calculate the Standard Mortality Ratio (SMR), which is the ratio of observed counts (y_i) to expected counts (e_i):

$$\hat{\theta}_i = \frac{y_i}{e_i}, \tag{4.1}$$

where the expected count is calculated as the number of cases expected in the population under the assumption of uniform risk. Due to the simplicity of the SMR there are many drawbacks. Because it is a ratio, a relatively small change in the expected value can create a large change in the SMR.

For more complex data a modelling approach should be used (Wakefield 2007). The classic disease mapping model assumes a Poisson distribution for disease count:

$$y_i \sim \mathsf{Poisson}\left(e_i\theta_i\right), \tag{4.2}$$

where θ_i is the relative risk for the ith region, e_i is the expected count or rate and y_i is defined

as above. Because the goal of disease mapping is to see if the observed counts are different from the background level of disease, it is important to include confounding factors in the calculation of e_i. When this is not possible the Poisson model can be expanded to include the confounding factors as fixed effects on the relative risk in a regression context (Lawson et al. 2003):

$$\theta_i = \mathsf{exp}\left\{x_i^T \boldsymbol{\beta}\right\}, \tag{4.3}$$

where x_i is a vector of covariates and $\boldsymbol{\beta}$ is a vector of parameters.

It is also important to account for unobserved confounding effects, for example spatial dependence. By modelling spatial dependence we are trying to model unobserved covariates, by using location as a substitute (Wakefield 2007). Generally for spatial models, we assume that areas close to each other will be more similar than those further away. Therefore the location provides us with more information which will provide us with more reliable relative risk estimates. Spatial dependence is modelled in the form of random effects. While random effects can be modelled using frequentist statistics it is more natural to model them within a Bayesian framework, as random effects have prior distributions (Lawson et al. 2003). The Bayesian framework also allows for the fitting of more complex models, therefore we choose a Bayesian approach and will fit our model using an MCMC approach as outlined in Section 2.3.

The classic Bayesian disease mapping model is an extension of Equation 4.2. The model is extended by applying prior distributions to the parameters. The common prior distribution for θ_i is the gamma distribution, thus producing a Poisson-gamma model. Further building the hierarchy, hyper-priors (h) can be applied to parameters a and b of the gamma prior. Hence,

$$y_i|\theta \sim \mathsf{Poisson}\left(e_i\theta\right) \tag{4.4}$$

$$\theta_i|a,b \sim \mathsf{Gamma}\left(a,b\right) \tag{4.5}$$

$$a|\nu \sim h_a\left(\nu\right) \tag{4.6}$$

$$b|\rho \sim h_b\left(\rho\right) \tag{4.7}$$

However, the gamma prior assumes independence between $\theta_i's$, which does not allow for spatial correlation between neighbouring regions. To allow for this the log-normal model was developed. These are discussed below.

Besag, York and Mollié model

The most common model to account for spatially correlated variation in disease mapping is the Besag, York and Mollié (BYM) model (Best et al. 2005, Besag et al. 1991). Here y_i is modelled as in Equation 4.4, however θ is composed as follows:

$$\mathsf{log}\theta_i = \alpha + \delta_i + v_i, \tag{4.8}$$

where α is the overall relative risk, δ_i represents correlated heterogeneity and v_i represents the uncorrelated heterogeneity. Correlated and uncorrelated heterogeneity account for the variance over-dispersion of the model. Uncorrelated heterogeneity accounts for the aspatial variation, while correlated heterogeneity models variation due to spatial dependence. A normal prior is assumed for uncorrelated heterogeneity:

$$v_i \sim \mathsf{Normal}\left(0,\tau_v^2\right). \tag{4.9}$$

This type of data lends itself to the Gaussian Random Field process. The Gaussian Random Field process assumes that any particular area or point depends on the other points elsewhere in the space. It assumes normality with local correlation specified in a particular

way. The conditional autoregressive (CAR) model is a Gaussian Random Field process which accounts for correlated heterogeneity by specifying the correlation for a given point to be defined in terms of just its direct neighbours. The CAR model was first suggested by Besag (1974), however its wider use only began in the 1990s (Besag et al. 1991), due to increase in computing power and the ease of Gibbs sampling and MCMC methods (Banerjee et al. 2004).

As mentioned, the CAR prior allows for a spatial correlation structure to be specified where each value depends on the values of its neighbours. In doing so it reduces the variance of estimates by allowing strength to be borrowed from neighbouring regions. In a conditional model the neighbouring regions must be defined. In simple cases a neighbour can be defined by sharing a common boundary, and an adjacency matrix can be created to this effect. The denoted adjacencies allow for a weighted average to be generated from the surrounding loci. This information is then used by the CAR prior detailed below in Equations 4.10-4.12. The CAR model is as follows:

$$[\delta_i|\delta_j, i \neq j, \sigma^2_{\delta_u}] \sim \mathsf{Normal}(\bar{\delta_i}, \sigma^2_{\delta_i}) \tag{4.10}$$

where

$$\bar{\delta_i} = \left(\frac{1}{\sum\limits_{j\neq i} w_{ij}}\right) \sum_{j\neq i} w_{ij}\delta_j \tag{4.11}$$

$$\sigma^2_{\delta_i} = \frac{\kappa}{\sum\limits_{j\neq i} w_{ij}} \tag{4.12}$$

where $w_{ij} = 1$ if i and j are adjacent neighbours and $w_{ij} = 0$ otherwise.

Bernardinelli, Clayton & Montomoli (1995) recommends assigning inverse χ^2 or inverse gamma prior distributions to parameters τ_v^2 and κ as it is computationally convenient. If the distribution from which we are sampling is assumed to be normal then the full conditional posterior distribution can be derived and therefore sampled using Gibbs sampling. If κ has a prior distribution $\kappa \sim \mathsf{InverseGamma}(\frac{1}{2}, \frac{e}{2})$ (where e is a small constant, in our case 0.01 (Besag et al. 1991)), then the conditional posterior for κ is given by:

$$[\kappa|\delta_j \forall j] \propto \kappa^{-n/2} \mathsf{exp}\left\{-\frac{1}{2\kappa}\left[e + \sum_{i \neq j} w_{ij}(\delta_i - \delta_j)^2\right]\right\} \tag{4.13}$$

(Besag et al. 1991, Banerjee et al. 2004).

The model has been further extended by Bernardinelli, Clayton, Pascutto, Montomoli, Ghislandi & Songini (1995) to include spatial/time correlations. This involves adding a time trend (β) and a space time interaction (η_i):

$$\mathsf{log}\theta_{ik} = \alpha + \delta_i + v_i + \beta t_k + \eta_i t_k, \tag{4.14}$$

where a similar type of CAR prior is assigned to the space-time effects, η_i. Thus $\bar{\eta_i}$ and $\sigma^2_{\eta_i}$ are defined as for $\bar{\delta_i}$ and $\sigma^2_{\delta_i}$.

$$[\eta_i|\eta_j, i \neq j, \sigma^2_{\eta_i}] \sim \mathsf{Normal}(\bar{\eta_i}, \sigma^2_{\eta_i}). \tag{4.15}$$

While CAR priors are most commonly used to account for spatial correlation in the disease mapping setting, CAR models have also been applied in a wide range of other disciplines. Eckert et al. (2010) use CAR models to link spatio-temporal trends in avalanches to climate change over a 60 year period in the northern French Alps. Sims et al. (2008) used CAR models to map fisheries by-catch data, where events are rare and the fishing effort is highly variable both over space and time. CAR models have also been applied to investigate the spatial relationships between traffic congestion and road accidents (Wang et al. 2009).

4.3 Developing the Model

Disease mapping techniques can be applied to longitudinal sequences of visual field data, with the hope of reducing noise and correctly accounting for spatial correlation. The traditional disease mapping techniques, described above, cannot indiscriminately be applied to the map of the eye. Therefore in this section we explain the adaption made to disease mapping methodology in order to account for the differing characteristics of visual field data, as well as

the physiology of the eye (see Section 3.6.3 in the previous chapter). We exclude the two loci close to the blind spot, thus we develop our model for the 52 remaining loci of the visual field. Because we implement our model within a Bayesian statistical framework, we are required to assign prior distributions that summarize any available prior information for our model parameters. In the subsections following we discuss individual components of our model and explain the rationale behind our modelling choices. We then present the model as a whole, and outline model implementation.

4.3.1 Modelling Observed Visual Field Thresholds

The visual field map (see Figure 3.4), like an areal disease map, has one value for each region on the map. However, unlike traditional disease mapping our value is a score as opposed to a count. We can assume that our threshold score will act similarly to a count, however unlike a count it has the potential to become negative, and is theoretically continuous.

For a given individual, we let y_{ik} denote the threshold measurement taken at visual field grid location i in the kth visual field, where the grid cells are numbered by row from top left to bottom right. We denote by $t_1, \ldots, t_k, \ldots t_{n_i}$ the time (in years) at which the visual field measurements where obtained. If two fields are taken the same day this corresponds to $t_k = t_{k+1}$. We assume that these observed values follow a normal distribution, but are censored to the left at zero. This accounts for the idea that a dB threshold of below zero could be obtained if a higher light intensity was tested on the Humphrey visual field analyser. For example an intensity of 1000asb corresponds to 10dB (cf. Table 3.1). The maximum intensity tested is 10,000asb and corresponds to a dB score of 0. Therefore if we tested an intensity brighter than 10,000 we would get a negative dB threshold (remembering that dB are measured on the log scale). Parameter y^*_{ik} is the notional values of y_{ik} that would be obtained if one could test unlimited intensities of light. That is, $y_{ik} = \mathsf{max}(y^*_{ik}, 0)$ where

$$y^*_{ik} | \mu_{ik} \sim \mathsf{Normal}(\mu_{ik}, \sigma^2_\epsilon). \tag{4.16}$$

The y^*_{ik} are conditionally independent given μ_{ik}. The spatial and temporal dependence in the data are generated by the manner in which we model μ_{ik}. Specifically we model the spatio-temporal variation in the visual field by

$$\mu_{ik} = \alpha + \delta_i + \beta t_k + \eta_i t_k. \tag{4.17}$$

The overall mean of the eye is represented by α, and the overall temporal trend is described by β. Spatially correlated heterogeneity is accounted for by δ_i and thus adjusts α to provide a mean for each point, i, on the eye. Spatio-temporal variation is described by η_i which adjusts the overall trend, β, to allow each locus on the eye to vary over time. The uncorrelated heterogeneity term seen in Equations 4.8 and 4.15 is not seen in Equation 4.17. This is because we are using a normal model, therefore uncorrelated heterogeneity is incorporated into the normal variation as per Equation 4.16.

4.3.2 Modelling Measurement Error

Measurement error is described by our uncorrelated heterogeneity term, σ^2_ϵ. We chose to model measurement error as a constant (see Section 3.7), therefore we assign an inverse gamma prior as recommended by Besag et al. (1991). Thus:

$$\sigma^2_\epsilon \sim \mathsf{InverseGamma}(5.6, 43.2). \tag{4.18}$$

An informative prior was allocated to σ^2_ϵ, as we had pre-existing information about the measurement error within the visual field. A representative subset of 411 full threshold visual fields from 34 eyes (17 subjects) were examined. This subset was chosen due to the availability of the repeated measures data. The difference in re-tested loci within each field were used to estimate variation. A mean variance of 9.4 dB (variance of variance 36 dB) was calculated for visual field readings at individual loci (y_{ik}) (as discussed in Section 3.7). Values for the prior were converted to create an informative inverse gamma prior centred around 9.4dB with a scale parameter of 36dB.

4.3.3 Specification of the Mean

A normal prior was employed for alpha, $\alpha \sim \mathsf{Normal}(24, 18^2)$, where 24 is the mean threshold score and 18 is two times the standard deviation across all loci of all eyes at time point zero. The standard deviation was doubled to make it less informative.

4.3.4 Specification of the Trend

We investigated the relationship of visual field scores with time (see Section 3.6.4). A linear trend appears appropriate in general, with the suggestion of non-linearity appearing in a small number of very long series of visual field data. We centre time at zero and rescale to have a standard deviation of 0.5. We assign a Cauchy prior to our trend parameter, β. When little is known of the trend Gelman et al. (2008) suggests using $\mathsf{Cauchy}(0, 2.5)$ as an uninformative prior (refer to Section 2.3.2). However we know that a change of 10 dB per year would be at the extreme end of clinical experience therefore we employ $\beta \sim \mathsf{Cauchy}(0, 5)$ instead of a scale 22.5 (2.5 times the sd of scaled y_{ik}).

4.3.5 Modelling the Spatial Correlation

In a review of Bayesian spatial models Best et al. (2005) found that the BYM was a good choice for modelling the spatial variation of a single disease (see Section 4.2.1). Best et al. (2005) noted that this model and others which preformed well all used the adjacency matrix as a means for specifying spatial structure. This seemed to allow 'flexibility for capturing appropriate features of the risk surface' (Best et al. 2005). We therefore decided the BYM was a suitable for our data structure. The parameters δ and η are specified as in Equations 4.11 - 4.13.

Specification of the Adjacency Matrix

As discussed in Section 4.2.1, when using a CAR prior an adjacency matrix must be specified to describe the spatial dependence of the region. In most cases this requires specifying an $n \times n$ matrix $X = (x_{ij})$, where $x_{ij} = 1$ if regions i and j are neighbouring and $x_{ij} = 0$ if they are not. However in our case the blind spot acts independently from the rest of loci on the VF

(as discussed in Section 3.6.3). Therefore in addition to not modelling the two loci nearest the blindspot, we must also specify the the loci neighbouring the blind spot as non-adjacent to their neighbours (ie $x_{ij} = 0$). Additionally the neighbouring loci along the mid-line of the visual field map, do not map directly to the optic disk. Therefore infero-temporal, supero-temporal, and supero-nasal loci along the mid-line are also defined as non-adjacent. (See section 3.6.3, Figure 3.4 for more details.)

Specification of Weights

A CAR model also requires the specification of weights (w_{ij}) which dictate how much influence the neighbouring regions have on the region being evaluated (see Equations 4.11-4.12). In Section 3.6.3 we discuss how the nerve fibres from the same sector group together on the optic disk. This means adjacent loci within a sector are more dependent than adjacent loci between sectors. We base our weight choice on expert opinion, and therefore based on our adjacency matrix described earlier, we assign $w_{ij} = 1$ if i, j are adjacent and of the same sector, $w_{ij} = 0.3$ if i, j are adjacent and not of the same sector, and $w_{ij} = 0$ if they are non-adjacent. Alternative weighting schemes are considered later in Section 4.4.6.

4.3.6 The Final Model

In this section the final model is presented as a whole, where y_{ik} is the observed visual field threshold for the ith location and the kth timepoint. The parameters are summarized in Table 4.1.

$$
\begin{aligned}
y_{ik} &= \max(y^*_{ik}, 0) && (4.19)\\
y^*_{ik}|\mu_{ik} &\sim \mathsf{Normal}(\mu_{ik}, \sigma^2_\epsilon) && (4.20)\\
\mu_{ik} &= \alpha + \delta_i + \beta t_k + \eta_i t_k && (4.21)\\
[\delta_i|\delta_j, i \neq j, \sigma^2_{\delta_i}] &\sim \mathsf{Normal}(\bar{\delta_i}, \sigma^2_{\delta_i}) && (4.22)\\
[\eta_i|\eta_j, i \neq j, \sigma^2_{\eta_i}] &\sim \mathsf{Normal}(\bar{\eta_i}, \sigma^2_{\eta_i}). && (4.23)
\end{aligned}
$$

The priors for the model are:

$$\alpha \sim \mathsf{Normal}(24, 18) \tag{4.24}$$

$$\beta \sim \mathsf{Cauchy}(0, 5) \tag{4.25}$$

$$\sigma_\epsilon^2 \sim \mathsf{InverseGamma}(5.6, 43.2) \tag{4.26}$$

where

$$\bar{\delta_i} = \left(\frac{1}{\sum_{j \neq i} w_{ij}}\right) \sum_{j \neq i} w_{ij}\delta_j \tag{4.27}$$

$$\sigma_{\delta_i}^2 = \frac{\kappa_\delta}{\sum_{j \neq i} w_{ij}} \tag{4.28}$$

$$\bar{\eta_i} = \left(\frac{1}{\sum_{j \neq i} w_{ij}}\right) \sum_{j \neq i} w_{ij}\eta_j \tag{4.29}$$

$$\sigma_{\eta_i}^2 = \frac{\kappa_\eta}{\sum_{j \neq i} w_{ij}} \tag{4.30}$$

and

$$\kappa_\delta \sim \mathsf{InverseGamma}_{\dagger_{20}}(\frac{1}{2}, \frac{e}{2}) \tag{4.31}$$

$$\kappa_\eta \sim \mathsf{InverseGamma}(\frac{1}{2}, \frac{e}{2}) \tag{4.32}$$

where e is a small constant (0.01 as per (Besag et al. 1991)) and subscript $\dagger_{20}$ indicates the distribution is truncated at 20 for computational efficiency to overcome large values during the burn-in period.

We note that the trend in Equation 4.21 appears overparameterized, with identifiability issues existing between α and $\boldsymbol{\delta}$, and β and $\boldsymbol{\eta}$. This can be countered by applying mean constraints on $\boldsymbol{\delta}$ and $\boldsymbol{\eta}$. We describe how this can be achieved in the context of MCMC sampling in the following section.

Table 4.1: Parameters for final model.

Parameter	Definition
y^*_{ik}	Latent sensitivity for ith location at kth time
y_{ik}	Observed sensitivity for ith location at kth time
μ_{ik}	Mean sensitivity at ith location and kth time
σ^2_ϵ	Measurement error variance
α	Overall intercept for individual's eye
δ_i	Adjustment for spatially correlated heterogeneity
β	Temporal trend coefficient
t_k	Time in years at kth time point
δ_i	Spatially correlated adjustment to overall eye mean α
η_i	Spatially correlated adjustment to temporal trend
$\bar{\delta}_i$	Conditional mean for δ
$\bar{\eta}_i$	Conditional mean for η
$\sigma_{\delta^2_i}$	CAR variance for δ_i
$\sigma_{\eta^2_i}$	CAR variance for η_i
w_{ij}	Weights for adjacent neighbours at locations i,j
κ_δ	CAR error scaling parameter for $\boldsymbol{\delta_i}$
κ_η	CAR error scaling parameter for $\boldsymbol{\eta_i}$
e	Small positive constant 0.01 (Besag et al. 1991)
n	Number of loci

4.3.7 Model Implementation

Models were fitted in R (R Core Team 2013) using Markov Chain Monte Carlo (MCMC) methods. As outlined in Section 2.4, MCMC works by drawing simulations of model parameters from a Markov chain whose stationary distribution matches the required posterior distribution (Gamerman & Lopes 2006). We use the Metropolis-Hastings (MH) algorithm to sample values from the Markov chain. Random walk MH, where candidate values are sampled dependent on the current value of the chain, was used for sampling α, β and σ_ϵ. The remaining parameters were sampled using the Gibbs step.

We implemented the following component-wise transition MCMC algorithm:

1. Set iteration counter $z = 0$ and initialize $\boldsymbol{\delta}^{(0)}, \alpha^{(0)}, \boldsymbol{\eta}^{(0)}, \beta^{(0)}, \boldsymbol{\sigma_\delta}^{(0)}, \boldsymbol{\sigma_\eta}^{(0)}, \sigma_\epsilon^{(0)}$ as specified below.
2. Set $z = z + 1$
3. For i in 1-52, sample $\delta_i^{(z)}$, conditional on $\delta_{1:i-1}^{(z)}, \delta_{i+1:52}^{(z-1)}, \alpha^{(z-1)}, \boldsymbol{\eta}^{(z-1)}, \beta^{(z-1)}, \boldsymbol{\sigma_\delta}^{(z-1)}, \boldsymbol{\sigma_\eta}^{(z-1)}, \sigma_\epsilon^{(z-1)}$ using Equation 4.22
4. Sample $\alpha^{(z)}$ conditional on $\boldsymbol{\delta}^{(z)}, \boldsymbol{\eta}^{(z-1)}, \beta^{(z-1)}, \boldsymbol{\sigma_\delta}^{(z-1)}, \boldsymbol{\sigma_\eta}^{(z-1)}, \sigma_\epsilon^{(z-1)}$ from proposal distribution $\mathsf{Normal}(\alpha^z, 0.3^2)$ using random walk MH
5. For i in 1-52, sample $\eta_i^{(z)}$ conditional on $\eta_{1:i-1}^{(z)}, \eta_{i+1:52}^{(z-1)}$, $\boldsymbol{\delta}^{(z)}, \alpha^{(z)}$ $\beta^{(z-1)}, \boldsymbol{\sigma_\delta}^{(z-1)}, \boldsymbol{\sigma_\eta}^{(z-1)}, \sigma_\epsilon^{(z-1)}$ using Equation 4.26
6. Sample $\beta^{(z)}$ conditional on $\boldsymbol{\delta}^{(z)}, \alpha^{(z)}, \boldsymbol{\eta}^{(z)}, \boldsymbol{\sigma_\delta}^{(z-1)}, \boldsymbol{\sigma_\eta}^{(z-1)}, \sigma_\epsilon^{(z-1)}$ from proposal distribution $\mathsf{Cauchy}(\beta^z, 0.4^2)$ using random walk MH
7. Sample $\kappa_\delta^{(z)}$ using Gibbs sampling (as in Equation 4.13) conditional on $\boldsymbol{\delta}^{(z)}, \alpha^{(z)}, \boldsymbol{\eta}^{(z)}, \beta^{(z)}, \boldsymbol{\sigma_\eta}^{(z-1)}, \sigma_\epsilon^{(z-1)}$ and hence calculate $\boldsymbol{\sigma_\delta}^{(z)}$
8. Sample $\kappa_\eta^{(z)}$ using Gibbs sampling (as in Equation 4.13) conditional on $\boldsymbol{\delta}^{(z)}, \alpha^{(z)}, \boldsymbol{\eta}^{(z)}, \beta^{(z)}, \boldsymbol{\sigma_\delta}^{(z)}, \sigma_\epsilon^{(z-1)}$ and hence calculate $\boldsymbol{\sigma_\eta}^{(z)}$
9. Sample $\sigma_\epsilon^{(z)}$ conditional on $\boldsymbol{\delta}^{(z)}, \alpha^{(z)}, \boldsymbol{\eta}^{(z)}, \beta^{(z)}, \boldsymbol{\sigma_\delta}^{(z)}, \boldsymbol{\sigma_\eta}^{(z)}$ using random walk MH on the log scale with step size $\mathsf{Normal}(0, 0.075^2)$
10. Repeat steps 2-9 until required number of iterations is completed.

Lack of formal identifiability of α and δ_is, and β and η_is, is handled by replacing α with $\alpha + \frac{\sum \delta_i}{n}$ and re-centering so that $\sum \delta_i = 0$ $\left(\text{ie. } \delta_i \leftarrow \delta_i - \frac{\sum \delta_i}{n}\right)$ (Besag & Kooperberg 1995). In other words we are constraining the random effects to sum to zero and adding the difference to the intercept term, so that μ_{ik} remains unchanged. Parameters $\sum \eta_i$ and β are handled in the same way.

Models were run for 100,000 iterations, with burn in period of 12000, and a thinning factor of 40. Parameter α was initialized at 23, the mean sensitivity of all eyes over all time points. Trend β was initialized at zero. Parameters κ_δ and κ_η were initialized at 5 and 1 respectively based on our experience with the variability of the visual field measurements. Parameter $\boldsymbol{\eta}$ is initialized at 0, δ_i at centred y_{i1}.

4.4 Results

This section presents the results of our model, which we term SPROG for Spatial Progression. We present overall results as well as providing results for two example eyes, one where glaucoma is progressing and one where glaucoma is stable. We also compare our results to current PLR methods and a clinical reference. Finally we present our discussion and conclusions.

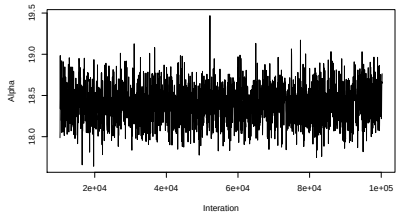
4.4.1 Convergence Diagnostics

The Deviance Information Criterion (DIC) was calculated and used to test the goodness of fit of our models (Gamerman & Lopes 2006). The calculation of DIC was based on the posterior mean estimates for parameters with a focus on random effects (see Section 2.4.4). Like other Information Criteria it penalizes models which provide a poor fit and/or are overly complex with superfluous parameters. For each eye, significant progression was defined as a one sided Bayesian p-value of ≤ 0.05 for the overall eye trend, β.

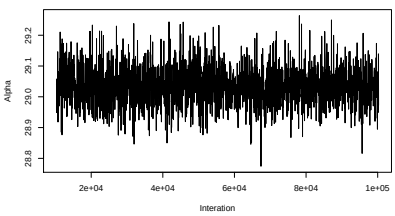
Convergence was judged using the Geweke diagnostic at the 5% significance level (see Section 2.4.3), implemented in R's coda package (Geweke 1992, Plummer et al. 2006, R Core Team 2013). Of the models 194 models run, 192 converged within the 12,000 iteration burn in period. Acceptance rates for all parameters sampled through MH random walk were close to 50%. Figure 4.1 shows trace plots for parameters α, β, σ_ϵ, κ_δ, and κ_η for a progressing eye and a stable eye. Trace plots for the 52 δs and ηs can be found in Appendix A. Similarly autocorrelation plots for δs and ηs can be found in Appendix A, while the autocorrelation plots for the remaining parameters are displayed in Figure 4.2. Following thinning with a factor of 40, no parameters showed indications of autocorrelation.

4.4.2 Posterior Summaries for Model Parameters

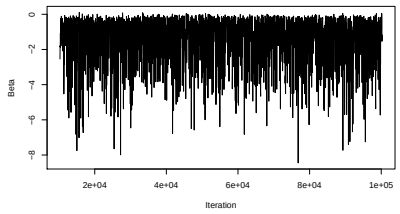
Means, 95% credible intervals and Bayesian p-values are calculated from the posterior distributions of the model parameters. Table 4.2 summarizes the parameters for a stable and progressing eye. We are particularly interested in the overall eye mean α and the parameter



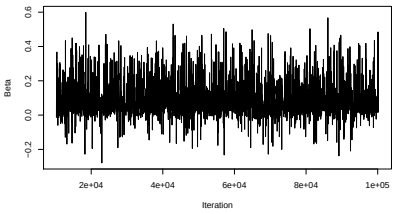
(a) α for Progressing Eye



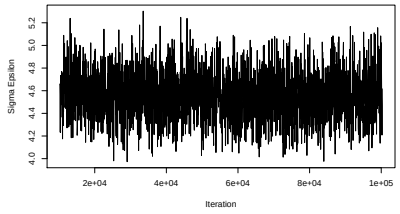
(b) α Stable Eye



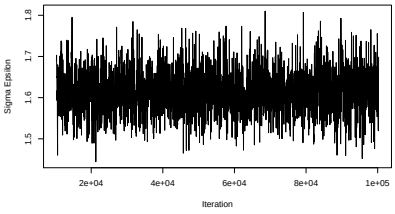
(c) β (Progressing Eye)



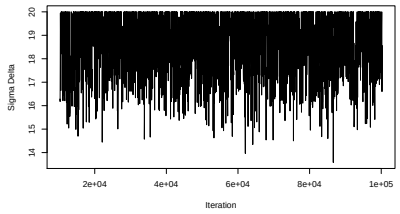
(d) β Stable Eye



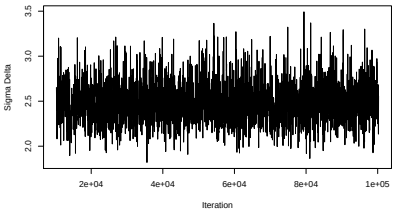
(e) σ_ϵ (Progressing Eye)



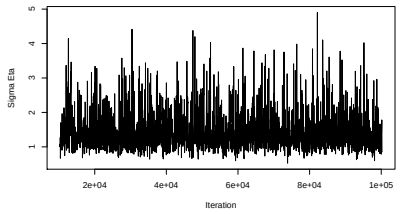
(f) σ_ϵ (Stable Eye)



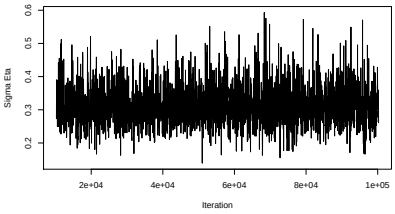
(g) κ_δ (Progressing Eye)



(h) κ_δ (Stable Eye)

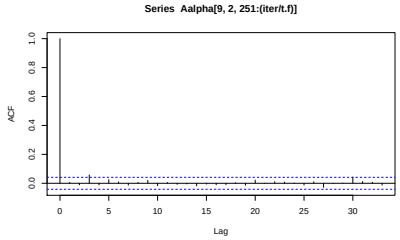


(i) κ_η (Progressing Eye)

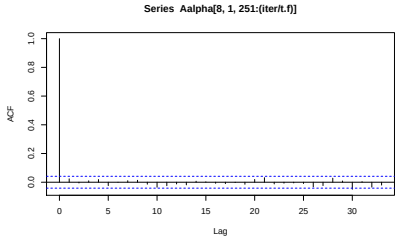


(j) κ_η (Stable Eye)

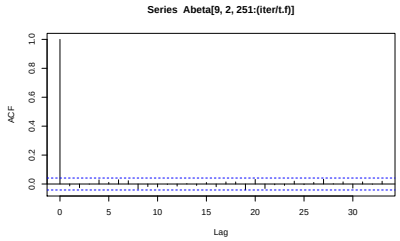
Figure 4.1: Trace plots of model parameters for a progressing (left column) and stable eye (right column). Excludes burn in of 12,000 iterations.



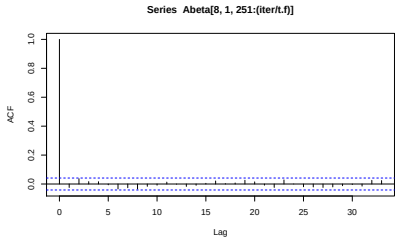
(a) α for Progressing Eye



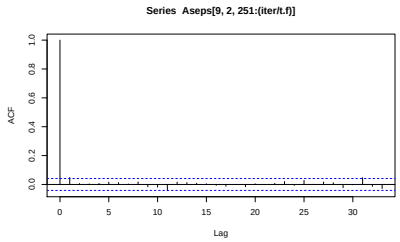
(b) α Stable Eye



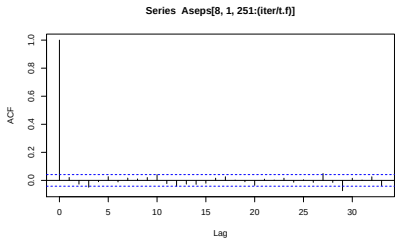
(c) β (Progressing Eye)



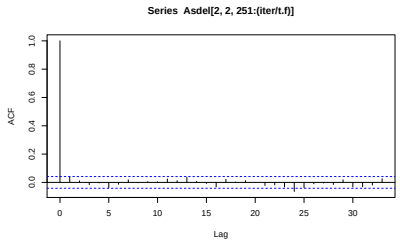
(d) β Stable Eye



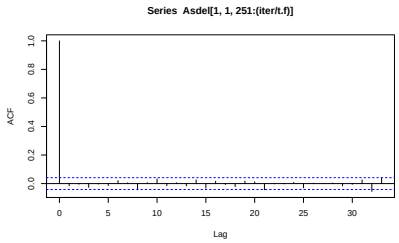
(e) σ_ϵ (Progressing Eye)



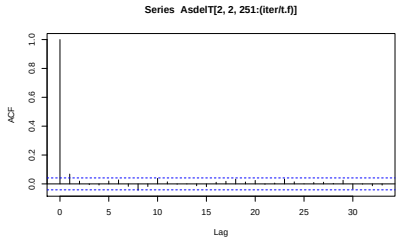
(f) σ_ϵ (Stable Eye)



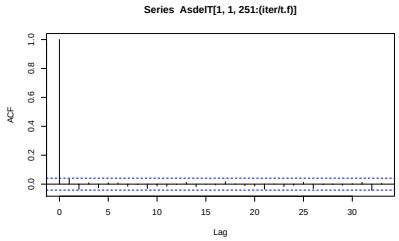
(g) κ_δ (Progressing Eye)



(h) κ_δ (Stable Eye)



(i) κ_η (Progressing Eye)



(j) κ_η (Stable Eye)

Figure 4.2: Autocorrelation plots (after thinning) for model parameters for a progressing (left column) and stable eye (right column). Excludes burn in of 12,000 iterations.

β which identifies progressing eyes. Individual loci means ($\alpha + \delta_i$) and slopes ($\beta + \eta_i$) are presented in Appendices B-C for the same stable and progressing eyes. Overall means (α) and slopes (β) for all eyes are presented in Appendix C. Using the SPROG progression criteria $\beta < 0$, with statistically significant slopes (Bayesian p-value < 0.05), our model showed 42% of the eyes progressing, at an average rate of -0.78 (range -0.27,-1.76) dB per year.

Table 4.2: Parameter summary for a typical progressing and stable eye.

	Stable Eye			Progressing Eye		
	Mean	95% C.I.	Bayesian P-value	Mean	95% C.I.	Bayesian P-value
α	29.04	(28.91,29.16)	<0.0001	18.40	(17.94,18.86)	<0.0001
β	0.09	(-0.08,0.38)	0.8444	-1.53	(-5.15,-0.06)	0.0076
σ_ϵ	1.61	(1.51,1.72)	<0.0001	4.55	(4.10,5.02)	<0.0001
κ_δ	2.49	(2.06,3.03)	<0.0001	18.51	(15.52,20.00)	<0.0001
κ_η	0.31	(0.21,0.45)	<0.0001	1.45	(0.78,3.07)	<0.0001
δ_1	-1.41	(-1.61,-1.20)	<0.0001	-5.91	(-6.61,-5.15)	<0.0001
δ_2	-0.54	(-0.76,-0.32)	<0.0001	-2.06	(-2.79,-1.28)	<0.0001
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
δ_{53}	0.84	(0.62,1.04)	<0.0001	4.59	(3.82,5.31)	<0.0001
δ_{54}	0.71	(0.50,0.93)	<0.0001	2.79	(1.99,3.57)	<0.0001
η_1	-0.01	(-0.28,0.24)	0.4658	-0.20	(-1.73,1.18)	0.3787
η_2	-0.01	(-0.25,0.23)	0.4720	-0.14	(-1.47,1.12)	0.3978
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
η_{53}	-0.01	(-0.26,0.22)	0.4653	0.17	(-0.93,1.54)	0.4178
η_{54}	-0.02	(-0.29,0.23)	0.4418	0.24	(-0.91,1.89)	0.4049

4.4.3 Predicted versus Observed

Figure 4.3 shows the fit of our model by sector compared to the observed thresholds, for a typical progressing eye. We can see that the model reduces the effect of overall variation in the eyes by allowing the predicted loci to be influenced by their neighbouring loci. This also means the result is not overly affected by the occurrence of outliers. However, the model is still flexible enough to allow a certain locus within a sector to differ if the mean of a certain locus in the eye is consistently lower (Figure 4.3, temporal sector). Figure 4.4 shows the flat trend of typical eye with stable glaucoma.

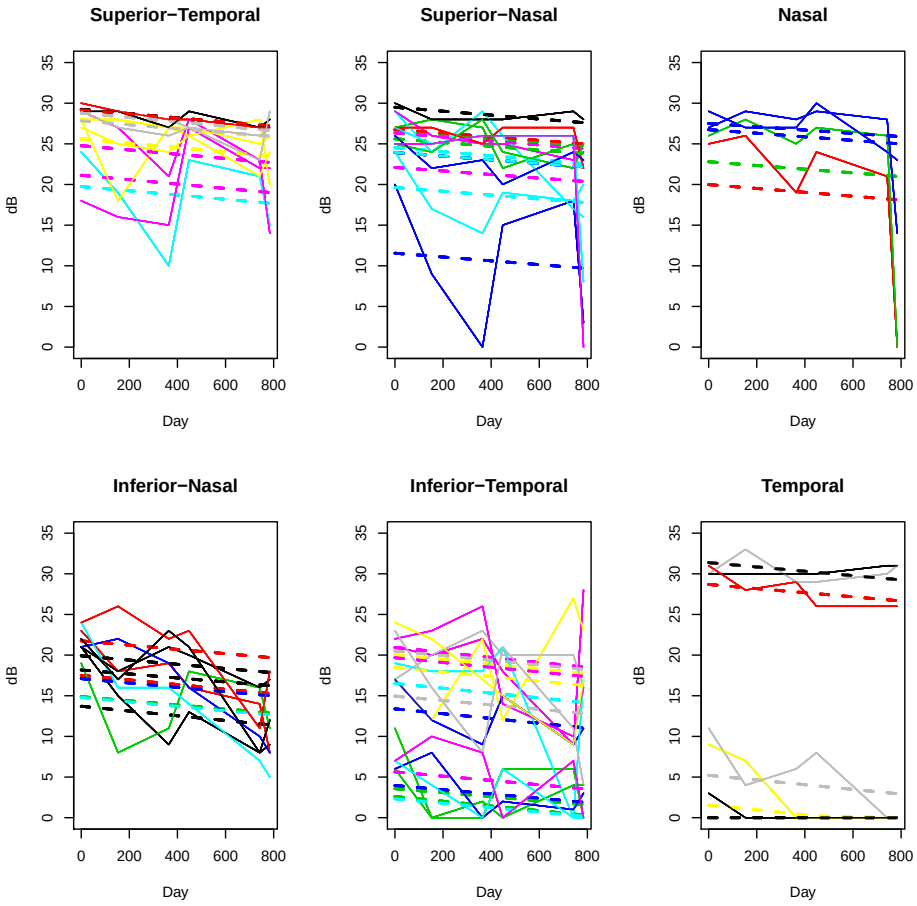


Figure 4.3: Observed versus predicted sensitivities of the 52 analysed loci by sector of the visual field for a progressing eye. Observed and predicted lines are colour matched for each locus within each sector. Solid lines depict observed data and dashed lines represent fitted values from the SPROG model.

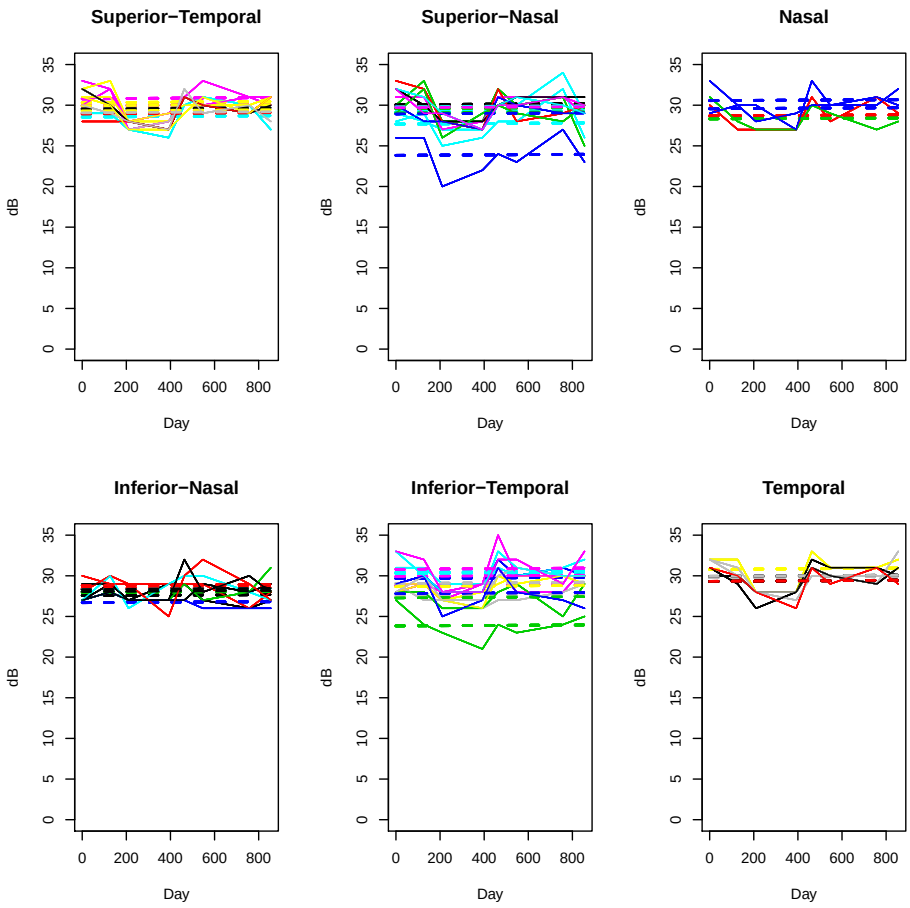


Figure 4.4: Observed versus predicted sensitivities of the 52 analysed loci by sector of the visual field for a stable eye. Observed and predicted lines are colour matched for each locus within each sector. Solid lines depict observed data and dashed lines represent fitted values from the SPROG model.

4.4.4 Heat Maps

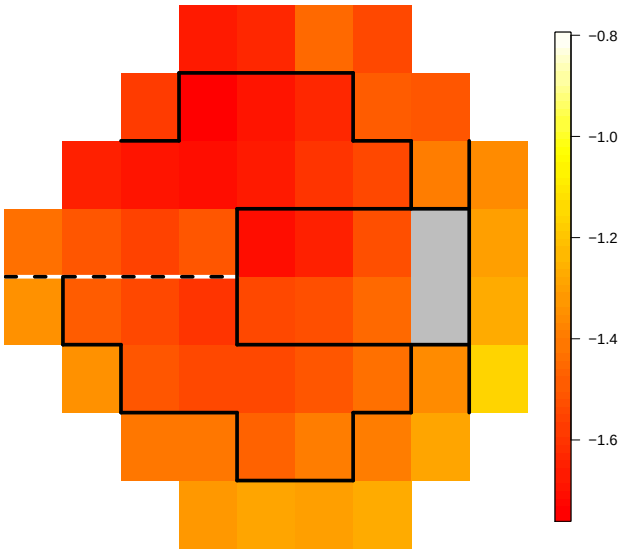
The spatial nature of the visual field can be seen in Figure 4.5, where the individual locus slopes ($\beta + \eta_i$) are presented by sector. This eye in (a) is progressing at an average rate of -1.5 dB per year. As expected, neighbouring loci are more similar than loci further apart. By plotting the rates of progression over the eye the clinician is able to see which loci or regions within the eye are progressing the fastest. In Figure 4.5 (b) the eye is progressing at an average rate of -1.1 dB per year. By examining the plot we can see that the temporal sector is progressing at a higher rate than the remainder of the eye (-1.23 dB per year).

4.4.5 Clinical Judgement of Progression

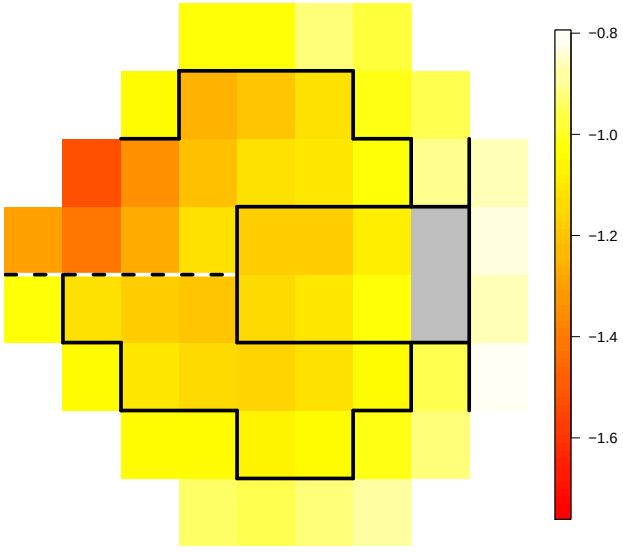
In the absence of a gold standard we compare our model with a clinical reference. Visual fields were evaluated by two independent clinicians using only the visual field information. A third clinician adjudicated any discrepancies. There were 13 discrepancies out of the 194 eyes evaluated. The expert clinicians were given access to the overview print format of the Humphrey 24-2 fields and Glaucoma Progression Analysis printout and asked to make a judgement as to whether the field was definitely progressing or not. Currently this is considered the mod acceptable standard to compare new methods to (Heijl et al. 2008). No specific guidelines or criteria were given to the clinicians. A progression rate of 26% (50/194) was diagnosed by the clinical experts. The same rate of progression (45/176) was seen when the discrepant eyes were removed from the data set.

4.4.6 Alternative Weighting Scheme

Initially our weight choice was based on clinical expertise. We investigated whether different weights for between and within sector correlations would optimize determining progression. This was carried out by running our model with different intra-sector weights and using the Receiver Operating Characteristics (ROC) when compared to the clinical reference. Briefly, ROC curves are plotted by comparing the fraction of true positives over actual positives (Sensitivity), to the fraction of false positives over actual negatives (1 - Specificity). The level of significance for classification of progression is adjusted to calculate these values across the



(a) Heat map of progressing eye



(b) Heat map of progressing inferior−temporal sector

Figure 4.5: Heat maps of trends in a progressing eye (a) and an eye showing progression in one sector (b). Rates are in dB/year.

full range of probabilities. In our case the actual positives and negatives are given by the clinical reference (see Section 4.4.5). The true and false positives are calculated by varying the significance level for classifying progression for parameter $\beta$for the various weight schemes.

Figure 4.6 shows the ROC curves for various weights and the area under the curve (AUC) is presented in Table 4.3. It can be seen from these results that the optimal intra-sector weight is 0.3. Therefore, according to our adjacency matrix described earlier, $w_{ij} = 1$ if i, j are adjacent and of the same sector, $w_{ij} = 0.3$ if i, j are adjacent and not of the same sector, and $w_{ij} = 0$ if they are non-adjacent.

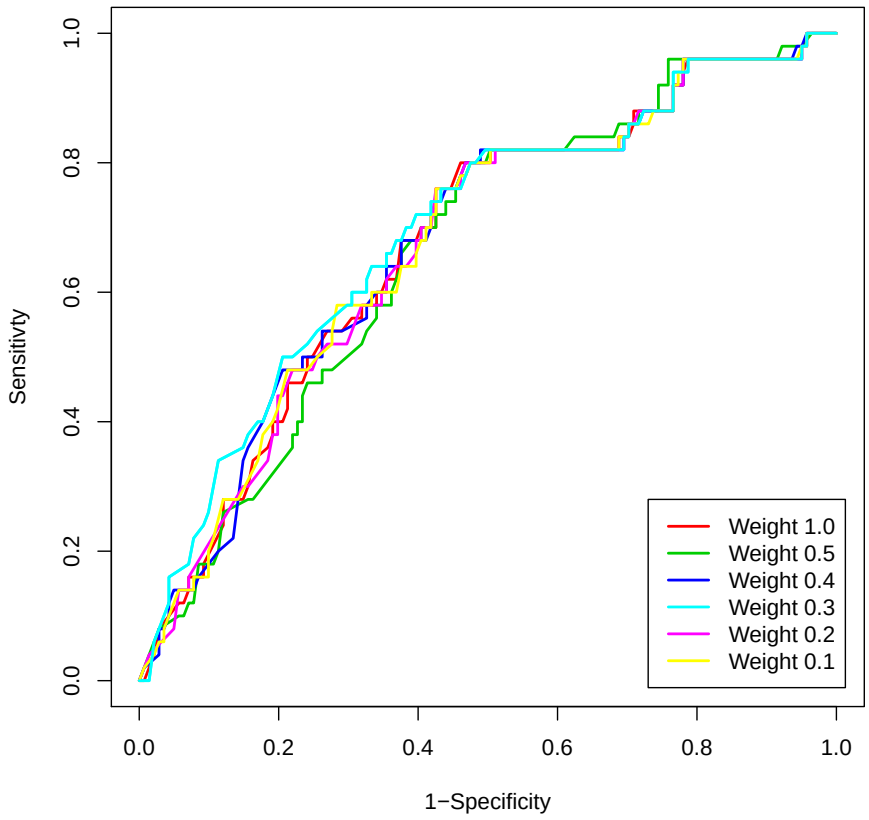


Figure 4.6: Receiver operating characteristic (ROC) curve for CAR spatial correlation weighting schemes compared to the clinical reference.

Table 4.3: Area under the curve for CAR spatial correlation weighting schemes compared to the clinical reference.

Weight	AUC
0.1	0.6760
0.2	0.6735
0.3	0.6934
0.4	0.6779
0.5	0.6661
1.0	0.6754

4.4.7 Comparison to Current PLR methods

To compare our model to the currently used PLR methods, we used the lm function in R (R Core Team 2013) for each of the 52 evaluated loci on the 24-2 field. No outliers were removed. Progression was evaluated at the final time point based on the values of the slope and their significance (p-values). Progression was judged using several criteria as defined as in Table 3.7. Our results may differ slightly from Fitzke et al. (1996), Viswanathan et al. (1997) and Birch et al. (1995) as PLR was used instead of the PROGRESSOR software. The ranking of these various methods can be seen in Figure 4.7.

PLR was then carried out within a Bayesian framework with flat priors in order to calculate the DIC to test model fit and provide a direct comparison to our new method. Observing pair-wise DIC results for individual eyes, our method SPROG, was preferred to PLR in 97% of cases, indicating SPROG better represents the observed data. The methods ranged from predicting $0\% - 75\%$ progression. The mean progression for the methods tested was 30%. Our method sits above the mean and ranks 8th out of 12 in terms of the number of eyes it predicts as progressing. Our method was one of the two closet methods to the clinical reference in terms of overall progression rate, and was by far the closest amongst the methods that did not under-diagnose in comparison to the clinical reference.

Figure 4.8 illustrates the difference between PLR and SPROG. In (a) the spatial correlation smooths the visual field, showing progression in all sectors of the visual field. In (b) the PLR method shows erratic estimates of the slopes which do not support the fact that if one loci is

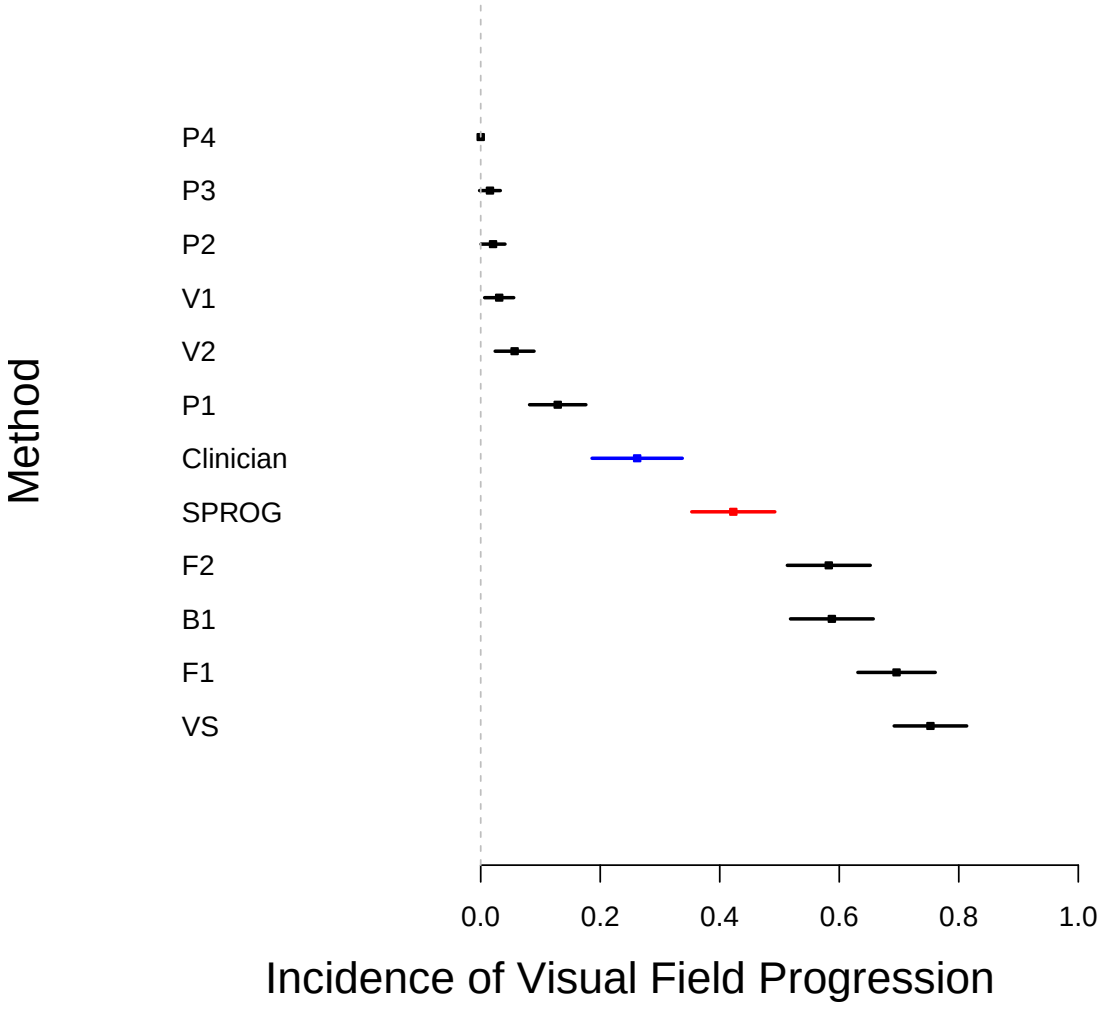


Figure 4.7: A ranking comparing PLR methods and SPROG (our Bayesian CAR approach) to determine visual field progression.

progressing, loci nearby are more likely to be progressing (Heijl et al. 1989). These maps can be compared to the observed data seen in Figure 4.3. SPROG shows a mean progression rate of -1.48 (range -1.15,-1.75). PLR produces a reasonably comparable estimate of the average progression rate at -2.13 (taken as the median of 52 local slopes, so as to control for extreme values), but gives much greater local variation (range -7.87, 4.77). One consequence of the erratic behaviour of the PLR method is that it indicates very significant vision improvement in some regions of the eye (where the slopes are positive), an implausible finding that is not seen when using the SPROG methodology. Figure 4.9 compares SPROG and PLR for an eye with stable disease. SPROG shows the stable eye has a mean slope of 0.09 (range 0.07-0.11). Again PLR shows a similar average progression rate at 0.06, but with a larger range (-1.25, 1.53).

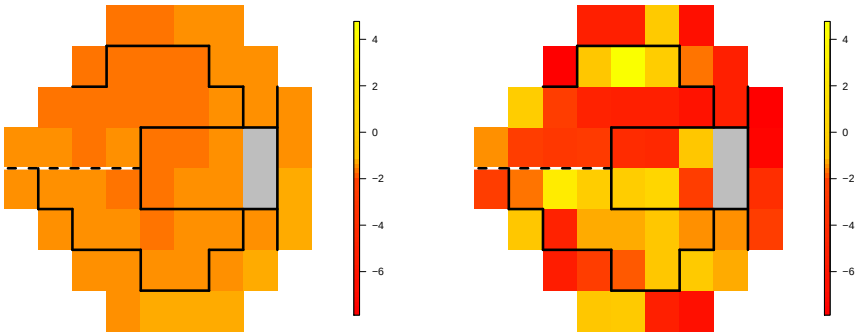


Figure 4.8: Heat maps showing slope parameters for a single progressing eye over the 52 evaluated loci of the visual field. Slopes show rates of sensitivity change in dB/year. Black lines separate the sectors and loci nearest the blind spot are shaded gray. A common colour scheme is applied in (a) and (b). (a) highlights the smoothed results of the SPROG method, while (b) shows the variation of modelling loci individually by the PLR method.

Sensitivity versus Specificity

Table 4.4 compares sensitivity and specificity of our method and the PLR methods compared to the clinical reference. Our model outperforms the other models in that it gives reasonable specificity and sensitivity. By default SPROG classifies an eye as progressing if $\beta < 0$ at the 5% significance level using a Bayesian p-value. We also investigated using a Bayesian

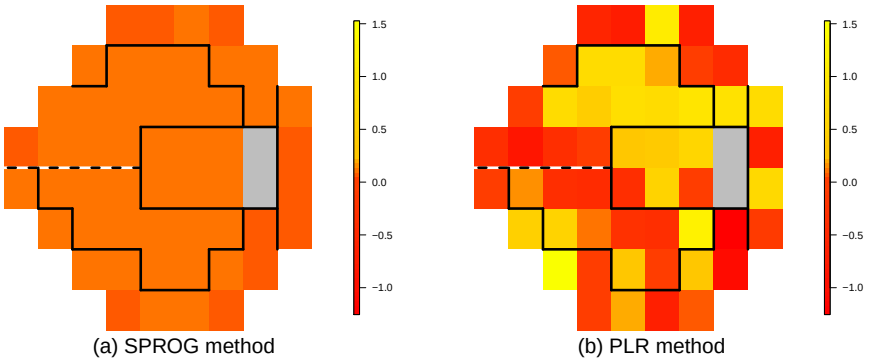


Figure 4.9: Heat maps showing slope parameters for a single eye with stable disease over the 52 evaluated loci of the visual field. Slopes show rates of sensitivity change in dB/year. Black lines separate the sectors and loci nearest the blind spot are shaded gray. A common colour scheme is applied in (a) and (b). (a) highlights the smoothed results of the SPROG method, while (b) shows the variation of modelling loci individually by the PLR method. Note that this figure has a different colour scale from Figure 4.8.

p-value of 0.1 and including a slope criterion for our model. We chose the value of -0.5 dB per year as generally any loss less than this is considered clinically significant progression. With the changed criteria SPROG still performed best. We also ran this analyses with the discrepant eyes excluded (see Section 4.4.5). Results were almost identical, and are therefore not included.

Receiver Operating Characteristic (ROC) curves are shown for SPROG and P1-P4 (Figure 4.10). These are obtained by varying the p-value thresholds, from the standard 0.05 or 0.01 for the slope parameters (see Section 4.4.6). For SPROG the true and false positives were calculated for the parameter β. For the PLR methods progression each of the 52 slopes was evaluated at each significance threshold, and the count of individual slopes classified as progressing was used to define overall progression as per the progression criterion is Table 3.7. We have not included ROC curves for methods with requirements on the numerical value of the slopes because these are not directly comparable. In particular, when the specificity is zero we will obtain a sensitivity of less than one because the slope conditions will not always be attained. Area under the curve (AUC) values, for the same methods, are presented in Table

Table 4.4: Sensitivity and specificity of our method compared with PLR methods against the clinical reference. Methods are as described in Table 3.7 and Section 4.4.5. For the SPROG model we have also looked at the effect of changing the p-value, as well as introducing a limit on the slope.

Method	Sensitivity	Specificity
$P4$	0.00	0.74
$P3$	0.04	0.99
$P2$	0.06	0.99
$V2$	0.12	1.00
$P1$	0.18	0.89
$V1$	0.18	0.99
$SPROG(p = 0.05)$	0.64	0.66
$SPROG(p = 0.1, slope < -0.5)$	0.68	0.73
$B1$	0.74	0.48
$SPROG(p = 0.1)$	0.76	0.55
$F2$	0.78	0.49
$F1$	0.86	0.36
VS	0.88	0.29

Table 4.5: Area under the curve for SPROG and PLR methods P1-P4 compared to the clinical reference.

Method	AUC	95% Confidence Intervals
P1	0.64	0.55-0.71
P2	0.67	0.57-0.75
P3	0.64	0.55-0.73
P4	0.67	0.58-0.76
SPROG	0.69	0.60-0.78

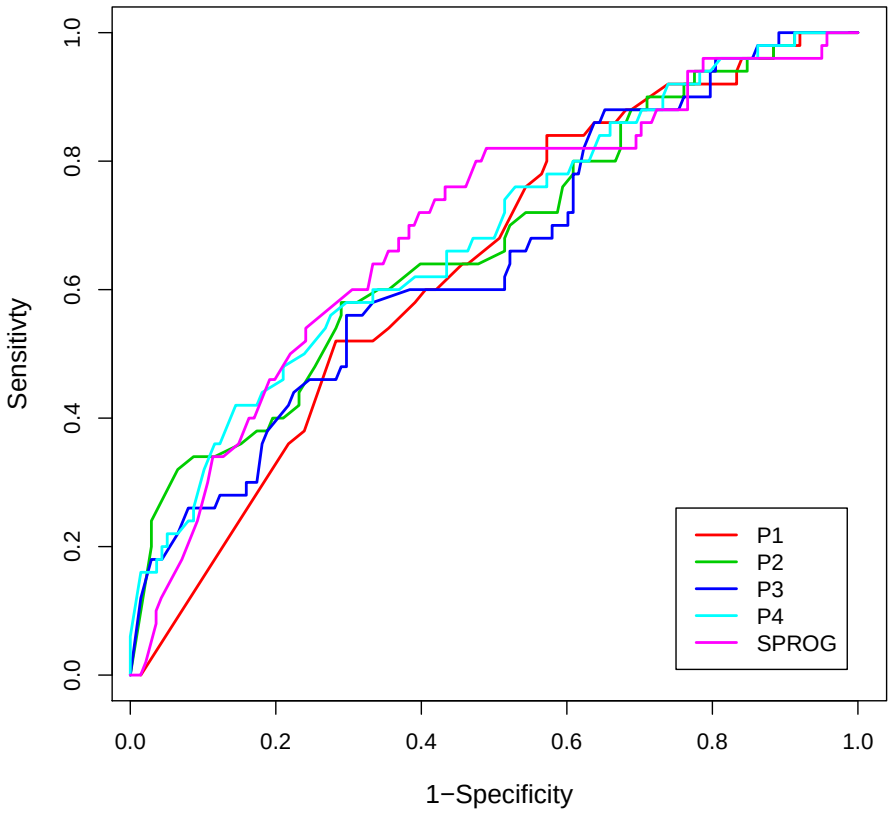


Figure 4.10: Receiver operating characteristic (ROC) curve for PLR methods and SPROG. Note that slope conditions have been excluded in producing these curves.

4.5. Confidence intervals were computed using the bootstrap percentile method. SPROG shows the largest AUC, however the confidence interval for SPROG overlap with the PLR methods. Nonetheless for the range which is most helpful to a clinician SPROG performs best. In practice a method which is operating in the middle of the ROC range is preferred. While we do not want to under diagnose we do not want to be constantly dealing with false positives.

4.5 Discussions and Conclusions

Our model provides a statistical method to model visual field progression, which takes account of spatial correlation within visual fields while respecting the relationship between the spatial arrangement of the field and the anatomy of the eye. It also provides a method which is robust to outliers, due to the pooling of strength within and between proximate sectors, and can overcome the large amounts of variation in a visual field dataset.

This model has the potential to provide more reliable results with higher specificity and sensitivity than existing methods. By borrowing information from surrounding regions, our model is able to smooth the occurrence of spikes in the data, while PLR methods struggle with the occurrence of spikes in a dataset. In contrast the PLR methods are highly influenced by outliers and identified eyes as improving that SPROG showed to be stable. Some methods (Smith et al. (1996) and Fitzke et al. (1996) see Table 3.7) even identified an eye as improving, while our method classified this eye as progressing. Criteria which require 3 consecutive fields to be significant before classifying progression under-estimate true progression. This was noted in our dataset for eyes which were classified as progressing for 2 consecutive points, stable at the next time point and then progressing for the following 2 time points. Because significant progression was not seen at the middle time point, most likely due to an outlier, this eye was not classified as progressing. For the eye seen in the figures both the clinician and SPROG identified this eye as progressing while only 50% of the PLR criteria classified this eye as progressing.

Currently our method uses a subset of our dataset to obtain a prior for measurement error and calculates the retest variation within each test. Ideally we would like to estimate measurement error from an independent dataset consisting of pairs of repeated tests carried out on two consecutive days (so that any change would not be true change, and subject fatigue would not be an issue). However because the information from the subset only contributes to the model through the specification of a prior distribution we expect any refinements to measurement error parameters to be have only a modest effect on the final results produced by SPROG.

Previous spatial modelling has focused on pre-processing the data by applying a spatial filter where a weighted average is calculated based on the surrounding loci (Fitzke et al. 1995, Spry et al. 2002, Gardiner et al. 2004). PLR was then used for classification of the eye. These methods can be regarded as an approximation to the statistically more principled SPROG method, since all in effect involve averaging over local neighbourhoods. However, our model has the additional advantage that these averages are calculated in conjunction with the other model terms (see equation 4.21) including measurement error, a temporal trend and a spatial-temporal trend. While previous spatial methods are able to account for spatial correlation and thus minimize measurement error their benefits are still confined by the use of PLR for progression classification.

We suggest that SPROG is more accurate than PLR methods which simply fit separate linear regressions of measured values (y_{ik}) over time for each location (i). PLR methods require subjective cut-offs which differ greatly in terms of what significance level is being used, the number of loci that must be seen progressing at each test, and the number of consecutive tests these loci must be identified as progressing for. This is largely due to the problem of multiple testing, 52 tests for each eye, and leads to either high false positive rates or low detection of true progression. A solution to multiple testing is to use the Bonferroni correction which maintains that in multiple test scenarios an adjustment to the significance level should be

made so that the null hypothesis has the same chance of being rejected whether testing an individual case or over multiple cases (Abdi 2007). This is done by dividing the significance level (see Table 3.7) by the number of hypotheses being tested ($n = 52$). Without the correction, when significance is evaluated at the 5% level over 52, loci we would expect 2-3 loci to show significant progression by chance; at 10% we would expect approximately 5 loci to be significant by chance.

We can compare our model to additional methods using a recent meta-analysis (Ernest et al. 2012*b*). While this does give an idea of high- and low-ranking methods it does not solve the problem of a lack of a gold standard (Ernest et al. 2012*b*). Other methods have to juggle between high sensitivity and high specificity by adjusting the various cutoffs. We believe our method reduces both false positives and false negatives by its novel approach to account for spatial correlation, thus reducing the effects of high variation which are largely responsible for false predictions. When comparing our model to the clinical reference we see that our model is the closest of the conservative methods, as well as having the best overall ROC properties. We argue that it is better to slightly over-estimate progression than under-estimate it, as doing so could lead to a delay in treatment and potential visual loss in a patient.

In addition to providing an overall estimate of whether the eye is progressing, our method also estimates the rate at which the eye is progressing. This provides the clinician with more information and can assist when deciding how aggressively to treat the glaucoma. Furthermore our model can provide estimated rates of progression for individual loci as well as subregions of the eye such as the sectors in Figure 3.4 or by hemifield. Another advantage of our method is that only 2 visual fields are required to run the model. However, it should be noted that with data from fewer time points there is considerable uncertainty around parameter values, and the ability to identify marginal cases of progression is reduced. PLR methods require a minimum of 3 visual fields for the model, and many criteria require significant results from 3 consecutive tests, which means that 5 visual fields must be completed before an eye can be classified as progressing. Given that in clinical practice on average 1.7 visual field tests are

taken in a year (Ernest et al. 2012*b*) this could lead to a long delay before an eye is shown to be progressing.

Another refinement to consider would be how the spatial correlation is modelled. The CAR model uses an adjacency matrix to define spatial correlation. This method of modelling spatial correlation was found to perform better and allow more flexibility than the alternative of accounting for spatial correlation by distance using parametric functions Best et al. (2005). However, the method of modelling spatial correlation is particularly important when considering defects that straddle the sector boundaries defined by Garway-Heath et al. (2000). While the CAR model allows for correlation between the sectors, we have placed less weight on correlations between sectors than within. The weighting 0.3 for between sectors was chosen as it gave the best combination of sensitivity and specificity overall the eyes. For eyes with defects straddling the sectors it may be of benefit to use a more even weighting scheme. Investigating modelling spatial correlation via a distance based method may prove worthwhile.

We believe that our method could be a valuable tool for integration into the clinical environment. Our method minimizes noise with a better combination of sensitivity and specificity. The heat maps present the slopes graphically for each location in a manner that makes sense to the clinician.

4.6 Chapter Summary

In this chapter we introduced disease mapping techniques. We explained how these techniques can be adapted to apply to visual field data. A CAR model was chosen as the best method to model spatial correlation. By using CAR priors we have modelled the spatial correlation in the eye. Combining this with physiological information we are able to provide a novel method for VF analysis. Because our model intrinsically accounts for the large variation of VF data, by adjusting for spatial correlation, the effects of outliers are minimized, and spurious trends are avoided. Model diagnostics, sensitivity and specificity show our model to be apparently

superior to current point-wise linear regression methods.

Chapter 5

Spline Modelling of the Progression of Visual Field Mean Deviation

5.1 Introduction

In this chapter we model mean deviation over series of visual field tests. As discussed in Section 3.2.1, mean deviation is the average difference of the 52 visual field loci thresholds compared to normal age matched controls. Mean deviation gives an indication of the overall sensitivity of the eye, as opposed to the localised defects identifiable through the individual loci threshold values. It is possible that very long series of visual fields may not follow a linear trend. While we choose to model the trend of VF thresholds using a linear trend in our SPROG model (Section 3.6.4), we acknowledge that in some cases, particularly for longer series of data, there may be some departure from linearity. Splines are a natural method to consider when modelling non-linear trends in time series data. However, they can allow too much flexibility in some cases, particularly when the trend is expected to be monotonic. In our application decreasing monotonicity is expected as we do not expect visual function to improve. Accordingly we selected 18 eyes, from 9 patients, whom had a minimum of 20 VFs taken. Both unconstrained and monotonic splines are fitted within a Bayesian framework and compared to asses the monotonicity of a mean deviation sequence.

We begin this chapter by providing a brief review of smoothing techniques, more specifically penalized spline smoothing. We then discuss how these are applied to our dataset and how we can test whether a spline is monotonic. We then provide our results, discussions and conclusions.

5.2 Modelling Background

In practice, series of data exist in which the mean function curve is more complex than can be represented by a global analytic function, such as a linear or polynomial regression. Therefore we consider spline smoothing which allows for the use of local or piecewise polynomial representations, while still maintaining the useful mathematical properties of the traditional global polynomial regression. The allowance for local representations permits the modelling of multiple or changing trends within one function.

The flexibility splines allow in modelling the shapes of relationships between variables has contributed to their extensive use in various medical applications. Splines have been used to model breast cancer survival data (Gray 1992), where the impact of covariates tumour size and number of nodes were investigated. Greenland (1996) highlighted the advantages of penalized splines when investigating the HIV incidence rate by subgroup. More recently splines have been used in the mapping of genetic quantitative trait locus (Das et al. 2012).

In this section we review the use of splines in ophthalmic literature. We then provide a brief theoretical background on spline models and their implementation in the mixed model context. As a person's vision cannot improve after glaucomatous damage occurs, we also discuss the idea of constraining splines to be monotonic.

5.2.1 Splines in Ophthalmology

Splines have been applied to ophthalmic data in several instances over the last 14 years. The earliest case of splines being applied to glaucoma data in literature is by authors Ramirez

et al. (1999). They applied cubic B-splines as a smoothing technique within a stereo vision model to investigate 3D tomography of the optic disc. Since then splines have been used in other medical imaging applications for glaucoma data (Sanfilippo et al. 2009, Knighton & Gregori 2013). Sanfilippo et al. (2009) use thin plate splines to investigate optic disc morphology. Knighton & Gregori (2013) use 2D penalised regression splines to model the thickness of the ganglion cell layer and the inner plexiform layer. From this they were able to calculate information about 'size, shape, and slope of the edge of' both the foveal depression and the macular ridge. Mederios, Zangwill, Bowd, Mansouri & Weinreb (2012) investigated the structure-function relationship between mean deviation and number of retinal ganglion cells in glaucoma patients. They used cubic smoothing splines useful to model the non-linear relationship between the variables, and suggest that interpretation of the slopes is dependent on the stage of disease.

5.2.2 Introduction to Splines

The following is a brief introduction to splines. For a more thorough overview consult texts such as Ruppert et al. (2003) and Fahrmeir & Knieb (2011).

Splines are a series of piecewise polynomials that satisfy continuity constraints where the neighbouring pieces are joined. Practically this involves partitioning the x-axis data into M contiguous intervals, bounded by $M + 1$ knots. For the continuous variable x_i the knots are:

$$x_{min} = \kappa_0 < \kappa_1 \ldots < \kappa_{M-1} < \kappa_M = x_{max}. \tag{5.1}$$

If the function $f(x)$ is to be defined as a spline of pth order it must meet the following conditions:

- $f(x)$ is a polynomial of degree $\leq p$ for $x \in [\kappa_m, \kappa_{m+1})$ where $m = 0, \ldots, M-1$ and
- $f(x)$ is $(p-1)$ times continuously differentiable at $x = \kappa_m$ for $m = 1, \ldots, M-1$.

The common splines are the linear spline where $p = 1$, quadratic ($p = 2$), and the cubic spline ($p = 3$). The linear spline, often referred to as the 'broken stick' model consists of line

segments joined at knots (see Figure 5.1) and can be written as:

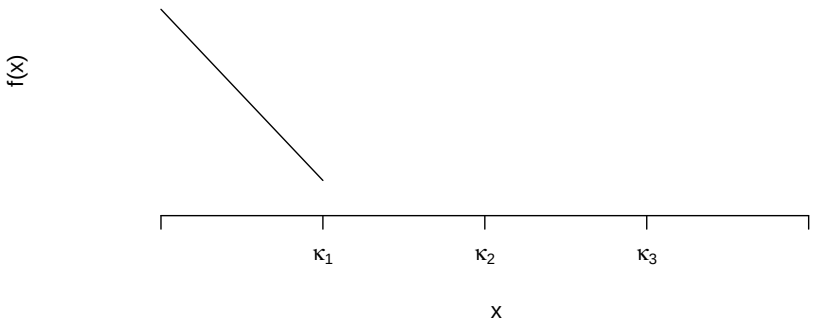


Figure 5.1: Linear spline with three knots.

$$f(x) = \beta_0 + \beta_1 x + \sum_{m=1}^{M} b_m (x - \kappa_m)_+. \tag{5.2}$$

To the left of the first knot (κ_1) the equation of the line is $f(x) = \beta_0 + \beta_1 x$. Accordingly the parameter b_m adjusts the slope for segment between κ_m and κ_{m+1}. Mathematically this is denoted by subscript + which represents the function:

$$z_+ = \begin{cases} z & z \geq 0 \\ 0 & z < 0 \end{cases}$$

where negative values are replaced by zero. Thus the coefficient of x between κ_m and κ_{m+1} is $\beta_1 + \sum_{l=1}^{m} b_l$.

Equation 5.2 is naturally extended to a quadratic polynomial spline:

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \sum_{m=1}^{M} b_m (x - \kappa_m)_+^2, \tag{5.3}$$

and a pth order spline:

$$f(x) \equiv f(x; \boldsymbol{\beta}, \boldsymbol{b}) = \beta_0 + \beta_1 x + \cdots + \beta_p x^p + \sum_{m=1}^{M} b_m (x - \kappa_m)_+^p, \tag{5.4}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)$ and $\boldsymbol{b} = (b_1, b_2, \ldots, b_M)$. We consider the knot placement to be fixed and so the dependence of f on κ in Equation 5.4 is suppressed in our notation.

Splines provide an attractive and flexible way of describing a complex trend function. Consider the non-parametric regression model:

$$y_i = f(x_i) + \epsilon_i, \tag{5.5}$$

where y_i is the mean deviation at time point i and $f(x_i)$ is a function of time, x, in years, which we model as a spline. As the goal with applying a spline is to smooth the data (y_i), we can optimize the amount of smoothing by minimizing the sum of squares and including a penalty for additional 'roughness'. Equation 5.6 balances the closeness of the model fit (seen in term one), and the smoothness of the fit through the roughness penalty (term two) (Fahrmeir & Knieb 2011). Therefore fitted values $\boldsymbol{\beta}$ and $\boldsymbol{b}$ minimize the criterion:

$$Q(\boldsymbol{\beta}, \boldsymbol{b}) = \sum_{i=1}^{n} \{y_i - f(x_i; \boldsymbol{\beta}, \boldsymbol{b})\}^2 + \alpha \int \{f''(x; \boldsymbol{\beta}, \boldsymbol{b})\}^2 dx \tag{5.6}$$

where α is a tuning parameter controlling the importance of fits versus smoothness. Note that the second term which controls the smoothness is based on the second derivative and hence measures the rate of change of gradient f.

So far we have discussed pth degree splines in general terms. By increasing the degree of the spline we can increase the smoothness of the function (e.g. linear, quadratic, cubic). Spline regressions can vary by the choice of knots, changing the estimation method of the roughness criterion, and constraints applied beyond the range of the data. Smoothing splines have knots at each unique x_i value. An example of a spline where constraints are placed beyond the

range of the data is the natural smoothing spline. Here the tails of the splines are constrained to be linear in the tails of the function either side of the boundary knots (Ruppert et al. 2003).

A popular choice of spline for the model in Equation 5.5 is the penalised regression spline. Penalised regression splines use a number of knots that is much less than the number of unique x_i values. Knot specification is known not to have a huge effect on the smoothness of a penalized spline (Wand 2003). Because of the smaller number of knots, penalised splines are also more computationally efficient. Therefore Wand (2003) suggests that a default rule for knot location is:

$$\kappa_m = \left(\frac{m+1}{M+1}\right) \text{th sample quantile of unique } x_i\text{'s}, \quad 1 \leq m \leq M \tag{5.7}$$

where M =min$(n/4, 35)$. While location of the knots for penalised regression splines is normally specified by quantile, some situations arise where it is more appropriate to manually specify knots (Ruppert & Carroll 2002).

Penalized splines are fitted by minimizing by the following penalized sum of squares criterion:

$$Q(\boldsymbol{\beta}, \boldsymbol{b}) = \sum_{i=1}^{n} \{y_i - f(x_i; \boldsymbol{\beta}, \boldsymbol{b})\}^2 + \alpha \sum_{m=1}^{M} b_m^2. \tag{5.8}$$

Similarly to equation 5.6, equation 5.8 also balances the goodness of fit versus the amount of smoothing through the parameter α (Ruppert et al. 2003).

5.2.3 Spline Regressions as Mixed Models

A natural way to implement penalised regression splines is within a linear mixed model framework. An identical model can be gained fitting a penalised regression spline in a mixed model, as when using the criterion in Equation 5.8 (see Speed (1991)). A penalised regression spline represented as linear mixed model is written:

$$y = X\boldsymbol{\beta} + Z\boldsymbol{b} + \epsilon \tag{5.9}$$

where

$$X = \begin{bmatrix} 1 & x_1 & \cdots & x_1^p \\ 1 & x_2 & \cdots & x_2^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_m & \cdots & x_m^p \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$$

and

$$Z = \begin{bmatrix} 1 & (x_1 - \kappa_1)_+^p & \cdots & (x_1 - \kappa_M)_+^p \\ 1 & (x_2 - \kappa_1)_+^p & \cdots & (x_2 - \kappa_M)_+^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & (x_m - \kappa_1)_+^p & \cdots & (x_m - \kappa_M)_+^p \end{bmatrix} \quad \boldsymbol{b} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_M \end{bmatrix}.$$

The vector of random effects has the following distribution $\boldsymbol{b} \sim \mathsf{Normal}(\mathbf{0}, G)$, where $G = \sigma_b^2 I$. The error term has a similar distribution $\boldsymbol{\epsilon} \sim \mathsf{Normal}(\mathbf{0}, R)$, where $R = \sigma_\epsilon^2 I$. The combined design matrix is typically denoted $C = [X, Z]$ and the covariance matrix $V = ZGZ^T + R$ (Ruppert et al. 2003).

The best linear unbiased predictors (BLUPs) of $\boldsymbol{\beta}$ and $\boldsymbol{b}$ minimise Equation 5.8 for a suitable choice of α, specifically when $\alpha = \sigma_\epsilon^2/\sigma_b^2$ (Speed 1991). Estimated BLUPs for $\boldsymbol{\beta}$ and $\boldsymbol{b}$ are defined:

$$\hat{\boldsymbol{\beta}} = (X^T \hat{V}^{-1} X)^{-1} X^T \hat{V}^{-1} \boldsymbol{y} \tag{5.10}$$

and

$$\hat{\boldsymbol{b}} = \hat{G} Z^T \hat{V}^{-1} (\boldsymbol{y} - X\hat{\boldsymbol{\beta}}). \tag{5.11}$$

where $\hat{G}$ and $\hat{V}$ approximate the covariance matrices G and V and are estimated through residual maximum likelihood estimation (REML). The covariance matrix of the estimated BLUPs is approximated by:

$$\mathsf{Var}\left(\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{b}} - \boldsymbol{b} \end{bmatrix}\right) \approx (C^T R^{-1} C + B)^{-1} \tag{5.12}$$

where

$$B = \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} \end{bmatrix}.$$

5.2.4 Monotonicity

In the context of some applications, such as growth curves, it is implausible for the mean function to decrease. Therefore when fitting a function to such a dataset monotonicity must be considered. Early contributions investigating monotonic constraints on splines include Wright & Wegman (1980) and Villalobos & Wahba (1987). More recently methods to fit monotonic splines have developed within the Bayesian framework. These include work by Holmes & Heard (2003), Neelon & Dunson (2004) and Shively et al. (2009). To enforce monotonicity the conditions on the spline coefficients $\boldsymbol{\beta}$ and $\boldsymbol{b}$ are specified. When using quadratic splines these constraints will be linear. It follows that if we employ priors that are zero whenever these constraints fail, then the resulting fit will be monotonic. To this end, Brezger & Steiner (2008) enforce the monotonicity constraint on B-splines by sampling the spline coefficients in an MCMC approach from a truncated multivariate normal distribution. Similarly Hazelton & Turlach (2011) propose a method within a semiparametic regression framework which allows for restrictions on the shape of penalized splines to be enforced by sampling from a truncated normal distribution.

In more detail, we can apply constraints on the shape of the relationship between x and y by enforcing restrictions on the coefficients ($\boldsymbol{\xi} = (\boldsymbol{\beta}^T, \boldsymbol{b}^T)^T$) of the spline f. For a linear spline to be considered monotonic non-decreasing the following requirements must hold for parameter vector for all $m = 1, \ldots, M$ $\boldsymbol{\xi} = (\beta_0, \beta_1, b, \ldots, b_M)$ for all $m = 1, \ldots, M$:

$$\beta_1 \geq 0 \tag{5.13}$$

$$\beta_1 + \sum_{i=1}^{m} b_i \geq 0. \tag{5.14}$$

Figure 5.2 shows a linear monotonic non-decreasing spline with 3 knots.

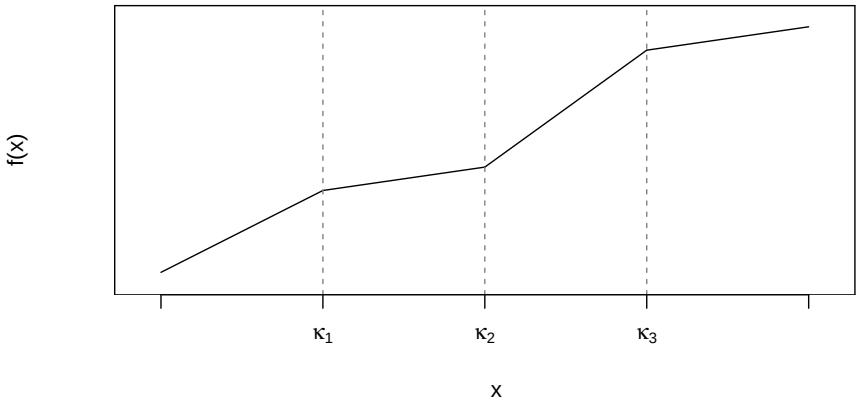


Figure 5.2: Monotonic non-decreasing linear spline with three knots.

The linear system of constraints in 5.13 and 5.14 can be written in matrix format as $A_p\xi \geq \boldsymbol{c}$, where A is an appropriate matrix of pth degree spline and $\boldsymbol{c}$ a vector, and the inequality is interpreted component-wise. For a linear spline

$$A_1 = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & 1 & \cdots & 1 \end{bmatrix}$$

and $\boldsymbol{c} = \mathbf{0}$.

Linear constraints, compared to say quadratic constraints, are very convenient as they are

computationally efficient while still allowing for the specification of monotonicity. For the modelling of many time series functions linear splines are not attractive because they are not smooth. Cubic splines require complex constraints to ensure monotonicity. This suggest we work with quadratic splines.

In the quadratic case the vector of parameters is $\boldsymbol{\xi} = (\beta_0, \beta_1, \beta_2, b, \ldots, b_M)^T$. For a non-decreasing monotonic function the following must hold:

$$\beta_1 + 2\beta_2\kappa_m + 2\sum_{k=1}^{m} b_k(\kappa_m - \kappa_k)_+ \geq 0 \tag{5.15}$$

for $m = 0, \ldots, M$. Hence in the linear inequality $\mathrm{A}\boldsymbol{\xi} \geq \mathbf{0}$ the matrix A becomes

$$A_2 = \begin{bmatrix} 0 & 1 & 2\kappa_0 & 0 & \cdots & 0 \\ 0 & 1 & 2\kappa_1 & 2(\kappa_2 - \kappa_1) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & 2\kappa_{M+1} & 2(\kappa_{M+1} - \kappa_1) & \cdots & 2(\kappa_{M+1} - \kappa_M) \end{bmatrix}.$$

The monotonic constraint discussed above is non-decreasing. Monotonic non-increasing constraints can also be fitted. This requires the reversal of the inequalities in the constraint Equations 5.13, 5.14 and 5.15. We discuss the implementation of these monotonicity constraints within a Bayesian framework as per Hazelton & Turlach (2011) in Section 5.3.3.

5.3 Application of Splines to VF Mean Deviation

We apply the methods discussed above to mean deviation scores from serial VF data. In the following section we describe the data. We then explain our model and discuss how it is implemented within a Bayesian framework.

5.3.1 Data

Our data consists of a subset of eyes from the dataset described in Chapter 3, Section 3.6. We selected a subset of 18 eyes from 9 patients, who had more than 20 measurements taken

for each eye, as a longer series is required to warrant fitting splines. The mean number of VF per subject was 23 over a period of 12 years. The mean mean deviation over all the eyes at all time points was -6.4.

5.3.2 Developing the Model

In order to investigate the shape of mean deviation over time we apply quadratic splines to series of mean deviations from individual eyes. The quadratic spline takes the form below:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \sum_{m=1}^{M} b_m (x_i - \kappa_m)_+^2 + \epsilon_i, \tag{5.16}$$

where y_i is Mean Deviation at time x_i (years). We fit our model within a linear mixed model framework as discussed in Section 5.2.3. The number of knots, κ_M, is chosen for each model by the rule of thumb $M = \min(N/4, 35)$ where N is the number of x measurements (Ruppert et al. 2003). The knots are evenly spaced by quantile from $\min(x)$ to $\max(x)$. For our data the mean number of knots for each spline function was 5, ranging from 4 to 7.

As discussed in Section 3.6.4, once a person has lost sight in a particular locus in their eye it is permanent. Therefore it is impossible for vision to improve, in principle. As we are modelling mean deviation, which is an average VF sensitivity adjusted for age (see Section 3.2.1) there may be some cases where the mean deviation increases due to stability in the VF compared to their age matched cohort. However in most cases we would not expect this to be the case. Hence we investigate the impact of enforcing a monotonic non-increasing spline on the dataset. In addition we fit an unconstrained spline to the same data to provide a comparison. The constrained spline and unconstrained spline have the same equation 5.16, however the unconstrained spline is not restricted by the monotonicity conditions laid out in Section 5.2.4. For comparison we also fit simple linear regression models.

5.3.3 Bayesian Implementation

Unconstrained penalised regression splines can be fitted using standard statistical software for fitting linear mixed models (such as nlme in R (Pinheiro et al. 2013, R Core Team 2013)). The estimates of σ_ϵ^2 and σ_b^2 are calculated through restricted maximum likelihood (REML) which automates the choice of smoothing parameter α, as discussed in Section 5.2.3. Constrained splines are much more easily handled within a Bayesian framework in order to sample from distributions specified by these constraints (Hazelton & Turlach 2011). Therefore we fit both constrained and unconstrained models within a Bayesian framework for comparability.

For our model the regression parameters are $\boldsymbol{\xi} = (\beta_0, \beta_1, \beta_2, b, \ldots, b_M)$. For the unconstrained model we employ vague priors for our model parameters:

$$\beta \sim \mathsf{Normal}(0, \sigma_\beta^2), \tag{5.17}$$

where σ_β^2 is the arbitrary large value, 25 times the estimated variance of the BLUPs, based on Equation 5.12. The random effects prior:

$$b \sim \mathsf{Normal}(0, \sigma_b^2), \tag{5.18}$$

is determined by random effects variance (σ_b^2) which for computational convenience is fixed at its REML value calculated from the mixed model. See Hazelton & Turlach (2011). The unconstrained model parameters can be sampled using Gibbs sampling.

For the constrained model, we employ truncated normal priors to enforce monotonicity. Specifically we follow Hazelton & Turlach (2011), and fit the model using an MCMC independence sampler. We want to employ a proposal distribution that reflects the posterior. This is achieved by fitting the unconstrained model using standard mixed model methods (using nlme (Pinheiro et al. 2013)). If we specify a multivariate normal proposal distribution centred on the parameter estimates, with a covariance matrix as defined by the variance-covariance

matrix of these estimates (see Equation 5.10-5.12), this will be approximately equal to the posterior distribution for the unconstrained model. By employing a truncated version of this distribution we hope to approximate the posterior for the constrained model. Therefore the candidate values $\boldsymbol{\xi}$ for the constrained model are sampled from truncated multivariate normal proposal distributions (Hazelton & Turlach 2011), where the monotonic non-deceasing spline function is subject the constraints specified in Equation 5.15.

It follows that the truncated normal proposal distribution is based on the estimated BLUPs $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{b}}$:

$$q = \mathsf{TN}_{\phi}(\hat{\boldsymbol{\xi}}, (C^T R^{-1} C + B)^{-1}), \tag{5.19}$$

where $\mathsf{TN}_{\phi}(\boldsymbol{\mu}, \Sigma)$ is a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix Σ, constrained so that $\phi \equiv \phi_{A,\boldsymbol{c}} = \{\boldsymbol{x} : A\boldsymbol{x} \geq \boldsymbol{c}\}$; $\hat{\boldsymbol{\xi}} = (\hat{\boldsymbol{\beta}}^T, \hat{\boldsymbol{b}}^T)^T$. The remaining terms are as defined in Section 5.2.3.

We implemented the following independence sampler using block updating as per Hazelton & Turlach (2011):

1. Initialize vectors $\boldsymbol{\xi}^0$ and ${\boldsymbol{\sigma}_{\boldsymbol{\epsilon}}}^0$. Set $t = 1$.
2. Generate candidate $\boldsymbol{\xi}^* \sim q$.
3. Calculate the acceptance probability using
   $\pi = \min\left[1, \frac{q(\xi^{t-1}) f(\boldsymbol{y}|\boldsymbol{\xi}^*, \boldsymbol{\tau}^{t-1}) f(\boldsymbol{b}_1^*|(\sigma_b^2)^{t-1}) f(\beta_1^*)}{q(\xi^*) f(\boldsymbol{y}|\boldsymbol{\xi}^{t-1}, \boldsymbol{\tau}^{t-1}) f(\boldsymbol{b}_1^{t-1}|(\sigma_b^2)^{t-1}) f(\beta_1^{t-1})}\right]$.
   Set $\boldsymbol{\xi}^t \leftarrow \boldsymbol{\xi}^*$ with probability π; otherwise $\boldsymbol{\xi}^t \leftarrow \boldsymbol{\xi}^{t-1}$.
4. Sample ${\sigma_{\epsilon}}^t$ conditional on $\boldsymbol{\xi}^t$, using the Metropolis-Hastings algorithm.
5. Set $t \leftarrow t + 1$. Repeat steps 2-5 until required number of iterations complete.

In addition to modelling a constrained and unconstrained spline, we implement linear regression model within a Bayesian context using an uninformative Cauchy distributions for the regression parameters and a vague inverse gamma distribution for the error.

5.3.4 Testing for Monotonicity

We extend the work of Hazelton & Turlach (2011) by testing for monotonicity by creating a Bayesian p-value. We begin by fitting an unconstrained model. The posterior probability that f is monotonic is $\mathsf{Pr}(f \text{ monotonic}|\text{data})$. Therefore, to test for the assumption of monotonicity $1-\mathsf{Pr}(f \text{ monotonic}|\text{data})$ is the Bayesian p-value. If f is a monotonic non-increasing function, $A\boldsymbol{\xi} \leq 0$, therefore the p-value for whether the function is monotonic or not becomes $\mathsf{Pr}(A\boldsymbol{\xi} \leq 0|\text{data})$. We then check this condition at each iteration of our MCMC algorithm, and hence approximate this posterior probability by the proportion of sampled $\boldsymbol{\xi}$ for which it holds. Hence $p = 1 - \frac{1}{N}\sum_{t=1}^{N} I(A\boldsymbol{\xi}^t \leq \mathbf{0})$ where N is the number of MCMC samples obtained.

5.4 Results and Discussion

Models were run on all 18 eyes. Simple linear regression was carried out for each model to identify outliers. Data-points with residuals > 4 were classified as outliers and removed. As mentioned in Section 3.4, variability due to phenomenon such as the learning effect may cause outliers in the dataset. In addition 5 other outliers were identified visually and removed. Convergence was observed for all 18 models using the Geweke Diagnostic (Geweke 1992), and acceptance rates were $\approx 50\%$. Figure 5.3 shows converging trace plots for an example eye. All plots are found in Appendix D. DIC was used to test whether linear regression (fitted using MCMC) or a spline model was preferred. A Bayesian p-value of $p < 0.05$ rejected the assumption of monotonicity, and showed preference for the unconstrained spline.

As expected the linear model was sufficient to describe the shape in the majority of cases (10/18). Of the 8 models in which splines provided a better fit 3 held to the assumption of monotonic-decreasing, while the remaining 5 preferred the unconstrained spline. Figure 5.4 shows the model fits for three eyes. We present an example of a linear fit (a), a monotonic fit (b), and an unconstrained fit (c). We also present a fourth case where the best fit is not as clear. Point-wise Bayesian 95% credible intervals were calculated for each model, and these are shown on the plots shaded in grey.

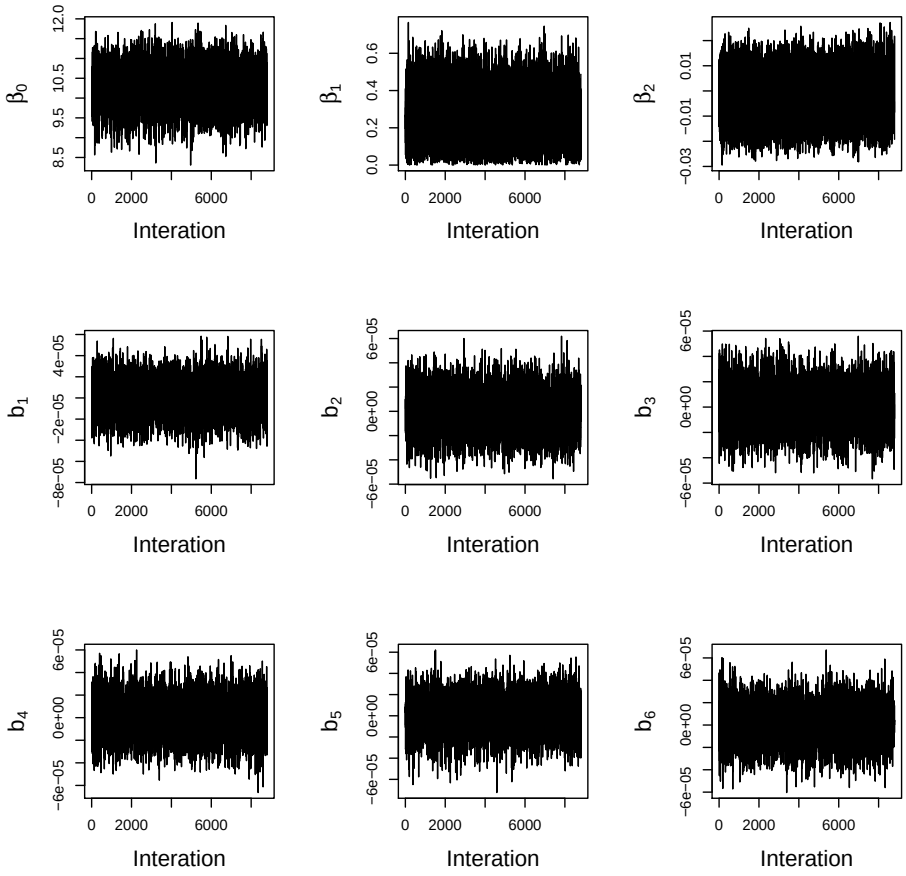
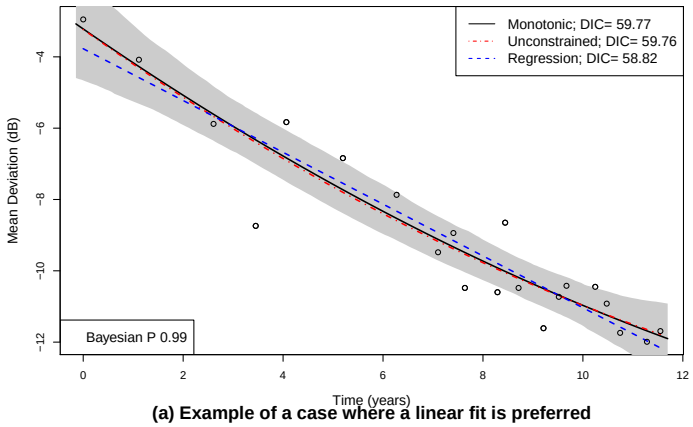
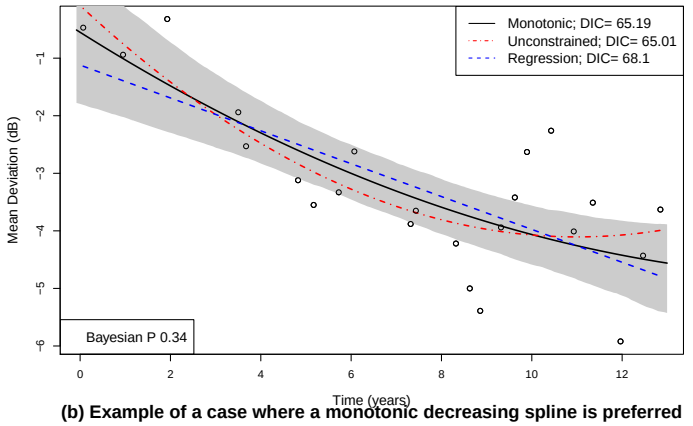


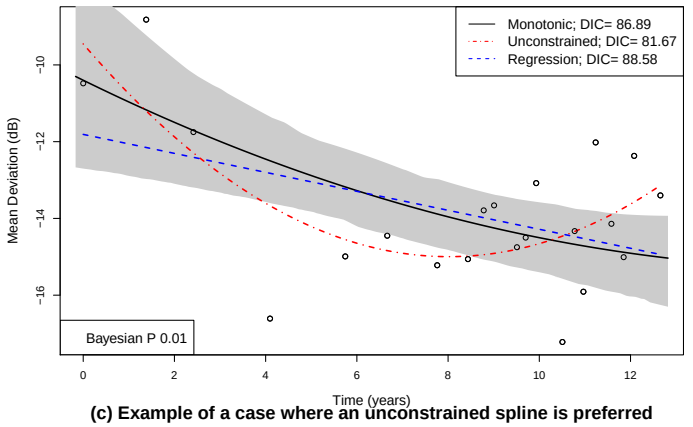
Figure 5.3: Trace plots for the 9 parameters for a constrained model of an example eye.



(a) Example of a case where a linear fit is preferred



(b) Example of a case where a monotonic decreasing spline is preferred



(c) Example of a case where an unconstrained spline is preferred

Figure 5.4: Examples of mean deviation over time which are best modelled by (a) linear regression, (b) monotonic spline and (c) unconstrained spline fits. Each figure shows the fits of all 3 models with the 95% Bayesian credible interval for the constrained model shown in grey. DICs for the models are given in the legend and the p-value for monotonicity shown in the left corner of the plot.

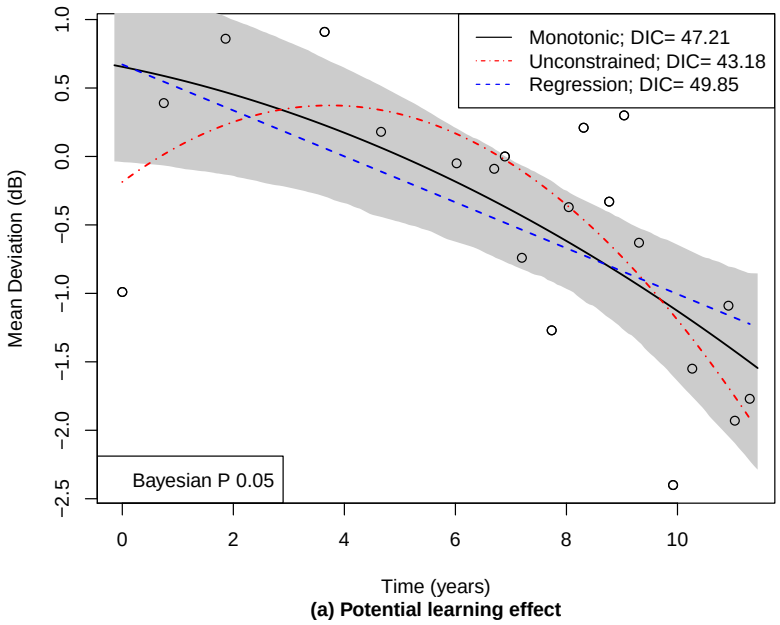
In Figure 5.5 (a) the DIC favours the unconstrained model, however $p = 0.07 > 0.05$. Looking at the plot we hypothesize that this may be due to the learning effect (see Section 3.4). When the initial point is removed and the curves are re-fitted, Figure 5.5 (b) shows that the DIC no longer favours the unconstrained spline. The assumption on monotonicity holds with ($P = 0.42$) and the lowest DIC value.

5.4.1 Post-hoc Analysis

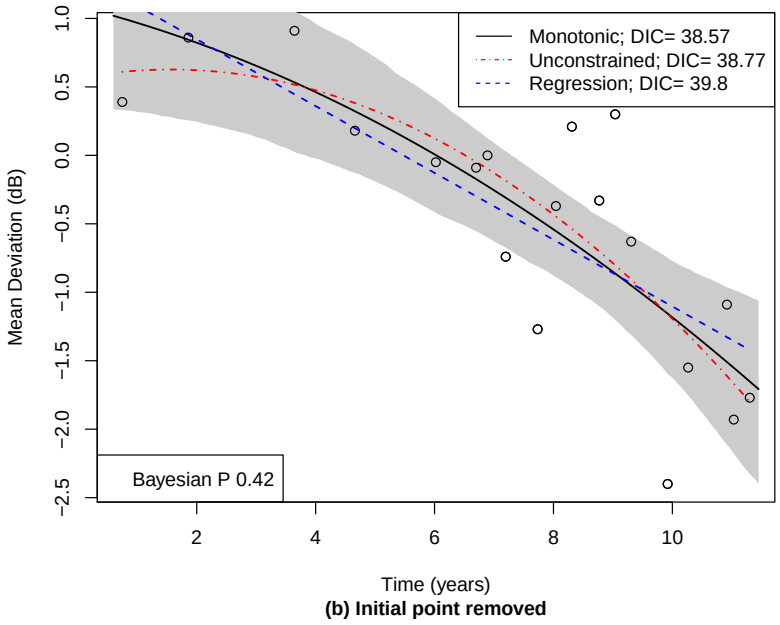
After fitting the models, a post-hoc analysis was carried out on the subjects who were best modelled by an unconstrained spline. The idea was to investigate whether any treatment or other circumstances may explain the unexpected behaviour of the curve. For one subject the curve shape is explained by the learning effect, and as above once the initial points are removed a linear relationship best describes the data. The remaining four cases are explored below.

Figure 5.6 shows a convex curve for the unconstrained curve. Post-hoc analysis shows that the subject received a needling of their left trabeculectomy bleb at 6.4 years after the initial field test. This is marked on the plot by the green vertical line. Additionally their medication was changed at this time point. The subject's IOP dropped from 19 to 12, and an IOP of 12 has remained since this intervention. It appears that the MD observations following this time point are influencing behaviour of this curve. While there is still a lot of fluctuation post intervention the trend is much flatter, indicating the field loss is stable. The large fluctuation in MD can be attributed to this subject's VF loss which is mainly central (close to fixation), and know to have a large amount of variability.

In the second case (Figure 5.7 (a)) the subject had a 'moderately dense' cataract removed 8.9 years after the initial VF. The convex curve can be explained by the decrease in MD as the cataract worsened, and then the increase in MD seen after the removal. In Figure 5.7 we look at the data prior to the surgery and we see a more typical shape relationship, which is adequately modelled by linear regression.



(a) Potential learning effect



(b) Initial point removed

Figure 5.5: Plots show linear regression, monotonic and unconstrained spline curves of mean deviation over time for an eye with a potential learning effect. 95% Bayesian credible interval for the constrained model shown in grey. DICs for the models are given in the legend and the p-value for monotonicity shown in the left corner of the plot.

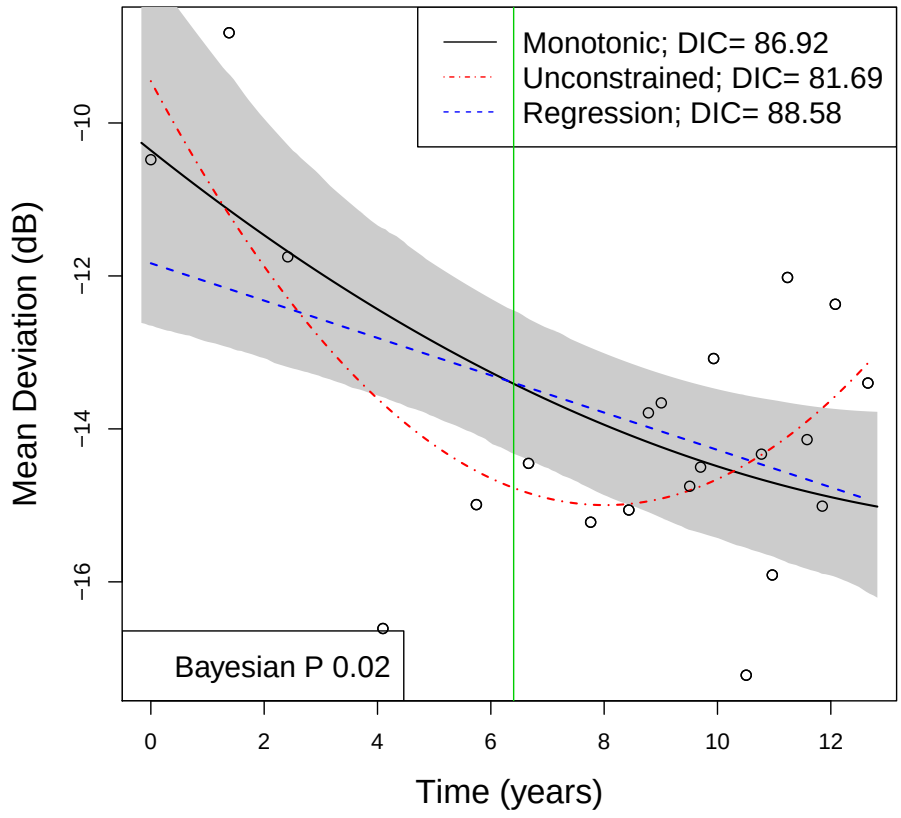


Figure 5.6: Spline curves for subject 1. The vertical green line indicates a clinical intervention.

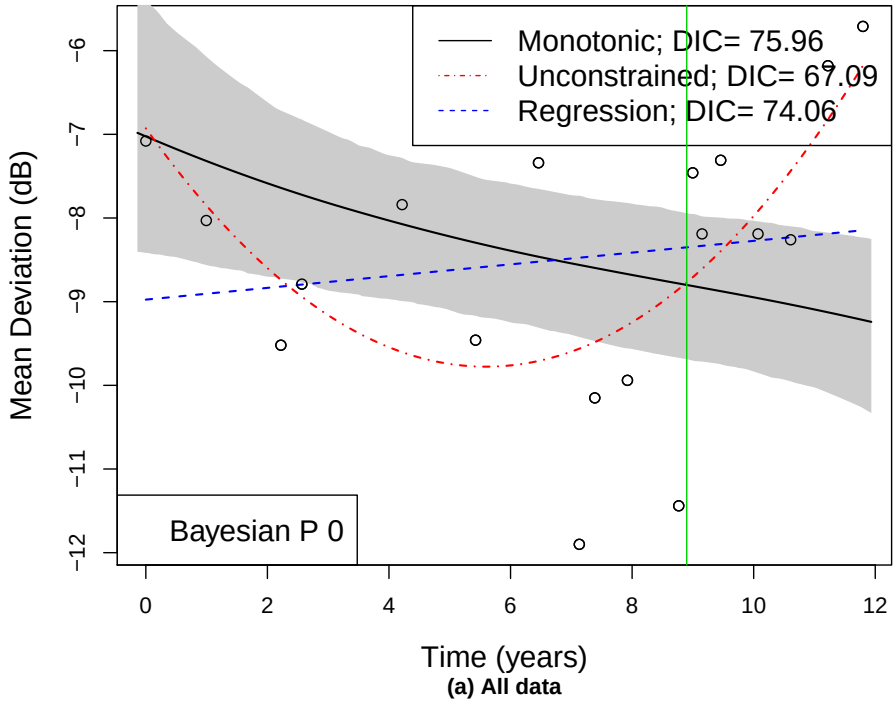
In Figure 5.8 we see a sigmoidal curve, which is highly influenced by two points at about 5 and 6 years post initial VF. There is also much variation seen in this subject's other eye. This may be due to the subject's alcohol dependency.

Figure 5.9 shows a sigmoidal curve. This subject has Alzheimer's disease which may explain some of the fluctuation of MD seen. In addition they had a fall and fracture around 4 years post initial test which may also contribute to their transient visual field performance.

5.5 Conclusions

In Chapter 4 our goal was to predict progression. However techniques in this chapter are not suited for predicting progression as we use local smoothing methods. Instead we propose this method as a screening and retrospective analysis tool. This methodology could be applied routinely in clinical practise, with the main benefit being that it would flag subjects which require further attention. The model is efficient enough to run within a clinical visit. This would be of particular benefit if the ophthalmologist was not familiar with the subject's history. Alternatively it could be run on all subject records within a clinic, flagging those which the ophthalmologist may wish to recall for follow up testing.

When the unconstrained model is preferred, the unusual shape appears to be explained by external circumstances. The concave curves were seen in patients where the fitted curve appeared to be influenced by lower initial points indicating a potential learning effect. A convex curve was seen when an intervention occurred at some point during the treatment, resulting in an 'improvement' in or a return to pre-condition mean deviation. A sigmoidal curve was viewed in subjects with a lot of fluctuation in their mean deviation which could not be attributed to any particular eye phenomenon.



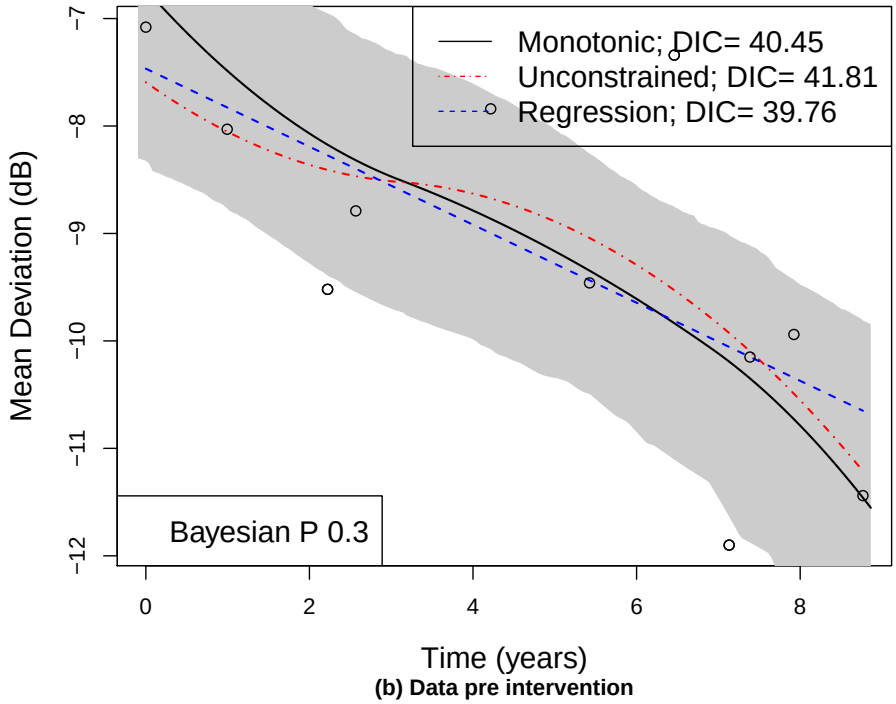


Figure 5.7: Spline curves for subject 2. The vertical green line in (a) indicates a clinical intervention.

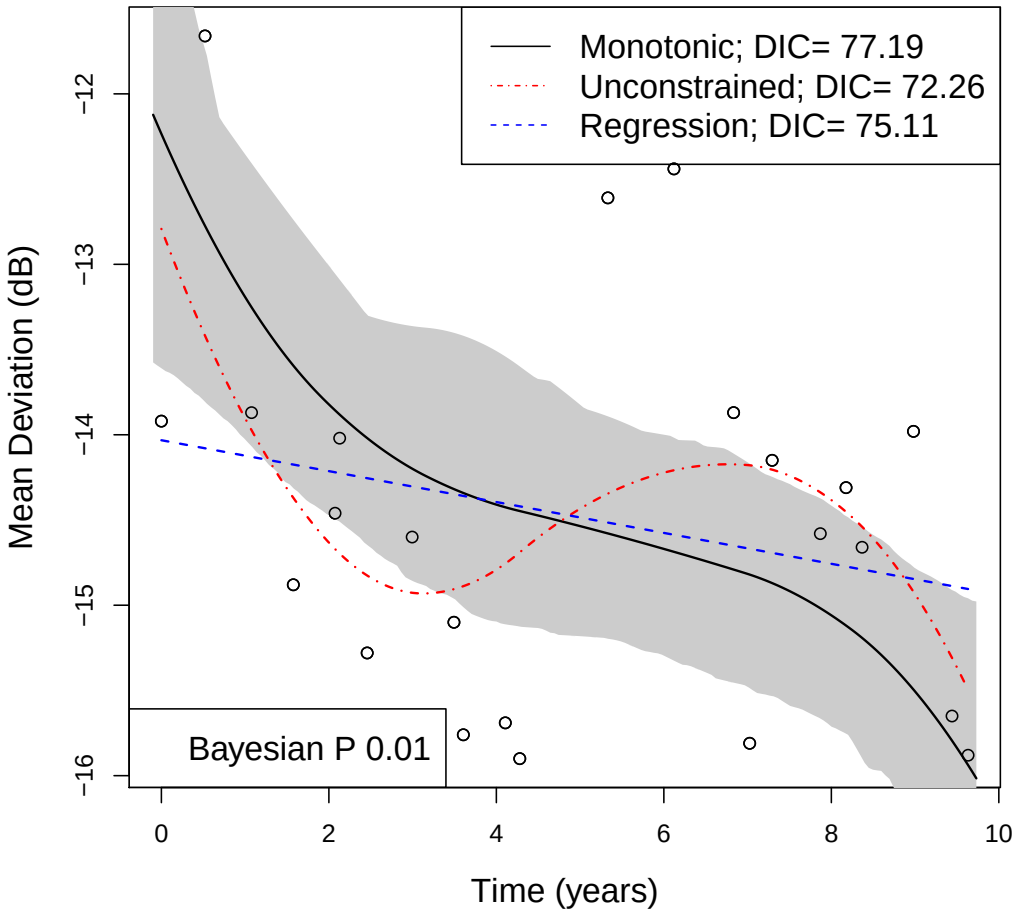


Figure 5.8: Spline curves for subject 3.

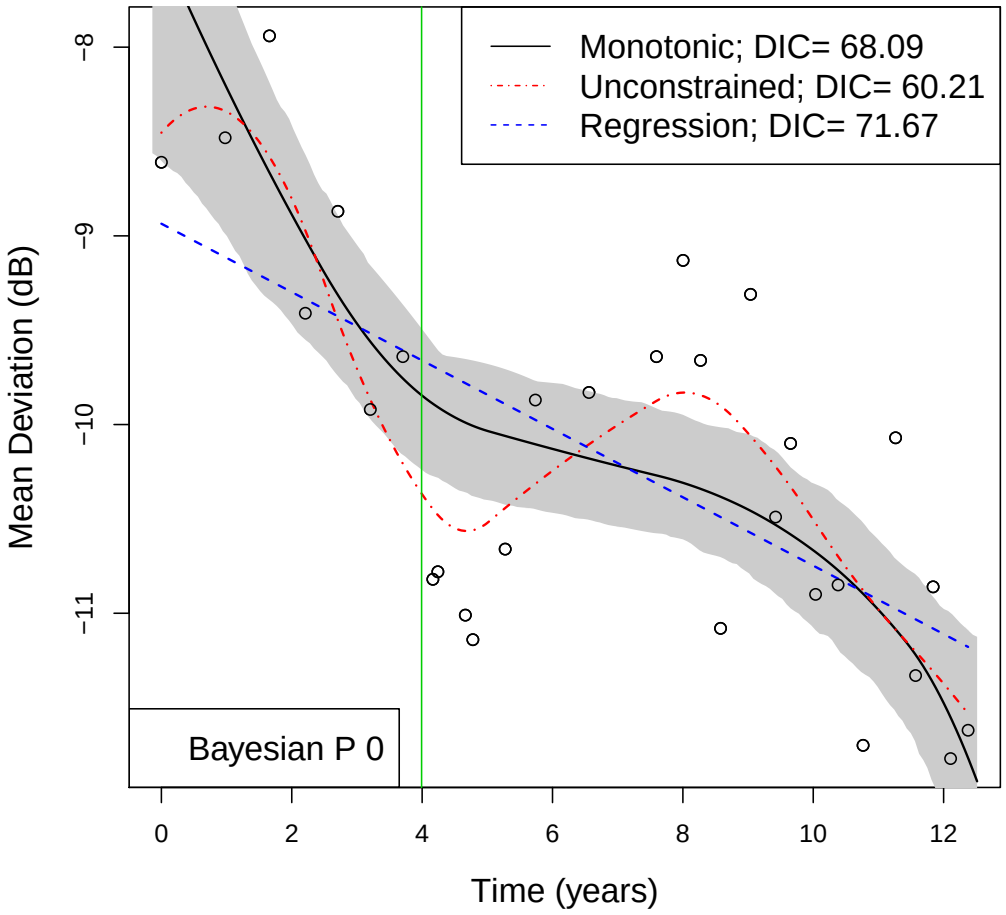


Figure 5.9: Spline curves for subject 4. The subject had an accident at the time point indicated by the vertical green line.

5.6 Chapter Summary

In this chapter we have looked at serial visual field mean deviation scores. We have shown that splines can be employed to investigate the more complex shapes observed in MD over time. We have extended the work of Hazelton & Turlach (2011) to develop a Bayesian p-value to test for the assumption of a monotonic spline and applied it to our analysis. Following this we discussed a post-hoc evaluation to account for the unexpected relationships observed in several subjects' mean deviation.

Chapter 6

Modelling Video Data: Retinal Vein Pulsation

6.1 Introduction

When the foundations of modern statistics were being developed early in the 20th century, datasets were small and simple, consisting of a few variables typically collected from designed experiments. Modern day datasets are often much larger and more complex, and include functional data, genomic datasets, and image data such as satellite images. Video data is another example, and within the field of biostatistics, modelling data extracted from contiguous video clips is in its infancy. Video data is becoming increasingly common in medicine in general, and certainly in ophthalmology.

Blood flow in the eye varies with the cardiac cycle, and the properties of this blood flow have been shown to be associated with various diseases. A brief clinical background is given in the first section of this chapter which provides motivation for the novel task of modelling retinal blood flow through video data. We then introduce statistical modelling of video data. Initially intensity data, relating to the volume of blood, must be extracted from the video sequences. We describe this image processing in Section 6.4. The second stage involves modelling these data as a time series, where the trend includes harmonic terms and an additional linear spline

component to account for the noise created by subject movement. The model is applied to 654 video sequences from 17 subjects. The associations between properties of the fitted trend curve and clinical variables are then investigated, as information from these curves, such as slope and amplitude, may have clinical importance. We examine the use of higher order frequency harmonic terms, and present our discussion and suggestions for future research.

6.2 Retinal Blood Flow through the Cardiac Cycle

Retinal blood flow is the only vascular blood flow that can be directly observed in vivo, as the retina can be seen through the pupil, whilst other blood vessels are not visible (Gao et al. 2000). Retinal vascular geometric characteristics, such as blood flow and vessel diameter, are known to be associated with blood pressure and diseases such as diabetes, coronary heart disease and ischemia heart diseases (Hao et al. 2012, Dumbskyj et al. 1995). Quantifying vessel diameter and modelling blood flow can be difficult due to the periodic nature of the data as the blood flow varies throughout the cardiac cycle (Paul & Molla 2012, Dumbskyj et al. 1995). The variation in blood flow corresponds to the diastolic and systolic phases of the heart beat. During the systolic phase (when the heart muscle contracts to pump blood into the arteries) an increase in blood flow through the ophthalmic vessels occurs, while a lower blood flow is observed during the diastolic phase (when the heart muscle relaxes and refills with blood) (Hao et al. 2012). At the peak of the systolic phase blood flow velocity is 3-4 times that of the end of the diastolic phase (Hao et al. 2012). In this section we discuss the periodic nature of blood flow and techniques for observing and modelling blood flow.

It is important to understand the variation in vessel and blood flow properties during the cardiac cycle, so that changes in the vascular properties are not attributed to pathological conditions when they are due to periodic physiological variation. Hao et al. (2012) analysed fundus images (see Section 2.2.2) synchronized to the cardiac cycle using an electrocardiogram. Semi-automatic software, requiring two points to be located manually, identifies vessels

by creating a binary image based on an intensity threshold (Gao et al. 2000). The software then identifies vessel measurements by plotting the intensity distribution over distance. The vessel intensity profile is modelled by twin Gaussian functions - where one function represents the blood filling the vessel and the other function represents error in the intensity measure created by light reflection in the eye (Gao et al. 2000). The diameter can then be calculated by measuring the distance between the two peaks of the second derivative of the Gaussian curve. Hao et al. (2012) found significant variation in vessel diameter during the cardiac cycle with maximum diameter for arteries occurring close to the R-wave (when depolarization of the heart ventricles begins), and veins halfway between the systolic and diastolic phase.

Burgansky-Eliash et al. (2012) used a retinal function imager to investigate retinal blood flow in subjects with diabetes. The retinal function imager detects the movement of blood cells by taking short video clips of length 8-24 frames. The video is taken at the same phase of the cardiac cycle to avoid heart beat pulsation bias which affects the velocity of blood flow. The blood flow is analysed through differential imaging (comparing two images, see section 6.3.2). They used a linear mixed model to account for the repeated measures data, and found that there was an increase in retinal blood flow velocity in subjects with diabetes, compared to those without (Burgansky-Eliash et al. 2012).

Blood flow through the choroid, a vascular layer in the eye, can be modelled indirectly by continuously monitoring IOP. When blood enters the eye during the systole, the volume of the eye changes creating an ocular pulse. This can then be measured by monitoring IOP using a Pulsatile Ocular Blood Flow (POBF) Tonograph, which also calculates pulse amplitude and measures pulse rate. Pulse volume can then be calculated using pulse amplitude and estimated ocular rigidity through the Friedenwald equation (Gunvant et al. 2005). The POBF is then calculated from the slope of IOP pulse. These equations have been investigated and POBF curves modelled using Fourier series analysis (Krakau 1992, Kraukau 1995). It should be noted that there is some controversy about the assumptions and accuracy of POBF.

Colour Doppler Imaging is an ultrasound technique used to calculate blood flow velocities. When an ultrasound frequency is reflected from a moving red blood cell, the frequency changes depending on how far away the blood cell is (Mitchell 1990). For example if the blood cell is moving towards the ultrasound beam the frequency reflected back is increased. The change in frequency is known as the Doppler shift, and from this the machine can calculate the blood velocity. Stalman et al. (2009) showed that the resistive index of pulsatile blood flow as measured by colour Doppler imaging is correlated with ocular pulse amplitude. In addition they found that subjects with POAG or normal tension glaucoma had reduced retrobulbar (the area behind the eye) blood flow velocities.

6.2.1 Summary

Blood flow can be measured by many methods including imaging techniques such as fundus photography and Doppler imaging or indirectly calculated by monitoring IOP. Most methods quantify blood flow velocity or vessel diameter, showing that change occurs during the cycle, however statistical modelling of these periodic changes does not seem to have been attempted. This may be partially due to the fact that the images analysed were taken at non-consecutive time points. Fourier series was used to create a mathematical model of POBF, however this was not validated with data. Our analysis therefore differs from previous analyses in that our data is extracted from contiguous image frames from video. Instead of just classifying the image as changing or quantifying how much the image is changing we fit statistical models to analyse the periodic time series nature of our data.

6.3 Statistical Modelling of Video Data

The amount of video data available is rapidly increasing due to the numbers of CTV and traffic cameras, as well as video from sources such as remote sensing, medical imaging and videos produced by the entertainment industry. Due to the large amount of data stored in

a video image it is also one of the most time consuming data types to process and analyse (Daniel & Chen 2003). Therefore it is desirable to have automatic techniques to extract, process and analyse data from video images. In following sections we define video data and discuss methods to detect change in video data. We then discuss quantifying this change through modelling the red-green-blue (RGB) colour intensity of each pixel.

6.3.1 Definition

We define video data as a sequence of M static images $\{\mathbf{I_1}, \mathbf{I_2}, ..., \mathbf{I_M}\}$. Each image then maps each pixel coordinate ($\mathbf{x}$) to a colour (or intensity) $\mathbf{I}(\mathbf{x}) \in \mathbb{R}^k$, where $k = 1$ for grey-scale images and $k = 3$ for RGB images (see Section 6.3.3 below) (Radke et al. 2005). The modelling of this data then focuses on the differences between the consecutive images.

6.3.2 Image Differencing

As video data can be thought of as a series of contiguous static images, the simplest analysis involves asking whether a pixel at location $\mathbf{x}$ on one image is the same as the pixel at location $\mathbf{x}$ on the precious image (i.e. $\mathbf{I}_{i+1}(\mathbf{x}) = \mathbf{I}_i(\mathbf{x})$). When the background scene in a video remains the same, this is known as image change detection. In image change detection the background scene remains the same, therefore the aim is to identify the pixels which are 'significantly different' from those of the previous image frame. The pixels identified as changing are referred to as the change mask (Radke et al. 2005).

In the basic case, 'significant' change is defined as $|D(\mathbf{x})| > \tau$, where $D(\mathbf{x}) = I_{i+1}(\mathbf{x}) - I_i(\mathbf{x})$. This is known as simple differencing. When detecting change it is important to differentiate between true change and unimportant change due to such factors as camera movement and variation in illumination (Radke et al. 2005). What is classified as 'unimportant' or 'significant' change varies depending on the application and therefore algorithms are often application specific (Radke et al. 2005).

Decision rules to detect change can also involve hypothesis testing, for example likelihood ratio tests. When type of change is important, such as an intensity change versus a movement of an object, probabilistic mixture models have been used. Spatial and spatial temporal models have also been used to exploit the correlation between neighbouring pixels. These methods will not be discussed further as for our purposes it is sufficient to look at the average change in pixel intensity. Other video processing methods include background modelling, where the aim is to identify pixels that belong to the background and are not of interest in terms of pixel change. We also bypass this section of literature as we will only be analysing the selected foreground removed from the total image.

Analysis of video data in ophthalmology

Image change detection has been used in ophthalmic research (Conrad et al. 1994, Dumbskyj et al. 1995). Conrad et al. (1994) create a montage of non-overlapping adjacent images taken from video data. They then use image processing methods to reduce the background noise and manually outline the area for quantification. A threshold for density of the vessels is then identified, and used to calculate vessel area via pixel counting. They found that using this method accurately identified small changes in vessel area. Dumbskyj et al. (1995) took a series of 12 non-consecutive fundus photographs (Section 2.2.2) throughout the duration of the cardiac cycle, synchronized with an electrocardiogram. An automated computer program calculated the transmittance profile and used this to identify the diameter of the vessels at various points of the cardiac cycle. Neither of these methods used contiguous video data.

6.3.3 Modelling RGB data

We are interested in quantifying how the pixels are changing with respect to the overall image. We can do this by looking at the red, green, blue (RGB) content of each pixel. RGB colours are defined by a triple of numbers ranging between 0 and 1. True red is therefore represented by $[1, 0, 0]$, true green $[0, 1, 0]$ and true blue $[0, 0, 1]$. As these are the primary

colours all other colours can be comprised with different values of RGB. For example pale orange is $[0.99, 0.75, 0.25]$, white is $[1, 1, 1]$ and black is $[0, 0, 0]$ (Koenderink 2010). Generally computers and other digital devices rescale the RGB 0 to 1 values to integers between 0 and 255. RGB values for an image can be extracted by software such as the R package `jpeg` (Urbanek 2013).

While there has been a lot of research into analysing video in terms of indexing, segmentation, and identifying background and foreground objects, there does not appear to be as much literature about modelling the objects of interest once they have been cleaned and identified. One example is the creation of a fire alarm through the analysis of RGB data. Huang et al. (2003) use RGB pixel information and create a model that automatically detects fire from a video sequence. They do this through a series of rules based on the colour and colour distribution of the pixels. As the colour of fire generally ranges from yellow to red, the first rule states $R \geq G \geq B$. In addition the value of R must be greater than threshold R_T, and to avoid the effects of background illumination, the saturation should be inversely proportional to R ($S \geq (255 - R) \cdot S_T/R_T$), where R_T and S_T are based on prior experiments. Rules that take into account the growth of the fire were also included by counting the number of fire coloured pixels within each frame, and including an algorithm that investigated the invariability of the centroid. If all the rules are breeched a fire alarm is triggered.

The next section describes the data and how the numerical information is extracted from the video images. This is followed by a section discussing the modelling of this pixel information.

6.4 Data

This section describes the subjects from whom the data was collected and the data collection method. We describe the data extraction, detailing how the video data is converted to numeric data. For the duration of this chapter we use the data from one eye as an example to illustrate

the processes involved in modelling video data, from the initial extraction of information to analysing the final models.

6.4.1 Subjects

Video data was available for a subset ($n = 17$) of the subjects described in Section 3.6.1. The majority of subjects (76%) had POAG in at least one eye, of these 8 had bilateral POAG. Of the remaining subjects, 3 had hydrocephalus (accumulation of cerebrospinal fluid in the brain) due to spina bifida, a cerebellar tumour and unknown causes. The final subject had bilateral pseudo-exfoliation glaucoma.

6.4.2 Video Collection

Figure 6.1 illustrates the set-up for recording the video and collecting the other key variables. Intracranial pressure (ICP) was recorded via intracranial transducers, and the length of a cardiac cycle was monitored using a pulse oximeter. An ophthalmodynometer (Meditron, Völklingen, Germany) was used to induce venous pulsation, if necessary, and venous pulsation was recorded by video for a minimum of 10 cardiac cycles. These videos of the 'IOP Goldman applanation tonometry mires and the central retinal vein pulsation' were recorded using a video slit lamp (Leica, Wetzlar, Germany). Video resolution was 720×526 pixels at 25 frames per second (Morgan et al. 2012).

6.4.3 Video Characteristics

Video sequences at varying ophthalmodynamic forces (ODFs - see Section 2.2.6) were recorded for each subject over several cardiac cycles. Videos for a typical subject are tabulated below (see Table 6.1). Our complete dataset consisted of 654 video sequences from 17 subjects. The mean number of sequences for each subject was 38 (range 21-60), with varying numbers from each eye. Three cardiac cycles were recorded for each sequence, with a mean of 23 frames per cycle. The mean ODF of all subjects was 16.7, but ranged from 1 to 46.

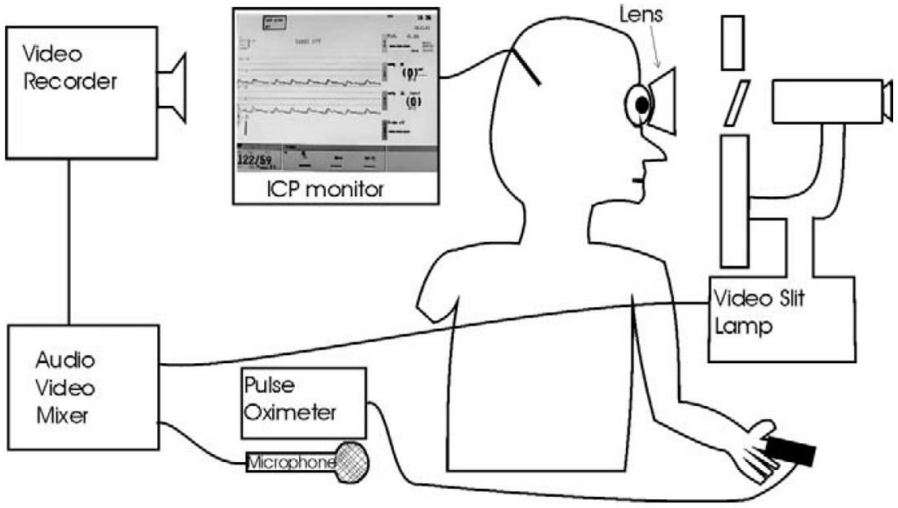


Figure 6.1: Schematic diagram showing the how the experimental data were recorded (Morgan et al. 2012).

Table 6.1: Video sequences obtained for an example subject at varying ODF levels.

		Number of Frames		
Eye	ODF	Cycle 1	Cycle 2	Cycle 3
Left	5	24	24	24
	6	25	24	24
	6	25	24	24
	7	24	24	24
	11	25	25	24
	15	23	24	24
	20	25	23	24
Right	2	24	23	23
	8	23	23	24
	10	23	24	23
	19	24	24	25
	24	24	26	24
	26	25	24	24

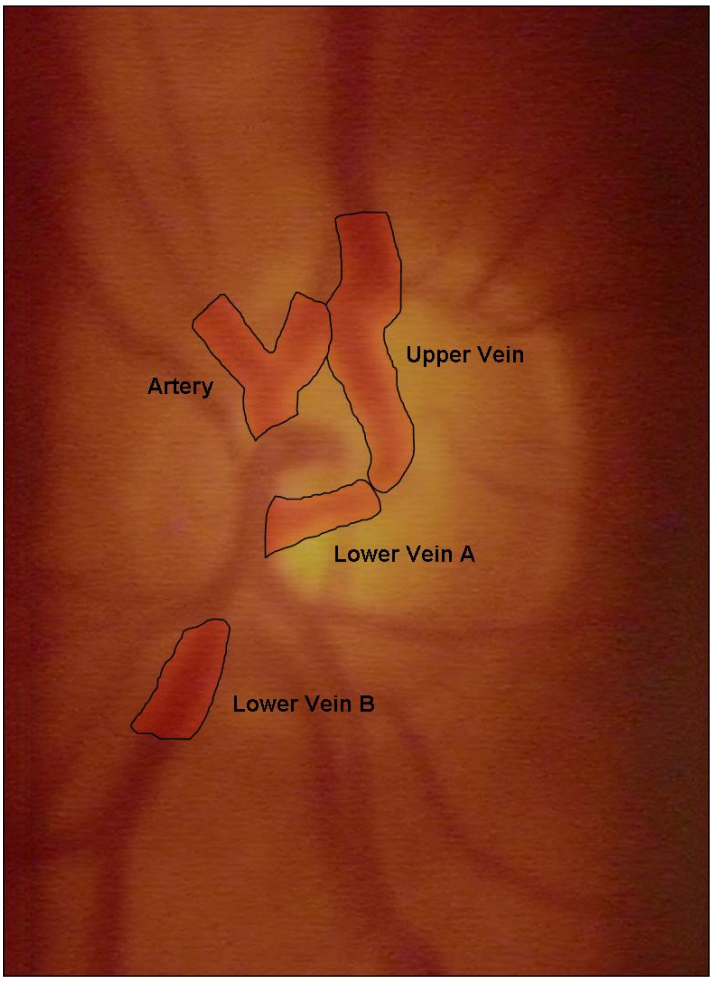


Figure 6.2: Image frame showing templates for the sections lower vein, upper vein and artery removed for further analysis.

6.4.4 Converting Video Data to Numeric Data

In this section we describe how numeric information was extracted from the video data. As mentioned above, each video sequence was broken up into individual JPEG image frames. From these, three templates were created for each sequence to select a section of the artery, upper and lower veins. An example of these templates can be seen in Figure 6.2. The templates three vessels are modelled independently of each other. Figure 6.3 shows the sequence of 70 JPEG frames of the artery through three cardiac cycles.

The individual vessel images created from the templates were then read into R (R Core Team

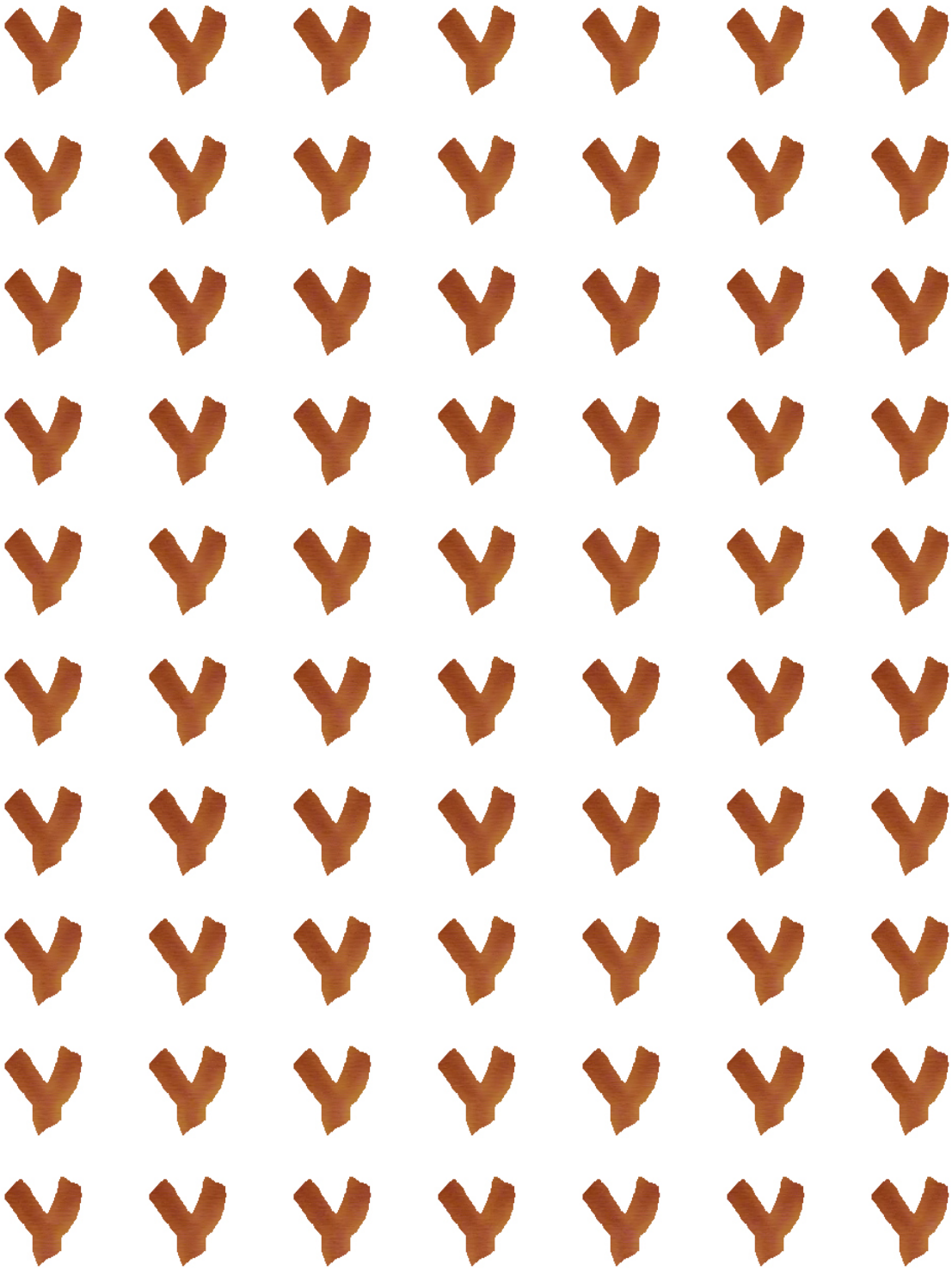


Figure 6.3: Sequence of video frames of the artery.

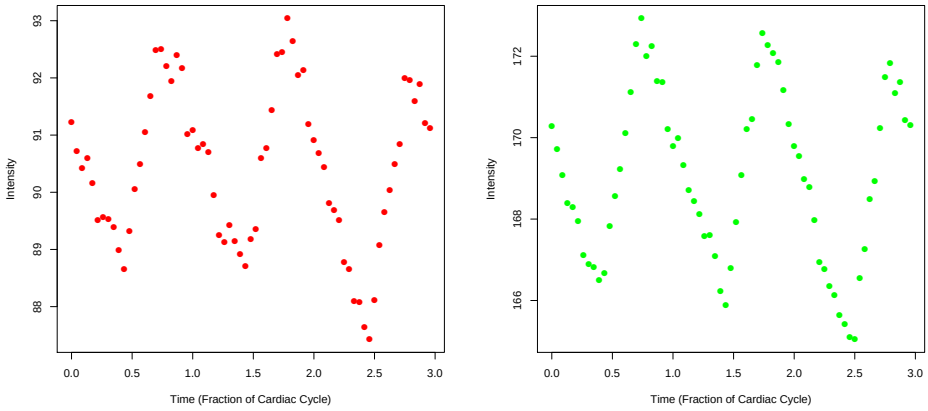
2013) using the package jpeg (Urbanek 2013) which extracts the RGB data from each image. As mentioned in Section 6.3.3, each pixel within the image is then assigned three values between 0 and 1, one for each of the three colour channels (red, green, blue). These values are then trimmed to exclude any 'bad' pixels, defined as being black or white. As black is [0,0,0] and white is [1,1,1] we defined a 'bad' pixel as $\sum$[R,B,G] < 0.03 or > 2.97. The mean of each colour channel was then calculated (excluding 'bad' pixels). This average can be thought of as a summary of the RGB intensity of the whole image. These values were then converted to the commonly used RGB 0-255 scale. The numeric representation of the 70 arterial images in Figure 6.3 can be seen in Table 6.2. Figure 6.4 shows the arterial data plotted overtime, where time is measured as a fraction of the cardiac cycle, rather than in seconds, so that each cardiac cycle would be of a standard length. In Figure 6.4 (a) and (b), showing the red and green channels respectively, we can observe a cyclical trend. The blue channel seems to be mostly noise and does not appear to contribute useful information with respect to blood volume. Similar plots were seen for the lower and upper veins.

For further modelling the green channel was selected. The green channel shows greater correlation with haemoglobin content (Yang et al. 2013), as haemoglobin absorbs green light better than red or blue (vanKampen & Zijlstra 1965, Verkruysse et al. 2008). In addition the Bayer sampling used on video cameras favours green as it has two green filters, one red, one blue, to better mimic the physiology of the human eye (Bayer 1976). The fact that there are twice as many green filters results in the green channel being the least noisy of the three colour channels.

As the green light passes through the tissues it is reflected back through the tissue structures to the camera sensor. As most of the light in blood vessels is absorbed through haemoglobin, the amount of light absorbed can be thought of as being proportional to the amount of haemoglobin in the light path. Therefore the average intensity of each template represents the haemoglobin content of this section of the vessel. A darker image - recorded as a higher intensity- means more light has been absorbed, indicating a higher concentration of

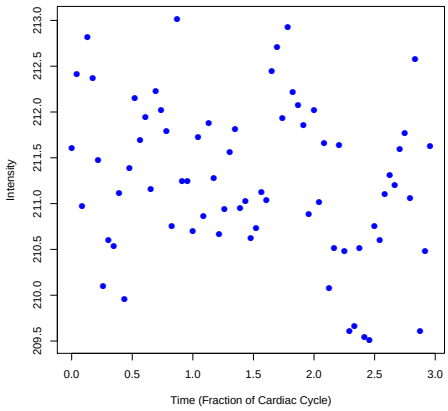
Table 6.2: Numeric data extracted from the sequence of JPEG artery images. Mean RGB values for each frame.

	frame	R	G	B	cycle	time
1	1	91.22	170.28	211.60	1	0.00
2	2	90.72	169.72	212.41	1	0.04
3	3	90.42	169.08	210.97	1	0.09
4	4	90.60	168.40	212.82	1	0.13
5	5	90.15	168.29	212.37	1	0.17
6	6	89.52	167.95	211.47	1	0.22
7	7	89.57	167.11	210.10	1	0.26
⋮	⋮	⋮	⋮	⋮	⋮	⋮
24	1	91.09	169.79	210.70	2	0.00
25	2	90.77	169.98	211.72	2	0.04
26	3	90.85	169.33	210.86	2	0.09
27	4	90.71	168.72	211.88	2	0.13
28	5	89.95	168.45	211.28	2	0.17
⋮	⋮	⋮	⋮	⋮	⋮	⋮
64	18	90.84	170.24	211.60	3	0.71
65	19	92.00	171.50	211.77	3	0.75
66	20	91.95	171.84	211.06	3	0.79
67	21	91.60	171.09	212.58	3	0.83
68	22	91.90	171.36	209.61	3	0.88
69	23	91.21	170.44	210.48	3	0.92
70	24	91.13	170.32	211.63	3	0.96



(a) Red Channel

(b) Green Channel



(c) Blue Channel

Figure 6.4: Raw data for example sequence from the artery. (a) and (b) show a cyclical trend, while (c) appears to show mostly noise.

haemoglobin. Therefore the change in mean image intensity, represented by the green channel, is proportional to the change in blood volume of the vessel. As mentioned earlier this change in blood volume is viewed in the eye as venous pulsation.

6.5 Modelling Retinal Venous Pulsation

For each subject's eye we model retinal venous pulsation in the three vessels individually. Figure 6.4 shows typical data obtained from a video sequence representing retinal venous pulsation in the artery due to the change in blood volume through the cardiac cycle. There are three main components to this data, the cyclical variation, the non-cyclical variation and the residual variation. The cyclical variation is due to the periodicity of the cardiac cycle. The non-cyclical trend is due to the subject movement during the filming of the video sequences. In this dataset residual variation must take into account the correlation between data points which occurs in time series data. We model mean image intensity for each vessel:

$$y_t = f(t) + \epsilon_t, \tag{6.1}$$

where y_t is the mean image intensity represented by the green channel and ϵ_t is the residual error process. The trend $f(t)$ can be further broken down into:

$$f(t) = f_p(t) + f_{np}(t), \tag{6.2}$$

where f_p is the periodic component and f_{np} is the non-periodic component.

As mentioned we model time as a fraction of the cardiac cycle. The idea is that by 'warping' time we standardise the length of the cardiac cycle. That is, the nominal time for the ith frame in the cth cycle is:

$$t = \frac{i}{n_c} + c - 1, \tag{6.3}$$

where n_c is the number of frames in the cth cycle. This means that if the number of frames in the cardiac cycle is constant, there is no time warping. In general, time warping will only

be slight (e.g. cycles vary 24,25,24 - see Table 6.1).

6.5.1 Cyclical Trend

Fourier series expansions are frequently used to model cyclical trends, and have been used in a wide range of statistical and mathematical applications. Fourier series have been used in economics, environmental science, engineering and image processing. In the medical field Fourier series have been used to estimate heart rate from data obtained during near-infrared spectroscopy (Trajkovic et al. 2011). Using a Fourier model, estimated heart rates were calculated, which highly corresponded to the measured rates from the electrocardiogram. Computing heart rate from spectroscopy eliminates the need of an additional device to measure heart rate.

Fourier Series in Modelling Blood Flow

Robertson et al. (2001) investigated simulated vessel pulsatile flow within a cylindrical PVC tube, set up in such a way as to mimic blood flow in vivo. Phase constant MRI was used to extract velocity data from the MRI pixels via an automatic algorithm. An analytical model using Womersley solution (a dimensionless expression of pulsatile frequency) and the Navier-Stokes equations which describe the motion of fluid were applied to the data (Robertson et al. 2001). While this is adequate to model a sinusoidal wave a more complex wave form can be modelled using Fourier series, which allows for the decomposition of the the wave 'into a series of simple harmonic terms'. The superimposition of the individual Fourier terms allows for the 'calculation of the temporal-spatial behaviour of the velocity for the flow waveform'. The analytical results were very similar to the experimental results. However, while the analytical method is useful in vitro, it is not generally applicable in vivo as not all the flow theory assumptions hold in a human blood vessels. Fourier series analysis has also been used within a mathematical model incorporating computational fluid dynamic equations to analytically model pulsatile flow during arterial narrowing (Paul & Molla 2012).

Fourier Series Defined

A Fourier series consists of potentially infinite linear combinations of the oscillating functions, sine and cosine. The Fourier expansion for the 2π-periodic function, $f_p(x)$, is written as:

$$f_p(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} [a_n \cos(nx) + b_n \sin(nx)] \tag{6.4}$$

where

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(nx) dx, \ n \geq 0 \tag{6.5}$$

$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin(nx) dx, \ n \geq 1 \tag{6.6}$$

(Brown & Churchill 2012, Stade 2005).

The Fourier series will model the periodic component ($f_p(x)$) of the trend. For many periodic functions, a good approximation can be obtained using the lowest frequency terms. Given the nature of the cardiac cycle we would not expect multiple turning points. This motivates the statistical modelling of blood flow throughout the cardiac cycle using just low order terms. We expand the first order Fourier series to a second order Fourier series with higher frequency terms, and an intercept term, (a_0):

$$f_p(t) = a_0 + a_1 \cos(2\pi t) + a_2 \sin(2\pi t) + a_3 \cos(4\pi t) + a_4 \sin(4\pi t). \tag{6.7}$$

6.5.2 Non-cyclical Trend

Slight subject movement during the cardiac cycle could introduce changes in the RGB intensity values that are due to noise rather than reflecting physiological changes. We mentioned this could be seen in Figure 6.4 (a) and (b), however some subjects have more pronounced movement as seen in Figure 6.5. Therefore we account for this by adding a linear trend to each cycle, 'tied together' to form a linear spline through time.

Our full model for the trend $f(x)$ from Equation 6.1 hence becomes:

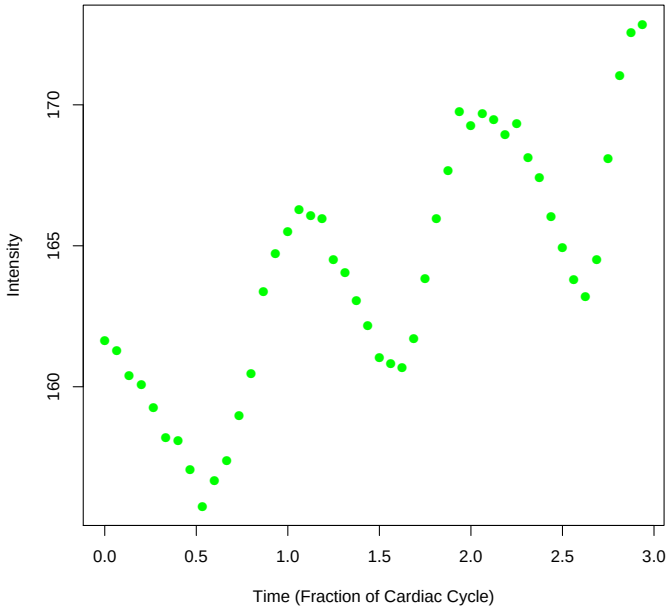


Figure 6.5: Green channel for subject sequence with movement during cardiac cycles.

$$\begin{aligned} f(t) &= f_p(t) + f_{np}(t) \\ &= a_0 + a_1\cos(2\pi t) + a_2\sin(2\pi t) + a_3\cos(4\pi t) + a_4\sin(4\pi t) \\ &\quad + b_1 t + b_2(t-1)_+ + b_3(t-2)_+ \end{aligned} \tag{6.8}$$

where

$$z_+ = \begin{cases} z & z > 0 \\ 0 & z \leq 0 \end{cases}$$

as detailed in Chapter 5.

6.5.3 Residual Variation

As we are dealing with time series data, we are likely dealing with correlated errors. We decided to account for this correlation using an autoregressive process. Our first order au-

toregressive model is:

$$y_t = f(t) + \epsilon_t \tag{6.9}$$

$$\epsilon_t = \rho_1 \epsilon_{t-1} + u_t \tag{6.10}$$

where $f(t)$ is the trend from Equation 6.8 and ϵ_t models the error. This process states that the current error (ϵ_t) is constant multiple of the error from the previous time period ($\rho_1 \epsilon_{t-1}$) plus an unstructured random error (white noise) term, u_t, assumed to be normally distributed with zero mean and variance σ_u^2. Therefore $\rho_1 \epsilon_{t-1}$ accounts for the autocorrelation in the response between consecutive time periods (Congdon 2001, Bowerman et al. 2005).

6.5.4 Model Fitting

Our models are fitted using generalized least squares, which is a version of the least squares method used to estimate regression parameters when the data contains correlated observations or heterogeneous variability (Carroll & Ruppert 1988). The model coefficients $(a_0, \ldots, a_4, b_1, \ldots, b_3)$ are estimated by minimizing the squared Mahalanobis distance of the residual vector, $y - f(t)$, where $f(t)$ is the trend as in Equation 6.8. The Mahalanobis difference differs from the Euclidean difference in that it accounts for correlation within the data. For more details refer to texts such as Carroll & Ruppert (1988) or Bates (1988). The autoregressive parameter ρ is fitted by REML. We fit our models in R (R Core Team 2013).

6.6 Model Implementation and Analysis

6.6.1 Analysis of One Sequence

Our model with autocorrelated error terms, Equations 6.8-6.10, was run for each sequence of video frames. Table 6.3 shows the results for one sequence of video frames. The model had a residual standard error of 0.494 (95% confidence interval: 0.364, 0.595) and the autocorre-

lation between terms was 0.615 (95% confidence interval: 0.395,0.769). Figure 6.6 shows the fit of the model against the raw data for the same sequence. The red, green and blue points differentiate between the 3 cardiac cycles for which we have data.

Table 6.3: Results for Fourier Series model (Equation 6.8) with AR(1) error structure for example sequence.

Parameter	Estimate	Std. Error	Pvalue
a_0	169.41	0.35	<0.0001
a_1	-2.41	0.16	<0.0001
a_2	1.52	0.15	<0.0001
a_3	0.16	0.12	0.1899
a_4	-0.62	0.11	<0.0001
b_1	0.12	0.51	0.8104
b_2	-0.72	0.85	0.4018
b_3	-0.33	0.87	0.7038

Fitted models are examined using residual plots. When dealing with autocorrelated data it is not sufficient to only use a marginal residuals plot. This is because the variance of the marginal residuals ($m_t = y_t - f(t)$), will still show autocorrelation which interferes with other patterns in the residuals indicating model mis-specification (Fraccaro et al. 2000). Fraccaro et al. (2000) provide a solution in examining the orthogonal residuals. Orthogonal residuals are defined as $v_t = \Sigma^{-1} m_t$ where Σ is the covariance matrix of the error. Properties of the matrix Σ^{-1} give the orthogonal errors low autocorrelation, with v_t only being correlated with orthogonal errors that come directly before or after it.

The residual plots discussed are shown in Figure 6.7 for our example sequence. In (a) the marginal residuals are shown and there appears to be a cyclical trend. Due to the autocorrelation of our time series data this is expected in a marginal residuals plot. Therefore we use the orthogonal residuals in (b) to validate the regression model as mentioned above (Fraccaro et al. 2000). As the cyclical trend can no longer be seen in the orthogonal residual plot, the AR(1) process adequately models the correlation in our data.

We investigate the periodic component ($f_p(t)$) of the trend by calculating the amplitude,

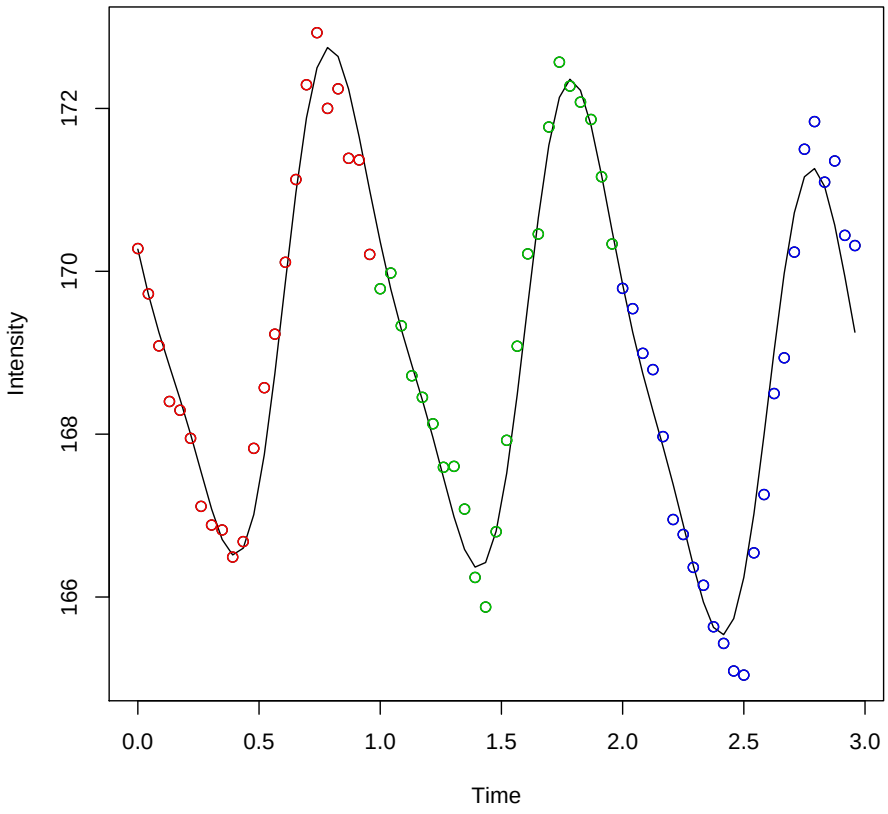
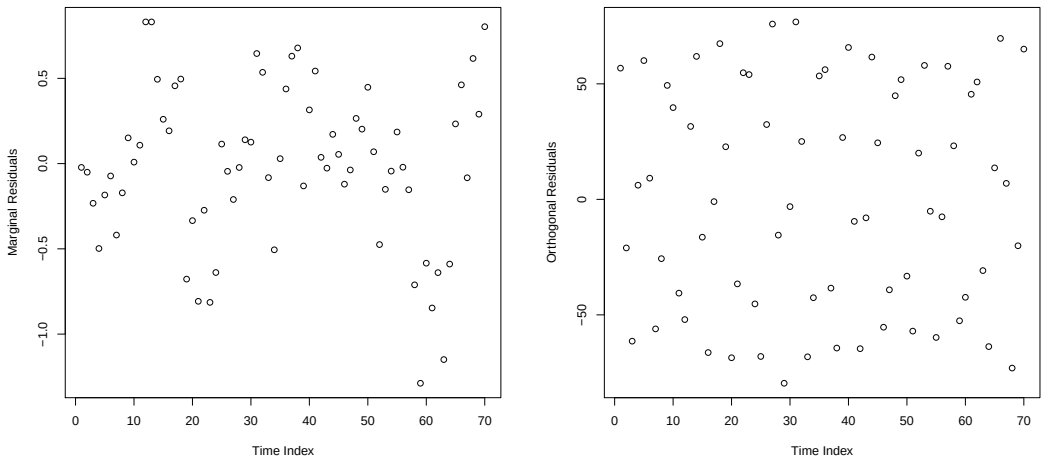


Figure 6.6: Fourier Series with linear adjustments for example sequence. Different coloured points differentiate between the three cardiac cycles.



(a) Marginal Residuals (b) Orthogonal Residuals

Figure 6.7: Marginal and orthogonal residual plots for example artery model.

maximum (maximum upwards) and minimum (maximum downwards) slopes, as these curve parameters are known to be correlated with retinal blood flow (See Figure 6.8). These values are tabulated in 6.4 and can be seen in Figure 6.9 for all three vessel types.

Table 6.4: Amplitude, maximum and minimum slopes for an example eye. Time as a fraction of the cardiac cycle.

	Amplitude	Max. Slope (Time)	Min Slope (Time)
Upper Vein	7.43	32.86 (0.68)	-19.70 (0.01)
Lower Vein	18.45	80.33 (0.66)	-56.22 (0.00)
Artery	6.16	25.88 (0.60)	-14.46 (0.94)

6.6.2 Analysis of All Sequences

As mentioned in Section 6.2, blood flow characteristics are associated with many diseases. Therefore we investigate the relationship between the curve parameters, relating to blood flow, and clinical characteristics. The mean curve parameters are summarized below in Table 6.5.

As discussed earlier in the chapter the video sequences were filmed at different levels of ODF for each eye. The relationship between ODF and characteristics of retinal venous pulsation is

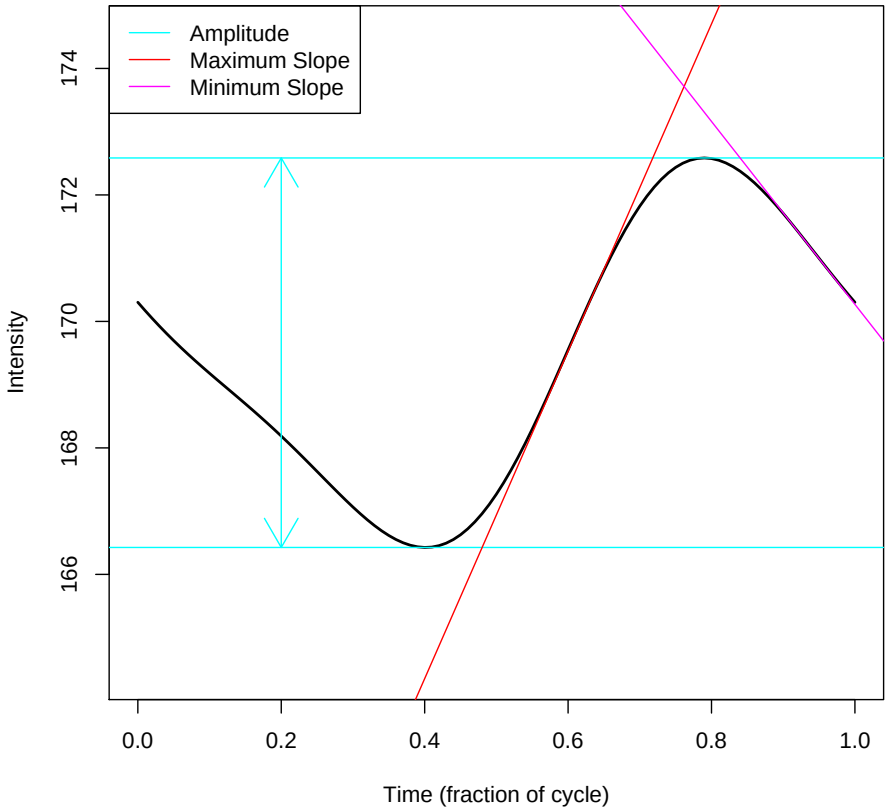


Figure 6.8: Fitted periodic component of trend for the example sequence showing computation of curve parameters. Amplitude is shown by the vertical arrow in light blue. Maximum slope can be seen in orange and minimum slope in pink.

Table 6.5: Mean amplitude, maximum and minimum slopes for all sequences.

	Amplitude (Range)	Max Slope (Range)	Min Slope (Range)
Upper Vein	7.87 (0.83,39.02)	33.87 (1.88,171.20)	-23.48 (-124.50,-3.55)
Lower Vein	8.42 (0.39,35.04)	35.98 (1.57,136.40)	-24.95 (-118.80,-1.88)
Artery	5.10 (1.06,44.37)	21.88 (3.36,224.30)	-15.98 (-201.30,-2.83)

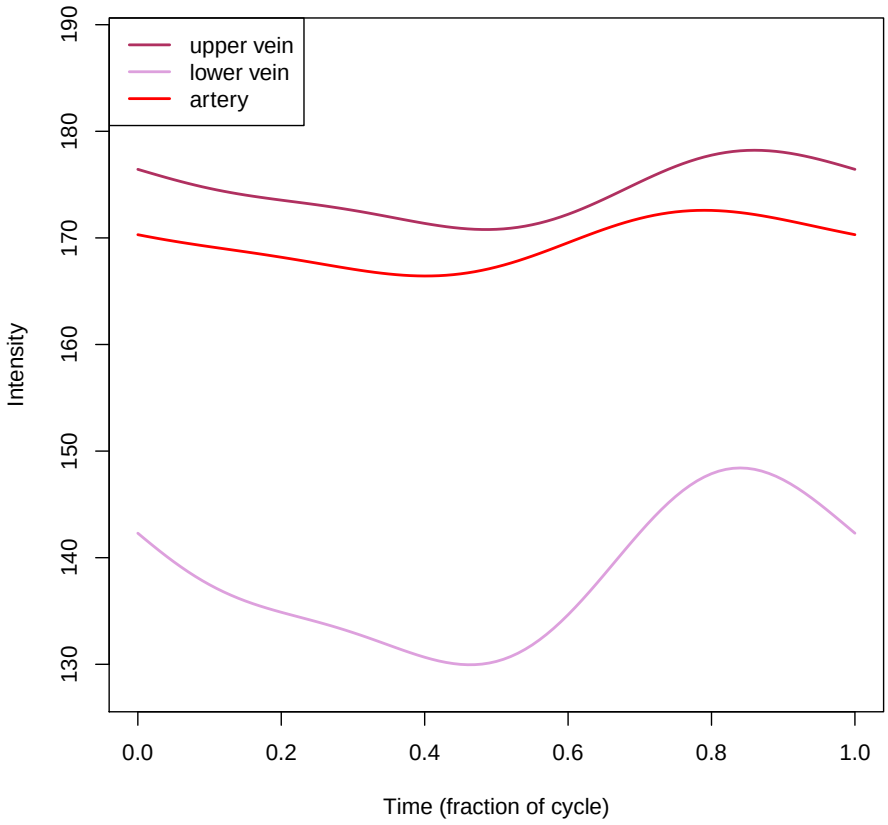


Figure 6.9: Plots showing the periodic function of the model for 3 sequences; Upper vein, lower vein and artery from one example eye.

of clinical interest. Therefore we compared the curve parameters for each eye to the clinical variable ODF (Figure 6.10). In addition we looked at the relationship between vessel type and curve parameters (Figure 6.11). The scatter plots indicate there is no or a very weak correlation between variables. Linear regression results support this observation, in that while all were significantly associated with ODF ($p<0.05$), the R^2 values did not exceed 0.03. Upper and lower veins were statistically different from arteries with respect to all curve parameters, but again the R^2 values were extremely low (< 0.07).

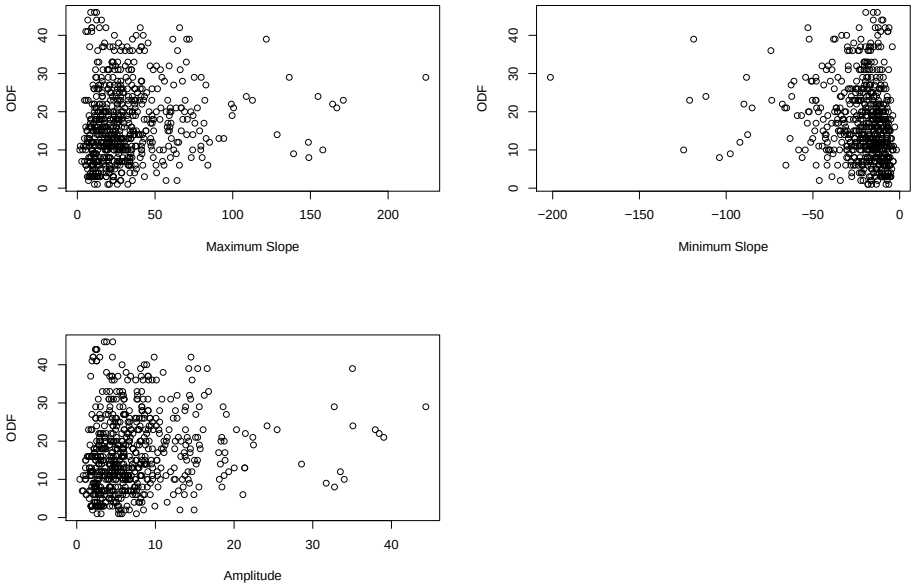


Figure 6.10: Scatter plots of the relationship between curve characteristics and ODF.

It should be noted that these parameters (slope and amplitude) are 'artificial' data points that are created parameters through previous modelling. In addition each slope/amplitude is subject to error. It is therefore unlikely the data upholds the assumptions required for statistical modelling. The complexities from combining regression parameters must be dealt with carefully (Becker & Wu 2007, LeBlanc & Tibshirani 1996). For this reason further statistical modelling such as multiple regression and mixed models to account for the multiple sequence measures on one eye is not entertained.

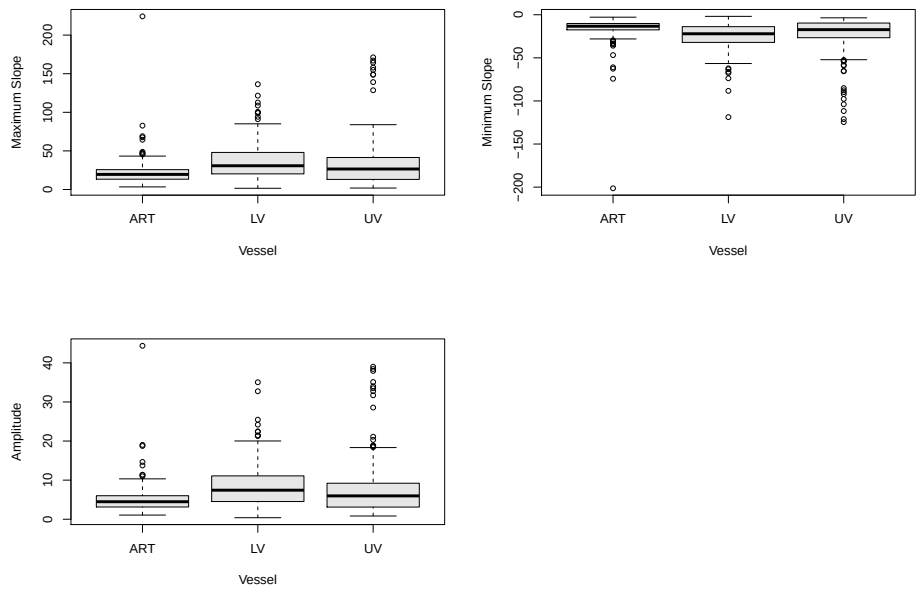


Figure 6.11: Box plots showing the relationship between vessel type and curve characteristics.

6.7 Higher Order Fourier Series

In some cases a second order Fourier Series did not seem adequate to model retinal venous pulsation. Therefore we decided to investigate higher order Fourier Series. The Akaike Information Criterion (AIC) was used to compare which order of model fits best for each sequence of data (Ritz & Streibig 2008, Box et al. 2008):

$$AIC = 2k - 2\mathsf{log}(L), \tag{6.11}$$

where k is the number of parameters in the model and L is the maximum likelihood. The AIC rewards goodness of fit and penalizes unnecessary model complexity. The preferred model has a lower AIC.

Figure 6.12 shows an example sequence with the first six order models plotted over the raw data. The second order model had the lowest AIC and therefore the best fit. A second figure (6.13) shows a sequence where the preferred order is 6. There appears to be more variation

in this sequence, therefore a higher order series is a closer fit to the data. Table 6.6 shows the number of series which preferred higher (and lower) order Fourier models. Here the second order Fourier series was preferred, based on minimum AIC, by 183 of the models fitted. We have assumed that an order two model is sufficient earlier in the chapter. In Table 6.7 we show the number of sequences for which we have tangible evidence that the optimal Fourier Series is different from order two. This was based on the difference in AIC being greater than 2 (Burnham & Anderson 2002).

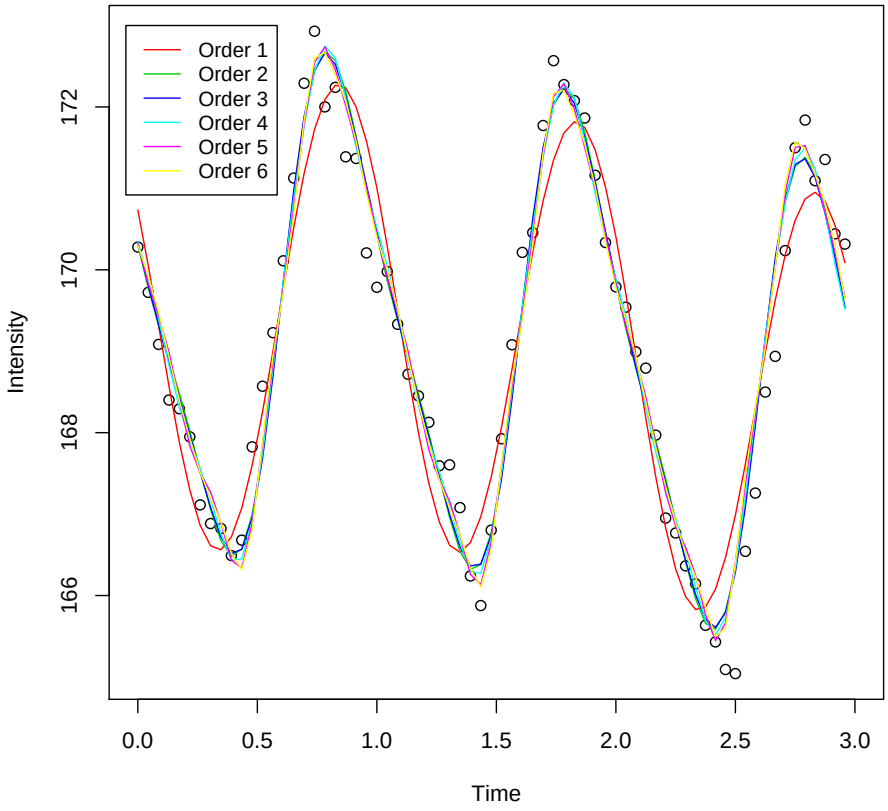


Figure 6.12: Example sequence showing Fourier Series from first to sixth order where preferred order is 2.

Table 6.6: Frequency of optimal Fourier series order for individual sequences classified by minimum AIC.

First	Second	Third	Fourth	Fifth	Sixth
104	183	134	84	77	72

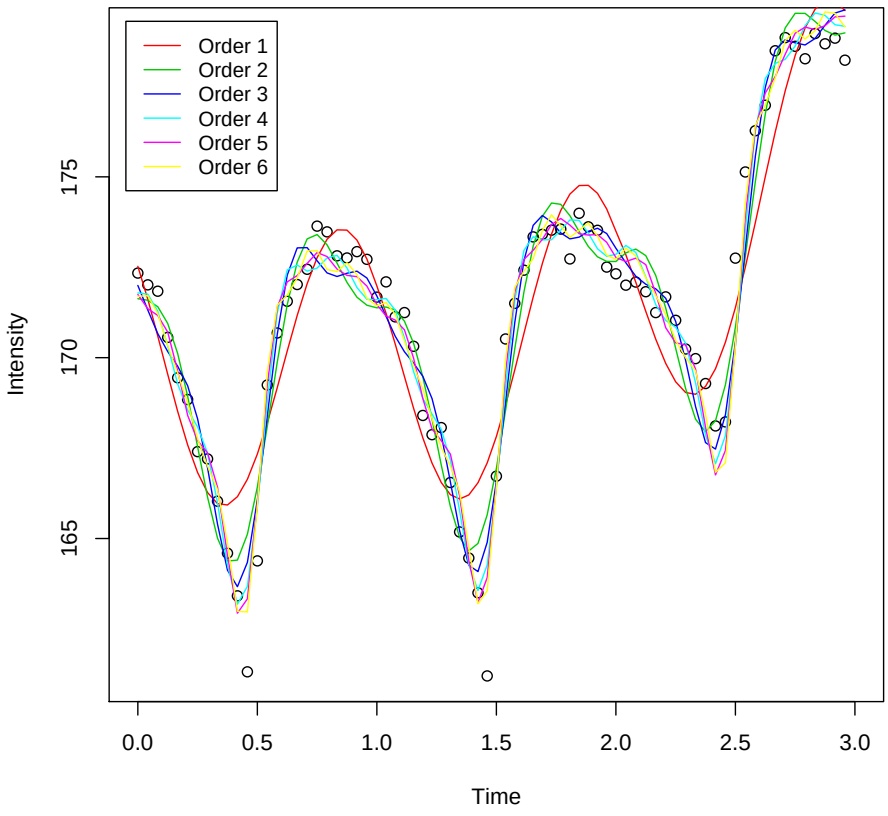


Figure 6.13: Example sequence showing Fourier Series from first to sixth order where preferred order is 6.

Table 6.7: Frequency of optimal Fourier series order for individual sequences notably different from order two (difference in AIC>2).

First	Second	Third	Forth	Fifth	Sixth
58	300	86	77	65	68

We used univariate regression to see whether there was a correlation between the Fourier Series order preferred and vein characteristics. ODF showed a significant positive association with preferred Fourier series order (P=0.014), however the R^2 for this model is very low (0.009). The type of vessel (lower vein, upper vein, artery) was not associated with Fourier Series order. The higher order models are preferred when there is more variation in the cyclical pattern (see Figure 6.13). In some cases it appears that the higher order models may be over-fitting by picking up noise rather than describing the cyclical trend of interest.

6.8 Discussion

We have developed a method to analyse RGB data extracted from video images. While this method was developed to model the pulsatility component of blood flow, it could be applied to other medical video datasets which extract information that is periodic in nature. With the increasing amount of video data it is important that methods be developed that can model information automatically extracted from image files, such as RGB data. Our model contributes to this area of modelling medical video data in that it models contiguous video data, where many previous methods have focused on a discrete set of non-consecutive images. In addition, instead of just quantifying the changes in images by pixel count, a statistical model has been fitted to the RGB data to quantify these changes over time. The issue of correlation is addressed through AR(1) process, which is applicable to other video datasets, due to the time series nature of the data.

From a clinical perspective this analysis provides information on the maximum slope and amplitude of the curve which represents blood flow through the retinal vessels. Many studies have shown the association of blood flow properties with various diseases (see Section 6.2), so it is reasonable to expect that slope and amplitude of 'blood flow' may also hold valuable

clinical insights. Krakau (1992) show theoretically that it is the steepest slope of the blood volume curve that affects the blood flow. We calculate the steepest maximum and minimum slope from our Fourier Series models. Here we can assume that amplitude corresponds to amplitude of vessel pulsation (change in blood volume) and maximum down-slope corresponds to the increased rate of flow out of the vessel. From a clinical perspective maximum down slope is of interest as it likely indicates degree of vessel obstruction. If we apply these methods to smaller sectors of the retina (ie. a 2x2 pixel area instead of the vessel template), we can create a heat map plotting maximal blood flow rates, which will aid in clinical assessment of damage in the eye. As explained the characteristics of these curves are regarded by clinicians as informative about the health of the eye. While regressions of these curves against the available clinical variables have revealed little, there is plenty of potential.

Information extracted from the same type of video was used by Morgan et al. (2012) to investigate the phase relationships of venous pulsation pressure, IOP and intracranial pressure. Morgan et al. (2012) showed a strong relationship between mean retinal venous pulsation and ICP, therefore phase timing of venous pulsation is also dependent on ICP. This suggests that 'ICP pulse pressure dominates the timing of venous pulsation.' It is hoped that further investigation of these properties will allow for a non-invasive estimation of ICP. When a bigger dataset is available it may be of benefit to see if there are any links between the curve properties and ICP, by incorporating regression synthesis techniques.

6.9 Chapter Summary

In this chapter the novel task of analysing contiguous medical video data is discussed. Relevant areas of video processing literature are reviewed and the clinical relevance of modelling retinal blood flow data is described. From the videos of retinal pulse pressure, we describe the extraction of the RGB data. With the increase of video information it is important to develop modelling techniques which can extract and analyse this particular type of data. A model using a Fourier series expression to model periodicity with AR(1) terms to model the error

process is developed which adequately fits the blood flow data. In addition we accounted for subject movement through the inclusion of a linear spline term. Information from these curves, such as slope and amplitude, may have clinical importance.

Chapter 7

Summary and Suggestions for Further Research

This thesis has investigated the application of statistical methods to three varying ophthalmic datasets. The first two arise from standard automated perimetry (SAP), which is the leading diagnostic technique for determining glaucoma progression. The third is video data of retinal venous pulsation. Conclusions and suggestions for further research are presented in this chapter.

Spatial Modelling of Visual Fields

Visual field thresholds are analysed by many techniques, however there is no gold standard. The techniques struggle to deal with the large amount of variation in SAP. The point-wise linear regression methods that are becoming more popular ignore the correlation structure of the eye and do not account for the testing of multiple hypotheses. We employ conditional autoregressive priors to correctly model spatial and spatial-temporal correlation. The priors are weighted to account for the fact that points within a sector are closer to each other than points between sectors. In addition the model is extended to adjust for physiologic features such as adjacent VF loci not mapping directly to adjacent optic disc regions and the presence of the blindspot. The CAR model borrows information from its neighbouring loci, and in doing so

can minimise the effect of outliers, thus the spurious trends observed in PLR methods are avoided. When compared to a clinical standard our model, SPROG, provided better measures of sensitivity and specificity than PLR methods. SPROG is both statistically and physiologically accurate, and provides the clinician with informative results that are easily interpretable.

SPROG modelled measurement error as a constant. Whilst this constant varied between eyes, it was not able to vary over the individual's visual field. While our dataset did not show strong dependence on location, previous research has suggested this. With a larger dataset within and between magnitude of measurement error could be investigated with respect to location and also how magnitude of measurement error varies with disease severity. Including a non-constant measurement error term in a CAR model would be complicated, and may not improve the predictivity of the model, however this could be investigated. A weighting scheme where less weight is placed on edge points could also be considered. This could be incorporated in a CAR model, or by using a different model to account for spatial correlation. While CAR models are more flexible and found to perform better than other methods to account for spatial correlation, another option would be to model spatial correlation using distance based parametric functions. Finally, linear trends have performed adequately in the majority of models, however as seen in our spline modelling chapter, longer sequences of visual fields have benefited from the incorporation of non-linear modelling. A variation of SPROG for longer fields incorporating a non-linear trend could prove interesting.

Spline Modelling of Mean Deviation

Longitudinal series of visual field mean deviation scores were modelled using quadratic splines. Splines provide flexibility to model non-linear relationships that are sometimes observed in longer series of visual field scores. We investigated constraining the splines and developed a Bayesian p-value to test the assumption of a non-increasing monotonic spline. We compared the unconstrained and constrained splines to a linear model. In the majority of cases the linear model was sufficient to model mean deviation. Relationships that were neither mono-

tonic non-increasing or linear were indicative of some intervention or other medical condition. We suggest that using the spline models would be useful as a screening tool in a clinical setting. Unusual curves could be flagged and the clinician could then re-evaluate the visual fields to familiarise themselves with the subject's history. It would be interesting to evaluate this method on a larger dataset.

Modelling Ophthalmic Video Data

As discussed in this thesis the number of medical video datasets is increasing, and the field of ophthalmology is no exception. We obtained video sequences of retinal pulse pressure, which is observed as change in blood through through the retinal vessels. Each video sequence was divided into a set of contiguous JPEG images. A set of three templates were created which extracted an area of the artery, lower vein and upper vein. RGB pixel information was then extracted from each sequence of template images. The mean pixel intensity for each template and each time point was calculated. The three vessels were modelled independently. The time series data was periodic due to the cardiac cycle. The periodic trend was modelled by a Fourier series. The non-periodic trend was modelled by a linear spline, which adjusted for patient movement during the cardiac cycle. An AR(1) error process was used to account for correlated measurement error.

Retinal blood flow has been linked to various diseases. Therefore relationships between characteristics of the curves and disease characteristics can be investigated. Because the curve characteristics are 'created' variables, special modelling techniques, such as regression synthesis, would need to be incorporated as these 'created' variables do not meet the normal assumptions required for many statistical models. In addition to the clinical applications linked to retinal blood change in the retina, these models are also applicable to other periodic medical video datasets.

Currently our method requires a person to select and draw a template of the individual ves-

sels from the JPEG images. This requires considerable time and also introduces subjectivity into the analysis. It would be useful, albeit difficult, to develop of way of modelling retinal pulse pressure from the whole image or alternatively use image processing techniques to select templates automatically from the images. We also model the three blood vessels, artery, lower vein, and upper vein independently. We could consider modelling the three vessels in a dependent manner, as blood flow between the vessels will be correlated.

7.1 Conclusions

This thesis has investigated the statistical modelling and inference of ophthalmic data. We have looked at three interesting datasets and developed new models, applying a variety of statistical techniques and adapting them to the novel aspects of the datasets. The statistical models account for the essential and desirable physiological characteristics of the data as well as appropriately modelling the statistical problems encountered such as correlation and error. The methods and models developed in this thesis are of clinical relevance, and we hope that they will have a significant impact on clinical practice for glaucoma suffers.

Bibliography

Abdi, H. (2007), Bonferroni test, *in* N. Salkind & K. Rasmussen, eds, 'Encyclopedia of Measurement and Statistics', Sage Publications, Inc., California, USA, pp. 104–107.

Anderson, D. (1992), Automated static perimetry, *in* 'Mosby Year Book', Mosby, St. Louis, USA, chapter Chapter 2.

Artes, P. & Chauhan, B. (2005), 'Longitudinal changes in the visual field and optic disc in glaucoma', *Progress in Retunal and Eye Research* **24**, 333–354.

Artes, P., Iwase, A., Kitazawa, Y. & Chauhan, B. (2002), 'Properties of perimetric threshold estimates from full threshold, SITA Standard and SITA Fast strategies', *Investigative Ophthalmology & Visual Science* **42**(8), 2654–2659.

Banerjee, S., Carlin, B. & Gelfand, A. (2004), *Hierarchical Modeling and Analysis for Spatial Data*, Chapman & Hall/CRC, Boca Raton, Florida.

Bates, D. (1988), *Nonlinear Regression Analysis and Its Applications*, John Wiley & Sons, NY, USA.

Bayer, B. (1976), *Color imaging array.* Patent.
URL: *http://www.google.com/patents/US3971065*

Bayes, T. (1763), 'An essay towards solving a problem in the doctrine of chances.', *Philosophical Transactions of the Royal Society of London* **53**, 370–418.

Becker, B. & Wu, M.-J. (2007), 'The synthesis of regression slopes in meta-analysis', *Statistical Science* **22**(3), 414–429.

Bengtsson, B. & Heijl, A. (2000), 'False-negative responses in glaucoma perimetry: Indicators of patient performance or test reliability?', *Investigative Ophthalmology & Visual Science* **41**(8), 2201–2204.

Benosman, R. (2010), 'Vision without frames: A semiotic paradigm of event based computer vision', *Biosemiotics* **3**(1), 1–16.

Bernardinelli, L., Clayton, D. & Montomoli, C. (1995), 'Bayesian estimates of disease maps: How important are priors?', *Statistics in Medicine* **14**, 2411–2431.

Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, N. & Songini, M. (1995), 'Bayesian analysis of space-time variation in disease risk', *Statistics in Medicine* **14**, 2433–2443.

Besag, J. (1974), 'Spatial interaction and the statistical analysis of lattice systems (with discussion)', *Journal of the Royal Statistical Society* **B 36**, 192–236.

Besag, J. & Kooperberg, C. (1995), 'On conditional and intrinsic autoregressions', *Biometrika* **82**(4), 733–746.

Besag, J., York, J. & Mollie, A. (1991), 'Bayesian image restoration with two applications in spatial statistics', *Annals of the Institute of Statistical Mathematics* **43**, 1–59.

Best, N., Richardson, S. & Thomson, A. (2005), 'A comparison of Bayesian spatial models for disease mapping', *Statistical Methods in Medical Research* **14**, 35–59.

Birch, M., Wishart, P. & O'Donnell, N. (1995), 'Determining progressive visual field loss in serial Humphrey visual fileds', *Ophthalmology* **102**, 1227–1235.

Boden, C., Chan, K., Sample, S., Hao, J., Lee, T., Zangwill, L., Weinreb, R. & Goldbaum, M. (2007), 'Assessing visual field clustering schemes using machine learning classifiers in standard perimetry', *Investigative Ophthalmology & Visual Science* **48**(12), 5582–90.

Bowd, C., Hao, J., Tavares, I. M., Medeiros, F. A., Zangwill, L. M., Lee, T.-W., Sample, P. A., Weinreb, R. N. & Goldbaum, M. H. (2008), 'Bayesian machine learning classifiers for

combining structural and functional measurements to classify healthy and glaucomatous eyes', *Investigative Ophthalmology & Visual Science* **4**(3), 945–953.

Bowd, C., Lee, I., Goldbaum, M., Balasubramaian, M., Medeiros, F., Zangwill, L., Girkin, C., Liebmann, J. & Weinreb, R. (2012), 'Predicting glaucomatous progression in glaucoma suspect eyes using relevance vector machine classifiers for combined structural and functional measurements', *Investigative Ophthalmology & Visual Science* **53**(4), 2382–2389.

Bowerman, A., O'Connell, R. & Koehler, A. (2005), *Forecasting, Time Series and Regression*, Thomson BrooksCole, Belmont, CA.

Box, G., Jenkins, G. & Reinsel, G. (2008), *Time Series Analysis: Forecasting and Control*, John Wiley & Sons, Hoboken, NJ.

Brezger, A. & Steiner, W. (2008), 'Monotonic regression based on bayesian p-splines: an application to estimating price response functions from store-level scanner data', *Journal of Business & Economic Statistics* **26**(1), 90–103.

Brooks, S. (1998), 'Markov chain Monte Carlo and its application', *The Statistician* **47**(1), 69–100.

Brown, J. & Churchill, R. (2012), *Fourier Series and Boundary Value Problems*, McGraw-Hill, New York, NY.

Burgansky-Eliash, Z., Barak, A., Barash, H., Nelson, D., Pupko, O., A., L., Grinvald, A. & Rubinstein, A. (2012), 'Increased retinal blood flow velocity in patients with early diabetes mellitus', *Retina* **32**, 112–119.

Burnham, K. & Anderson, D. (2002), *Model Selection and Multinomial Inference: A Practical Information-Theoretic Approach*, Springer-Verlag, NY, USA.

Caprioli, J. (1991), 'Automated perimetry in glaucoma', *American Journal of Ophthalmology* **111**(2), 235–239.

Carroll, R. & Ruppert, D. (1988), *Transformation and Weighting in Regression*, Chapman and Hall, NY, USA.

Chauhan, B., Garway-Heath, D., Goni, F., Rossetti, B., Viswanathan, A. & Heill, A. (2008), 'Practical recomendations for measuring rates of visual field change in glaucoma', *British Journal of Ophthalmology* **92**(4), 569–573.

Chauhan, B. & Johnson, C. (1999), 'Test-retest variability of frequency-doubling perimetry and conventional perimetry in glaucoma patients and normal subjects', *Investigative ophthalmology & Visual Science* **40**(3), 648–656.

Coleman, A. (2009), 'The role of statistics in ophthalmology', *American Journal of Ophthalmology* **147**(3), 387–388.

Congdon, P. (2001), *Bayesian Statistical Modelling*, John Wiley & Sons, Ltd, West Sussex, England.

Conrad, T., Chandler, D., Corless, J. & Klintworth, G. (1994), 'In vivo measurement of corneal angiogenesis with video data aquisition and computerized image analysis', *Laboratory Investigatioon* **70**(3), 426–434.

Daniel, G. & Chen, M. (2003), 'Video visualization', *IEEE Visualization 2003 Proceedings* pp. 409–416.

Das, K., Li, J., Fu, G., Wang, Z., Li, R. & Wu, R. (2012), 'Dynamic semiparametric Bayesian models for genetic mapping of complex trait with irregular longitudinal data', *Statistics in Medicine* **32**, 509–523.

De Moraes, C., Ritch, R., Tello, C. & Liebmann, J. (2011), 'Modified visual field trend analysis', *Journal of Glaucoma* **20**(4), 203–206.

Donnelly, S. & Subramanian, P. (2009), 'Relationship of intraocular pulse pressure and spontaneous venous pulsation', *American Journal of Opthalmology* **147**, 51–55.

Dumbskyj, M., Aldington, S., Dore, C. & Kohner, E. (1995), 'The accurate assessment of changes in retinal vessel diameter using multiple frame electrocardiograph synchronised fundus photography', *Current Eye Research* **15**, 625–632.

Eckert, N., Parent, E., Kies, R. & Baya, H. (2010), 'A spatio-temporal modeling framework for assessing the fluctuations of avalanche occurrence resulting from climate change: Application to 60 years of data in the northern French Alps', *Climate Change* **101**, 515–553.

Ernest, P., Viechtbauer, W., Schouten, J., Beckers, H., Hendrikse, F., Prins, M. & Webers, C. (2012*a*), 'The evidence base to select a method for assessing glaucomatous visual field progression', *Acta Ophthalmologica* **90**, 101–108.

Ernest, P., Viechtbauer, W., Schouten, J., Beckers, H., Hendrikse, F., Prins, M. & Webers, C. (2012*b*), 'The influence of the assessment of visual field progression in glaucoma: A network meta-analysis', *Acta Ophthalmologica* **90**, 10–19.

Fahrmeir, L. & Knieb, T. (2011), *Bayesian Smoothing and Regression for Longitudinal, Spatial and Event History Data*, Oxford University Press, NY, USA.

Fan, Q., Teo, Y. & Saw, S. (2011), 'Application of advanced statistics in ophthalmology', *Investagative Ophthalmology & Visual Science* **52**(9), 6059–6065.

Fienberg, S. (2006), 'When did Bayesian inference become 'Bayesian'?', *Bayesian Analysis* **1**(1), 1–40.

Fitzke, F., Crabb, D., McNaught, A., Edgar, D. & Hitchings, R. (1995), 'Image processing of computerised visual field data', *British Journal of Ophthalmology* **79**, 207–212.

Fitzke, F., Hitchings, R., Poinoosawmy, D., McNaught, A. & Crabb, D. P. (1996), 'Analysis of visual field progression in glaucoma', *British Journal of Opthalmology* **80**, 40–48.

Flammer, J. (2003), *Glaucoma*, 2nd edn, Hogrefe & Huber Publishers, Germany.

Fraccaro, R., Hyndman, R. & Veevers, A. (2000), 'Residual diagnostic plots for the checking of model mis-specification in time series regression', *Australian and New Zealand Journal of Statistics* **42**(4), 463–477.

Gamerman, D. & Lopes, H. (2006), *Markov Chain Monter Carlo: Stochastic Simulation for Bayesian Inference*, 2nd edn, Chapman & Hall/CRC, FL, USA.

Gao, X., Brarath, A., Hughes, A., Chapman, N. & Thom, S. (2000), 'Quantification and characterisation of ateries in retinal images', *Computer Methods and Programs in Biomedicine* **63**, 133–146.

Gardiner, S., Crabb, D., Fitzke, F. & Hitchings, R. (2004), 'Reducing noise in suspected glaucomatous visual fields by using a new spatial filter', *Vision Research* **44**, 839–848.

Garway-Heath, D., Holder, G., Fitzke, F. & Hitchings, R. (2002), 'Relationship between electrophysiological, psychophysical, and anatomical measurements in glaucoma', *Investigative Ophthalmology & Visual Science* **43**(7), 2213–2220.

Garway-Heath, D., Poinoosawmy, D., Fitzke, F. & Hitchings, R. (2000), 'Mapping the visual field to the optic disc in normal tension glaucoma eyes', *Ophthalmology* **107**, 1809–1815.

Gelman, A. (2006), 'Prior distributions for variance parameters in heirachical models', *Bayesian Analysis* **1**(3), 515–533.

Gelman, A., Carlin, J., Stern, H. & Rubin, D. (2003), *Bayesian Data Analysis*, Chapman and Hall, London, UK.

Gelman, A., Jakulin, A., Pittau, M. & Su, Y. (2008), 'A weakly informative default prior distribution for logistic and other regression models', *Annals of Applied Statistics* **2**(4), 1360–1383.

Gelman, A. & Rubin, D. (1992), 'Inference from iterative simulation using multiple sequences (with discussion)', *Statistical Science* **17**, 457–472.

Geman, S. & Geman, G. (1984), 'Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images', *IEEE Transactions on Pattern Analysis anf Machine Intelligence* **6**, 721–741.

Geweke, J. (1992), Evaluating the accuracy of sampling-based approaches to calculating posterior moments., *in* J. Bernado, A. Dawid & A. Smith, eds, 'Bayesian Statistics 4', Claredon Press, Oxford, UK.

Goldbaum, M., Jang, G., Bowd, C., Hao, J., Zangwill, L., Liebmann, J., Girkin, C., Jung, T., Weinreb, R. & Sample, P. (2009), 'Patterns of glaucomatous visual field loss in SITA fields automatically identified using independent component analysis', *Trans Am Ophthalmol Soc* **107**, 136–144.

Goldbaum, M., Sample, P., Chan, K., Williams, J., Lee, T., Blumenthal, E., Girkin, C., Zangwill, L., Bowd, C., Sejnowski, T. & Weinreb, R. (2002), 'Comparing machine learning classifiers for diagnosing glaucoma from standard automated perimetry', *Investagitive Ophthalmology & Visual Science* **43**(1), 162–169.

Goldbaum, M., Sample, P., Zhang, Z., Chan, K., Hao, J., Lee, T., Boden, C., Bowd, C., Bourne, R., Zangwill, L., Sejnowski, T., Spinak, D. & Weinreb, R. (2005), 'Using unsupervised learning with independent component analysis to identify patterns of glaucomatous visual filed defects', *Investigative Ophthalmology & Visual Science* **46**(10), 3676–3683.

Gray, R. (1992), 'Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis', *Journal of the American Statistical Association* **87**(420), 942–951.

Greenland, S. (1996), 'Historical HIV incidence modelling in regional subgroups: Use of flexible discrete models with penalized splines based on prior curves', *Statistics in Medicine* **15**, 513–525.

Gunvant, P., Baskaran, M., Vijaya, L. & Hansenm, B. (2005), 'Comparison of pulsatile ocular blood flow in Indians and Europeans', *Eye* **19**, 1163–1168.

Hao, H., Sasongko, M., Wong, T., Azemin, M., Aliahmad, B., Hodgson, L., Kawasaki, R., Cheung, C., Wang, J. & Kumar, D. (2012), 'Does retinal vascular geometry vary with the cardiac cycle?', *InvestigativeOphthalmology & Visual Science* **53**(9), 5799–5805.

Hastings, W. (1970), 'Monte Carlo smpling methods using Markov chains and their applications', *Biometrika* **57**(1), 97–109.

Hazelton, M. & Turlach, B. (2011), 'Semiparametric regression with shape-constrained penalized splines', *Computational Statistics and Data Analysis* **55**, 2871–2879.

Heijl, A., Bengtsson, B., Chauhan, B., Lieberman, M., Cunliffe, I., Hyman, L. & Leske, M. (2008), 'A comparison of visual filed progression criteria of 3 major trials in early manifest glaucoma trial patients', *Ophthalmology* **115**, 1557–1565.

Heijl, A., Lindgren, A. & Lindgren, G. (1989), 'Test-restest variability in glaucomatous visual fields', *American Journal of Ophthalmology* **108**, 130–135.

Heijl, A., Lindgren, G. & Olsson, J. (1987), 'Normal variability of static perimetric threshold values across the central visual field', *Archives of Ophthalmology* **105**, 1544–1549.

Henson, D., Chudry, S., Artes, P., Faragher, B. & Ansons, A. (2000), 'Response variability in the visual field: comparison of optic neuritis, glaucoma, ocular hypertension, and normal eyes', *Investigative ophthalmology & Visual Science* **41**(2), 417–421.

Holmes, C. & Heard, N. (2003), 'Generalized monotonic regression using random change points', *Statistics in Medicine* **22**, 623–638.

Huang, P.-H., Su, J.-Y. & Pan, J.-S. (2003), 'A fire-alarming method bassed on video processing', *Proceedings of the 2006 INternational Conference on Intellegent Information Hiding and Multimedia Signal Processing* pp. 359–364.

Jampel, H., Singh, K., Lin, S., Chen, T., Franxis, B., Hodapp, E., Samples, J. & Smith, S. (2011), 'Assessment of visual function in glaucoma', *Ophthalmology* **118**, 986–1002.

Juzych, M. (1992), 'Statisitcal techniques in ophthalmic journals', *Archives of Ophthalmology* **110**(9), 1225–1229.

Keltner, J., Johnson, C., Quigg, G., Cello, K., Kass, M. & Gordon, M. (2000), 'Confirmation of visual field abnormalities in the ocular hypertension treatment study', *Archives of Ophthalmology* **118**(9), 1187–1194.

Klein, G. (2010), *Industrial Color Physics*, Springer Series in Optical Sciences, Springer, NY, USA.

Knighton, R. & Gregori, G. (2013), 'The shape of the gangolian cell plus inner plexiform layers of the normal human macula', *Investigative Ophthalmology & Visual Science* **53**(11), 7412–7420.

Koenderink, J. (2010), *Color for the Sciences*, The MIT Press, MA, USA.

Krakau, C. (1992), 'Calculation of the pulsatile ocular blood flow', *Investigative Ophthalmology & Visual Science* **33**(9), 2754–2756.

Kraukau, C. (1995), 'A model for pulsatile and steady ocular blood flow', *Graefe's Archive for Clinical and Experimental Ophthalmology* **233**, 112–118.

Kutzko, K., Brito, C. & Wall, M. (2000), 'Effect of instructions on conventional automated perimetry', *Investigative Ophthalmology & Visual Science* **41**(7), 2006–2013.

Kwon, Y., Fingert, J. & Greenlee, E. (2008), *A Patient's Guide to Glaucoma*, FEP International, IA, USA.

Kwon, Y., Fingert, J., Kuehnn, M. & Alward, W. (2009), 'Primary open-angle glaucoma', *The New England Journal of Medicine* **360**(11), 1113–1124.

Lawson, A. (2009), *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology*, Chapman & Hall/CRC Press, FL, USA.

Lawson, A. B. & Williams, F. L. (2001), *An Introductory Guide to Disease Mapping*, John Wiley & Sons, Ltd, West Sussex, UK.

Lawson, A., Browne, W. & Vidal Rodeiro, C. (2003), *Disease Mapping with WinBUGS and MLwiN*, John Wiley & Sons, Chichester, West Sussex.

LeBlanc, M. & Tibshirani, R. (1996), 'Combining estimates in regression and classification', *Journal of the American Statistical Association* **91**(436), 1641–1650.

Li, Y., Swift, S. & Tucker, A. (2013), 'Modelling and analysing the dynamics of disease progression from cross-sectional studies', *Journal of Biomedical Informatics* **46**(2), 226–274.

Liang, F., Paulo, R., Molina, G., Clyde, M. & Berger, J. O. (2008), 'Mixtures of g priors for Bayesian variabel selection', *Journal of the American Statistical Assosiation* **103**(481), 410–423.

Medeiros, F., Leite, M., Zangwill, L. & Weinreb, R. (2011), 'Combining structural and functional measurements to improve dectection of glaucoma progression using bayesian hierachical models', *Investigative Ophthalmology & Visual Science* **52**(8), 5794–5803.

Medeiros, F., Zangwill, L., Girkin, C., Liebmann, J. & Weinreb, R. (2012), 'Combining structural and functional measurements to improve estimates of rates of glaucomatous progression', *American Journal of Ophthalmology* **153**(6), 1197–1205.

Mederios, F., Zangwill, L., Bowd, C., Mansouri, K. & Weinreb, R. (2012), 'The structure function relationship in glaucoma: Implications for detection of progression and measurements of change', *Investigative Opthalmology & Visual Science* **53**(11), 6939–6946.

Mederios, F., Zangwill, L., Mansouri, K., Lisboa, R., Tafreshi, A. & Weinreb, R. (2012), 'Incorporating risk factors to improve the assessment of rates of glaucomatous progression', *Investigative Ophthalmology & Visual Science* **53**(4), 2199–2207.

Metropolis, N., Rosenbluth, A., Rosenbluth, M. & Teller, A. (1953), 'Equation of state calculations by fast computing machines', *The Journal of Chemical Physics* **21**(6), 1087–1092.

Mitchell, D. (1990), 'Color doppler imaging: Principles, limitations, and artifacts', *Radiology* **177**, 1–10.

Morgan, W., Hazelton, M., Chan, H., Azar, S., House, P., Cringle, S., Yu, D. & Balaratnasingam, C. (2004), 'Retinal venous pulsation in glaucoma and glaucoma suspects', *Ophthalmology* **111**, 1489–1494.

Morgan, W., Lind, C., Kain, S., Fatehee, N., Bala, A. & Yu, D. (2012), 'Retinal vein pulsation is in phase with intracranial pressure and not intraocular pressure', *Investigative Ophthalmology & Visual Science* **53**(8), 4676–4681.

Morgan, W., Yu, D. & Balaratnasingam, C. (2008), 'The role of cerebrospinal fluid pressure in glaucoma pathophysiology: the dark side of the optic disk', *Journal of Glaucoma* **17**(5), 408–413.

Neelon, B. & Dunson, D. (2004), 'Bayesian isotonic regression and trend analysis', *Biometrics* **60**, 398–406.

Paul, M. & Molla, M. (2012), 'Investigation of physiological pulsatile flow in moel arterial stenosis using large-eddy and direct numerical simulations', *Applied Mathmatical Modelling* **36**, 4393–4413.

Pearson, K. (1907), 'On the influence of experience on future expectation', *Philosophical Magazine* **13**, 365–378.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. & R Core Team (2013), *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-109.

Plummer, M., Best, N., Cowles, K. & Vines, K. (2006), 'CODA: Convergence diagnosis and output analysis for MCMC', *R News* **6**(1), 7–11.

R Core Team (2013), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: *http://www.R-project.org*

Radke, R., Andra, S., Al-Kofahi, O. & Roysam, B. (2005), 'Image change detection algorithms: A systematic survey', *IEEE Transactions on image processing* **14**(3), 294–307.

Ramirez, J., Mitra, S. & Morales, J. (1999), 'Visualization of the three-dimensional topography of the optic nerve head through a passive stereo vision model', *Journal of Electronic Imaging* **8**(1), 92–97.

Resnikoff, S., Pascolini, D., Etya'ale, D., Kocur, I., Pararajasegaram, R., Pokharel, G. & Mariotti, S. (2004), 'Global data on visual impairment in the year 2002', *Bulletin of the World Health Organization* **82**(11), 844–855.

Reus, J. & Lemij, H. (2005), 'Relationships between standard automated perimetry, HRT confocal scanning laser ophthalmoscopy, and GDx VCC scanning laser polimetry', *Investigative Ophthalmology & Visual Science* **46**(11), 4182–4188.

Ritz, C. & Streibig, J. C. (2008), *Nonlinear Regression with R*, Springer, NY,USA.

Rivera, J., Bell, N. & Feldman, R. (2008), 'Risk factors for primary open angle glaucoma progression: what we know and what we need to know', *Current Opinion in Ophthalmology* **19**, 102–106.

Robertson, M., Kohler, U., Hoskins, P. & Marshall, I. (2001), 'Quantitative analysis of pc mri velocity maps: pulsatile flow in cylindrical vessels', *Magnetic Resonance Imaging* **19**, 685–695.

Rudnicka, A., Jarrar, Z., Wormald, R., Cook, D., Fletcher, A. & Owen, C. (2012), 'Age and gender variations in are-related macular degeneration prevalence in populations of european ancestry: A meta-analysis', *Ophthalmology* **119**(3), 571–580.

Rudnicka, A. R., Mt-Isa, S., Owen, C. G., Cook, D. G. & Ashby, D. (2006), 'Variations in primary open-angle glaucoma prevalence by age, gender, and race: A bayesian meta-analysis', *Investigative Ophthalmology & Visual Science* **47**(10), 4254–4261.

Ruppert, D. & Carroll, R. (2002), 'Selecting the number of knots for penalised splines', *Journal of Computational and Graphical Statistics* **11**, 735–757.

Ruppert, D., Wand, M. & Carroll, R. (2003), *Semiparametric Regression*, Cambridge University Press, Cambridge, UK.

Russell, R., Malik, R., Chauhan, B., Crabb, D. & Garway-Heath, D. (2012), 'Improved estimates of visual field progression using bayesian linear regression to integrate structural infromation in patients with ocular hypertension', *Investigative Ophthalmology & Visual Science* **53**(6), 2760–2769.

Sample, P., Boden, C., Zhang, Z., Pascual, J., Lee, T., Zangwill, L., Weinreb, Crowston, J., Hoffmann, E., Medrios, F., Sejnowski, T. & Goldbaum, M. (2005), 'Unsupervised machine

learning with independent component analysis to identify areas of progression in glaucomatous visual fields', *Investigative Ophthalmology & Visual Science* **46**(10), 3684–3692.

Sample, P., Goldbaum, M., Chan, K., Boden, C., Lee, T., Vasile, C., Boehm, A., Sejnowski, T., Johnson, C. & Weinreb, R. (2002), 'Using machine learning classifiers to identify glaucomatous change earlier in standard visual fields', *Investigative Ophthalmology & Visual Science* **43**(8), 2660–2665.

Sanfilippo, P., Cardini, A., Hewitt, A., Crowaton, J. & Mackey, D. (2009), 'Optic disc morphology - rethinking shape', *Progress in Retinal and Eye Research* **28**, 227–248.

Shively, T., Sager, T. & Walker, S. (2009), 'A Bayesian approach to non-parametric monotone function estimation', *Journal of the Royal Statistical Society series B* **71**(1), 159–175.

Sims, M., Cox, T. & Lewison, R. (2008), 'Modeling spatial patterns in fisheries bycatch: improving bycatch maps to aid fisheries management', *Ecological Applications* **18**(3), 649–661.

Smith, S., Katz, J. & Quigley, H. (1996), 'Analysis of progressive change in automated visual fields in glaucoma', *Investigative Ophthalmology & Visual Science* **37**, 1419–1428.

Snow, J. (1936), *Snow on cholera: being a reprint of two papers*, The Commonwealth Fund, London.

Speed, T. (1991), 'Comment on 'that BLUP is a good thing: The estimation of random effects", *Statistical Science* **6**, 42–44.

Spiegelhalter, D., Best, N., Carlin, B. & van der Linde, A. (2002), 'Bayesian measures of model complexity and fit', *Journal of the Royal Statistical Society Series B - Statistical Methodology* **64**(4), 583–616.

Spry, P., Johnson, C., Bates, A., Turpin, A. & Chauhan, B. (2002), 'Spatial and temporal processing of threshold data for detection of progressive glaucomatous visual field loss', *Archives of Ophthalmology* **120**, 173–180.

Spry, P., Johnson, C., McKendrick, A. & Turpin, A. (2003), 'Measurement error of visual field tests in glaucoma', *British Journal of Ophthalmology* **87**, 107–112.

Stade, E. (2005), *Fourier Analysis*, John Wiley & Sons, NJ, USA.

Stalman, I., Harris, A., Fieuws, S., Zeyen, T., Vanbellinghen, V., McCranor, L. & Siesky, B. (2009), 'Color doppler imaging and occular pulse amplitude in glaucomatous and healthy eyes', *European Journal of Ophthalmology* **19**(4), 580–587.

Strouthidis, N., Scott, A., Peter, N. & Garway-Heath, D. (2010), 'Optic disk and visual field progression in ocular hypertensive subjects:detection rates, specificity and agreement', *Investigative Ophthalmology* **47**, 2904–2910.

Strouthidis, N., Scott, A., Viswanathan, A. C., Crabb, A. & Garway-Heath, D. (2007), 'Monitoring glaucomatous visual field progression: The effect of a novel spatial filter', *Investigative Ophthalmology & Visual Science* **48**(1), 251–257.

Trajkovic, I., Scholkmann, F. & Wolf, M. (2011), 'Estimating and validating the interbeat intervals of the heart using near-infrared spectroscopy on the human forehead', *Journal of Biomedical Optics* **16**(8), 087002.

Turpin, A. & McKendrick, A. (2011), 'What reduction in standard automated perimetry variability would lmprove the detection of visual field progression', *Investigative Ophthalmology & Visual Science* **52**(6), 3237–3245.

Urbanek, S. (2013), *jpeg: Read and write JPEG images.* R package version 0.1-4.
URL: *http://CRAN.R-project.org/package=jpeg*

vanKampen, E. & Zijlstra, W. (1965), Determination of hemoglobin and its derivatives, *in* H. Sobotka & C. Stewart, eds, 'Advances in Clinical Chemistry', Academic Press, NY,NY, pp. 141–187.

Verkruysse, W., Svaasand, L. & Nelson, J. (2008), 'Remote plethysmographic imaging using ambient light', *Opt. Express* **16**(26), 21434–21445.

Vesti, E., Johnson, C. & Chauhan, B. (2003), 'Comparison of different methods for detecting glaucomatous visual feild progression', *Investigative Ophthalmology and Visual Science* **44**(9), 3873–3879.

Villalobos, M. & Wahba, G. (1987), 'Inequality-constrained multivariate smoothing splines with application to the estimation of posterior probabilities', *Journal of the American Statistical Association* **82**(397), 239–248.

Viswanathan, A., Fitzke, F. & Hitchings, R. (1997), 'Early detection of visual filed progression in glaucoma: a comparison of PROGRESSOR and STATPAC2', *British Journal of Ophthalmology* **81**, 1037–1042.

Wakefield, J. (2007), 'Disease mapping and spatial regression with count data', *Biostatistics* **8**(2), 158–183.

Wall, M., Woodward, K., Doyle, C. & Artes, P. (2009), 'Repeatability of automated perimetry: A comparison between standard automated perimetry with stimulus size III and V, matrix and motion perimetry', *Investigative Ophthalmology & Visual Science* **50**, 974–979.

Walsh, T. J. (2011), *Visual Fileds: Examination and Interpretation*, 3rd edn, Oxford University Press, New York.

Wand, M. (2003), 'Smoothing and mixed models', *Computational Statisitcs* **18**, 223–249.

Wang, C., Quddus, M. & Ison, S. G. (2009), 'Impact of traffic congestion on road accidents: A spatial analysis of the M25 motorway in England', *Accident Analysis and Prevention* **41**, 798–808.

Wild, J., Searle, A., Dengler Harles, H. & O'Neill, E. (1991), 'Long-term follow-up of baseline learning fatigue effects in automated perimetry of glaucoma and ocular hypertensive patients', *Acta Ophthalmology* **69**, 210–216.

Wright, I. & Wegman, E. (1980), 'Isotonic, convex and related splines', *Annals of Statistics* **8**, 1023–1035.

Yang, X., Piety, N., Vignes, S., Benton, M., Kanter, J. & Shevkoplyas, S. (2013), 'Simple paper-based test for measuring blood hemoglobin concentration in resource-limited settings', *Clinical Chemistry* **In press**.

Yoko, A. & Walter, M. (2013), 'Genetics and environmental stress factor contributions to anterior segment malformations in glaucoma'. http://www.intechopen.com/books/glaucoma-basic-and-clinical-aspects/ngenetics-and-environmental-stress-factor-contributions-to-anterior-segment-malformations-and-glauco.

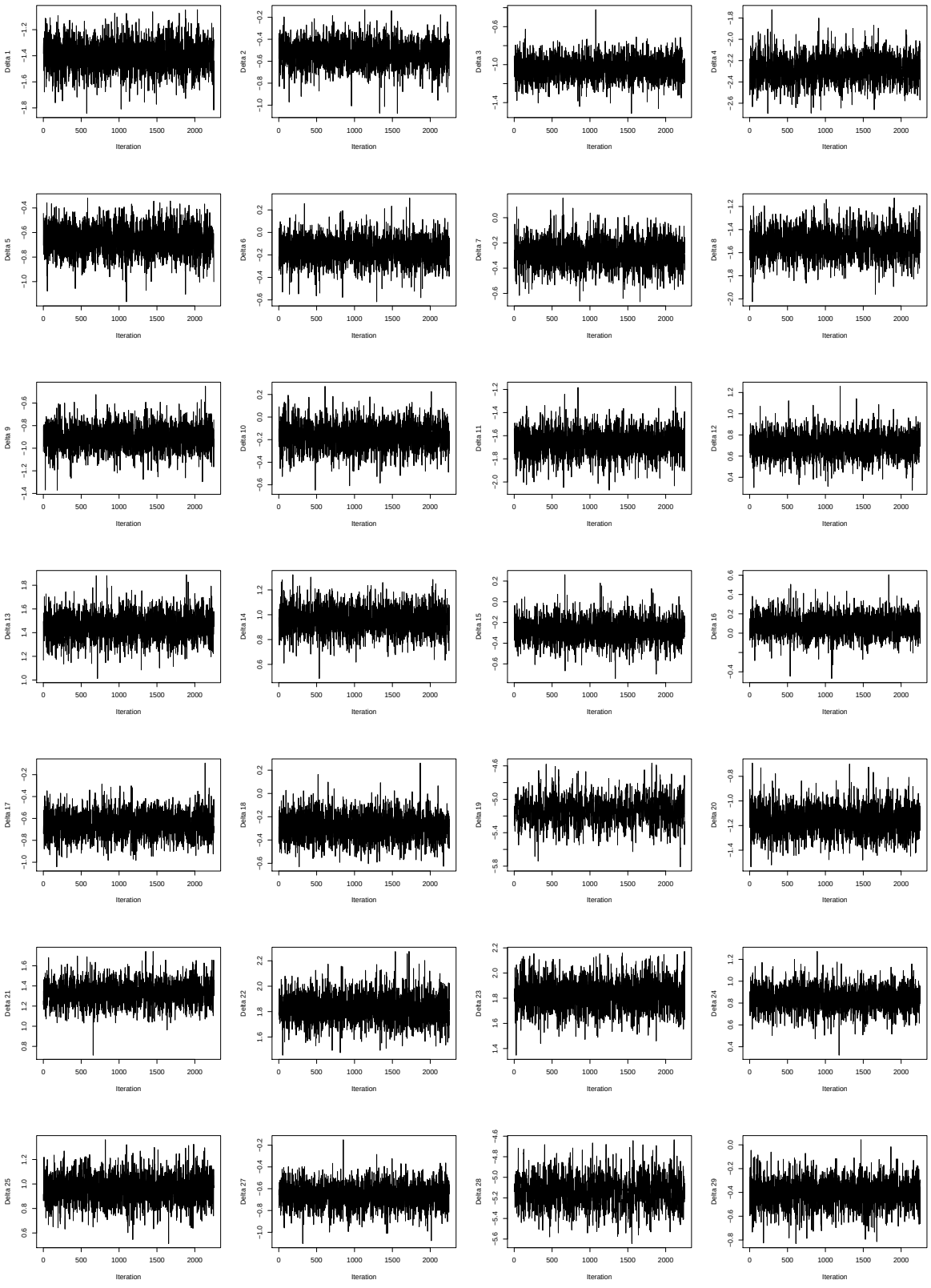
Zhu, H., Crabb, D., Fredette, M., Anderson, D. & Garway-Heath, D. (2011), 'Quantifying discordance between structure and function measurments in the clinical assessment of glaucoma', *Archive of Ophthalmology* **129**(9), 1167–1174.

Zhu, H., Crabb, D., Schlottmann, P., Lemij, H., Reus, N., Healy, P., Mitchell, P. Nad Ho, T. & Garway-Heath, D. (2010), 'Predicting visual function from the measurements of retinal nerve fiber layer structure', *Investigative Ophthalmology & Visual Science* **51**, 5657–5666.

Zulauf, M. & Caprioli, J. (1992), 'Unknown title', *Seminars in Ophthalmology* **7**(2), 130–146.

Appendix A

SPROG Convergence Plots


Delta 1
Delta 2
Delta 3
Delta 4
Delta 5
Delta 6
Delta 7
Delta 8
Delta 9
Delta 10
Delta 11
Delta 12
Delta 13
Delta 14
Delta 15
Delta 16
Delta 17
Delta 18
Delta 19
Delta 20
Delta 21
Delta 22
Delta 23
Delta 24
Delta 25
Delta 27
Delta 28
Delta 29
Iteration

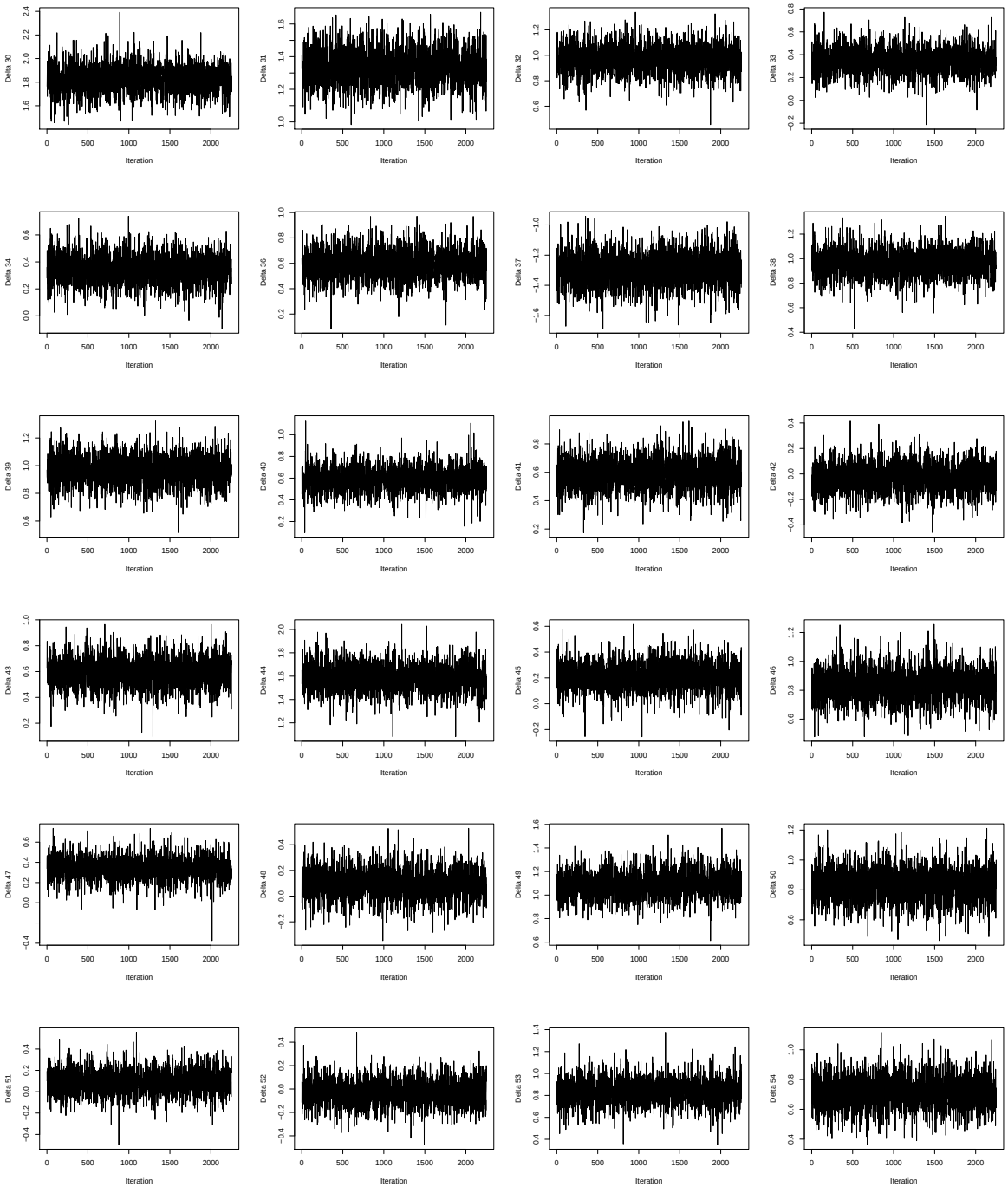


Figure A.1: Trace plots for δ_i's for a typical progressing eye.

Appendix B

Parameters for Stable Eye

Table B.1: Slope parameters ($\beta + \eta_i$) for a typical stable eye.

	Mean	95% C.I.	Bayesian P-value
1	0.08	(-0.25,0.47)	0.3507
2	0.08	(-0.23,0.45)	0.3431
3	0.09	(-0.23,0.48)	0.3284
4	0.08	(-0.24,0.47)	0.3511
5	0.08	(-0.24,0.47)	0.32
6	0.11	(-0.18,0.47)	0.276
7	0.11	(-0.19,0.47)	0.2653
8	0.1	(-0.18,0.45)	0.2916
9	0.09	(-0.21,0.46)	0.3187
10	0.08	(-0.22,0.47)	0.3098
11	0.1	(-0.2,0.45)	0.2827
12	0.1	(-0.18,0.45)	0.2671
13	0.1	(-0.18,0.46)	0.264
14	0.11	(-0.16,0.44)	0.2462
15	0.1	(-0.18,0.45)	0.2698
16	0.1	(-0.18,0.46)	0.2653
17	0.09	(-0.21,0.46)	0.3129
18	0.09	(-0.25,0.5)	0.3329
19	0.08	(-0.25,0.47)	0.34
20	0.08	(-0.22,0.45)	0.3213
21	0.09	(-0.2,0.45)	0.2916
22	0.09	(-0.21,0.46)	0.3071
23	0.1	(-0.19,0.47)	0.2773
24	0.1	(-0.18,0.46)	0.2716
25	0.1	(-0.19,0.46)	0.276
26			
27	0.07	(-0.24,0.43)	0.3684
28	0.09	(-0.26,0.49)	0.3316
29	0.09	(-0.22,0.46)	0.3067
30	0.1	(-0.21,0.46)	0.2911
31	0.09	(-0.2,0.46)	0.3116
32	0.09	(-0.2,0.45)	0.292
33	0.1	(-0.18,0.46)	0.2733
34	0.09	(-0.2,0.45)	0.2871
35			
36	0.08	(-0.24,0.44)	0.3382
37	0.09	(-0.21,0.47)	0.2996
38	0.09	(-0.19,0.44)	0.2978
39	0.09	(-0.2,0.44)	0.3022
40	0.1	(-0.19,0.45)	0.2853
41	0.09	(-0.21,0.44)	0.2978
42	0.1	(-0.22,0.46)	0.2813
43	0.08	(-0.23,0.46)	0.3191
44	0.07	(-0.25,0.46)	0.356
45	0.1	(-0.2,0.47)	0.2733
46	0.09	(-0.21,0.47)	0.2947
47	0.09	(-0.2,0.45)	0.2898
48	0.09	(-0.19,0.45)	0.3004
49	0.09	(-0.19,0.44)	0.3062
50	0.07	(-0.23,0.44)	0.3533
51	0.08	(-0.23,0.45)	0.3311
52	0.08	(-0.23,0.44)	0.3138
53	0.08	(-0.22,0.44)	0.3418
54	0.07	(-0.24,0.45)	0.3431

Table B.2: Loci mean ($\alpha + \delta_i$) for a typical stable eye.

	Mean	95% C.I.	Bayesian P-value
1	27.63	(27.39,27.87)	<0.0001
2	28.49	(28.26,28.73)	<0.0001
3	28	(27.76,28.25)	<0.0001
4	26.76	(26.5,27.02)	<0.0001
5	28.37	(28.12,28.61)	<0.0001
6	28.88	(28.65,29.12)	<0.0001
7	28.76	(28.53,28.98)	<0.0001
8	27.52	(27.27,27.78)	<0.0001
9	28.13	(27.9,28.35)	<0.0001
10	28.87	(28.62,29.12)	<0.0001
11	27.38	(27.13,27.61)	<0.0001
12	29.74	(29.5,29.97)	<0.0001
13	30.49	(30.25,30.72)	<0.0001
14	29.99	(29.75,30.22)	<0.0001
15	28.75	(28.51,28.98)	<0.0001
16	29.12	(28.87,29.35)	<0.0001
17	28.37	(28.15,28.61)	<0.0001
18	28.75	(28.5,28.98)	<0.0001
19	23.9	(23.57,24.21)	<0.0001
20	27.87	(27.63,28.1)	<0.0001
21	30.36	(30.12,30.6)	<0.0001
22	30.87	(30.63,31.11)	<0.0001
23	30.86	(30.62,31.11)	<0.0001
24	29.87	(29.63,30.11)	<0.0001
25	29.99	(29.76,30.24)	<0.0001
26			
27	28.38	(28.13,28.63)	<0.0001
28	23.9	(23.56,24.24)	<0.0001
29	28.63	(28.38,28.87)	<0.0001
30	30.86	(30.6,31.11)	<0.0001
31	30.37	(30.13,30.61)	<0.0001
32	30	(29.76,30.23)	<0.0001
33	29.38	(29.14,29.62)	<0.0001
34	29.38	(29.15,29.61)	<0.0001
35			
36	29.62	(29.37,29.87)	<0.0001
37	27.75	(27.5,28)	<0.0001
38	30	(29.76,30.24)	<0.0001
39	30	(29.76,30.23)	<0.0001
40	29.63	(29.4,29.86)	<0.0001
41	29.62	(29.39,29.85)	<0.0001
42	29	(28.77,29.25)	<0.0001
43	29.63	(29.39,29.86)	<0.0001
44	30.62	(30.36,30.89)	<0.0001
45	29.25	(29.02,29.49)	<0.0001
46	29.87	(29.64,30.1)	<0.0001
47	29.38	(29.14,29.61)	<0.0001
48	29.13	(28.9,29.37)	<0.0001
49	30.12	(29.9,30.36)	<0.0001
50	29.87	(29.63,30.12)	<0.0001
51	29.12	(28.88,29.38)	<0.0001
52	29	(28.75,29.24)	<0.0001
53	29.87	(29.63,30.11)	<0.0001
54	29.75	(29.51,29.99)	<0.0001

Appendix C

Parameters for progressing eye

Table C.1: Slope parameters ($\beta + \eta_i$) for a typical progressing eye.

	Mean	95% C.I.	Bayesian P-value
1	-1.73	(-5.68,0.49)	0.0951
2	-1.67	(-5.55,0.39)	0.0911
3	-1.45	(-5.35,0.67)	0.1578
4	-1.58	(-5.4,0.46)	0.1089
5	-1.65	(-5.42,0.45)	0.1004
6	-1.79	(-5.74,0.32)	0.0827
7	-1.74	(-5.65,0.39)	0.0982
8	-1.68	(-5.63,0.4)	0.0991
9	-1.53	(-5.41,0.47)	0.1133
10	-1.49	(-5.44,0.45)	0.1116
11	-1.7	(-5.77,0.46)	0.1133
12	-1.73	(-5.69,0.3)	0.0756
13	-1.75	(-5.61,0.28)	0.0756
14	-1.71	(-5.66,0.3)	0.0822
15	-1.63	(-5.58,0.39)	0.1022
16	-1.59	(-5.41,0.39)	0.0929
17	-1.42	(-5.29,0.48)	0.1222
18	-1.38	(-5.27,0.62)	0.1262
19	-1.45	(-5.51,0.73)	0.1613
20	-1.51	(-5.42,0.55)	0.1347
21	-1.6	(-5.52,0.38)	0.1053
22	-1.55	(-5.6,0.47)	0.116
23	-1.73	(-5.69,0.36)	0.0916
24	-1.69	(-5.55,0.33)	0.088
25	-1.56	(-5.5,0.58)	0.1249
26			
27	-1.36	(-5.31,0.51)	0.1209
28	-1.4	(-5.37,0.79)	0.16
29	-1.57	(-5.46,0.58)	0.1298
30	-1.62	(-5.6,0.45)	0.1196
31	-1.63	(-5.57,0.46)	0.1107
32	-1.6	(-5.41,0.46)	0.1124
33	-1.54	(-5.27,0.47)	0.12
34	-1.47	(-5.26,0.51)	0.1342
35			
36	-1.29	(-5.26,0.62)	0.1533
37	-1.43	(-5.33,0.53)	0.1498
38	-1.55	(-5.47,0.44)	0.1133
39	-1.58	(-5.55,0.43)	0.1102
40	-1.59	(-5.4,0.49)	0.1182
41	-1.55	(-5.48,0.47)	0.12
42	-1.43	(-5.32,0.61)	0.1489
43	-1.41	(-5.25,0.54)	0.1302
44	-1.19	(-4.92,0.75)	0.1693
45	-1.47	(-5.38,0.49)	0.1204
46	-1.47	(-5.32,0.51)	0.1258
47	-1.49	(-5.52,0.48)	0.1231
48	-1.46	(-5.28,0.48)	0.1267
49	-1.43	(-5.35,0.5)	0.1236
50	-1.3	(-5.04,0.62)	0.1542
51	-1.36	(-5.37,0.65)	0.1524
52	-1.36	(-5.16,0.65)	0.1564
53	-1.36	(-5.3,0.53)	0.1369
54	-1.29	(-5.15,0.53)	0.1311

Table C.2: Loci Mean ($\alpha + \delta_i$) for a typical progressing eye.

	Mean	95% C.I.	Bayesian P-value
1	12.49	(11.62,13.34)	<0.0001
2	16.34	(15.43,17.22)	<0.0001
3	13.83	(12.88,14.67)	<0.0001
4	16	(15.1,16.88)	<0.0001
5	13.65	(12.8,14.59)	<0.0001
6	19.66	(18.77,20.58)	<0.0001
7	18.83	(18.01,19.7)	<0.0001
8	19.34	(18.46,20.18)	<0.0001
9	18.81	(17.94,19.62)	<0.0001
10	20.65	(19.75,21.55)	<0.0001
11	1.4	(0.48,2.36)	0.004
12	12.14	(11.26,13.02)	<0.0001
13	15.29	(14.37,16.15)	<0.0001
14	18.47	(17.61,19.33)	<0.0001
15	17.32	(16.48,18.17)	<0.0001
16	13.84	(12.94,14.72)	<0.0001
17	17.16	(16.3,18.01)	<0.0001
18	19.01	(18.13,19.92)	<0.0001
19	2.56	(1.65,3.47)	<0.0001
20	2.85	(1.94,3.75)	<0.0001
21	1.2	(0.22,2.17)	0.0093
22	4.53	(3.63,5.44)	<0.0001
23	0.28	(-0.84,1.45)	0.3138
24	4	(3.04,5)	<0.0001
25	-4.33	(-5.84,-2.79)	<0.0001
26			
27	21.84	(20.96,22.71)	<0.0001
28	10.57	(9.67,11.49)	<0.0001
29	18.65	(17.77,19.54)	<0.0001
30	19.99	(19.14,20.84)	<0.0001
31	24.48	(23.65,25.31)	<0.0001
32	30.34	(29.33,31.29)	<0.0001
33	30.24	(29.23,31.3)	<0.0001
34	27.62	(26.67,28.58)	<0.0001
35			
36	25.83	(24.95,26.74)	<0.0001
37	18.66	(17.78,19.56)	<0.0001
38	23.68	(22.84,24.59)	<0.0001
39	26.82	(25.96,27.69)	<0.0001
40	27.67	(26.8,28.57)	<0.0001
41	28.16	(27.29,29.05)	<0.0001
42	28.15	(27.27,29.02)	<0.0001
43	25.67	(24.75,26.58)	<0.0001
44	26.65	(25.67,27.55)	<0.0001
45	23.50	(22.62,24.38)	<0.0001
46	25.33	(24.45,26.19)	<0.0001
47	24.51	(23.67,25.39)	<0.0001
48	26.84	(25.98,27.69)	<0.0001
49	28.50	(27.6,29.36)	<0.0001
50	25.82	(24.94,26.75)	<0.0001
51	24.65	(23.77,25.52)	<0.0001
52	23.00	(22.09,23.86)	<0.0001
53	22.99	(22.11,23.82)	<0.0001
54	21.19	(20.3,22.07)	<0.0001

Table C.3: Overall mean and slope parameters for all 194 eyes.

	α	95 % C.I.	P	β	95 % C.I.	P
1	28.3	(28.18,28.43)	<0.0001	-0.35	(-0.91,0)	0.0236
2	24.56	(24.25,24.87)	<0.0001	-0.81	(-2.33,-0.02)	0.0151
3	23.58	(23.29,23.87)	<0.0001	-0.79	(-2.25,-0.03)	0.0147
4	19.38	(19.03,19.72)	<0.0001	-0.81	(-2.32,-0.02)	0.016
5	21.67	(21.39,21.96)	<0.0001	-1.36	(-3.99,-0.06)	0.0036
6	24.98	(24.84,25.11)	<0.0001	-0.44	(-1.14,-0.01)	0.016
7	31.42	(31.32,31.52)	<0.0001	-0.34	(-0.85,0)	0.0222
8	19.68	(19.21,20.13)	<0.0001	-0.47	(-1.69,0.17)	0.1262
9	12.6	(12.09,13.06)	<0.0001	-0.66	(-1.98,0.03)	0.0378
10	21.18	(20.91,21.45)	<0.0001	-0.56	(-1.56,0.01)	0.0364
11	25.97	(25.66,26.24)	<0.0001	-0.64	(-1.87,0.02)	0.0373
12	29.45	(29.29,29.62)	<0.0001	-0.79	(-2.02,-0.03)	0.0116
13	28.15	(28.05,28.26)	<0.0001	-0.43	(-1.04,-0.02)	0.0062
14	20.21	(19.81,20.57)	<0.0001	-1.17	(-3.39,-0.03)	0.0098
15	29.29	(29.17,29.41)	<0.0001	-0.01	(-0.21,0.15)	0.464
16	27.21	(27.01,27.4)	<0.0001	-0.46	(-1.33,0.01)	0.0387
17	21.94	(21.68,22.19)	<0.0001	-0.84	(-2.33,-0.03)	0.0071
18	22.01	(21.71,22.3)	<0.0001	-0.86	(-2.53,-0.03)	0.0107
19	27.53	(27.29,27.77)	<0.0001	0.2	(-0.13,0.73)	0.8378
20	19.69	(19.15,20.14)	<0.0001	-0.18	(-0.81,0.3)	0.2489
21	25.79	(25.62,25.97)	<0.0001	-0.66	(-1.75,-0.02)	0.0133
22	23.27	(23.03,23.51)	<0.0001	-0.85	(-2.29,-0.04)	0.0071
23	26.14	(26.02,26.25)	<0.0001	-0.1	(-0.41,0.04)	0.1524
24	22.9	(22.66,23.15)	<0.0001	0.19	(-0.11,0.73)	0.8444
25	27.07	(26.81,27.34)	<0.0001	-0.51	(-1.51,0.03)	0.0493
26	28.75	(28.57,28.93)	<0.0001	-0.69	(-1.77,-0.01)	0.0147
27	30.13	(29.97,30.31)	<0.0001	0.06	(-0.18,0.38)	0.6916
28	28.61	(28.42,28.81)	<0.0001	-0.71	(-1.87,-0.02)	0.0142
29	26.25	(25.99,26.53)	<0.0001	-0.03	(-0.44,0.38)	0.4342
30	21.79	(21.35,22.21)	<0.0001	-1.06	(-3.29,0)	0.0249
31	22.26	(22.03,22.49)	<0.0001	-0.77	(-2.1,-0.02)	0.0133
32	17.07	(16.62,17.48)	<0.0001	-0.01	(-0.59,0.56)	0.5084
33	31.06	(30.96,31.16)	<0.0001	-0.25	(-0.66,0.01)	0.0373
34	30.18	(30.05,30.31)	<0.0001	-0.65	(-1.6,-0.01)	0.0116
35	17.55	(16.71,18.19)	<0.0001	-0.35	(-1.27,0.18)	0.1507
36	14.61	(13.41,15.55)	<0.0001	-0.47	(-1.62,0.26)	0.1342
37	21.74	(21.37,22.13)	<0.0001	-0.44	(-1.57,0.15)	0.1218
38	28.25	(28.09,28.39)	<0.0001	-0.07	(-0.4,0.14)	0.2787
39	27	(26.75,27.25)	<0.0001	-0.65	(-1.85,-0.02)	0.0156
40	27.51	(27.28,27.75)	<0.0001	0.01	(-0.33,0.38)	0.5404
41	31.01	(30.78,31.23)	<0.0001	-0.04	(-0.43,0.31)	0.4053
42	31.48	(31.26,31.71)	<0.0001	-0.05	(-0.46,0.28)	0.3627
43	29.17	(29,29.34)	<0.0001	0.69	(0,1.83)	0.9764
44	30.5	(30.31,30.69)	<0.0001	-0.23	(-0.75,0.06)	0.1049
45	28.25	(28.07,28.43)	<0.0001	-0.45	(-1.23,0)	0.0276
46	26.54	(26.28,26.81)	<0.0001	0.38	(-0.06,1.2)	0.9298
47	22.14	(21.81,22.46)	<0.0001	-0.3	(-1.04,0.14)	0.1396
48	16.64	(16.18,17.08)	<0.0001	-0.44	(-1.57,0.2)	0.1369

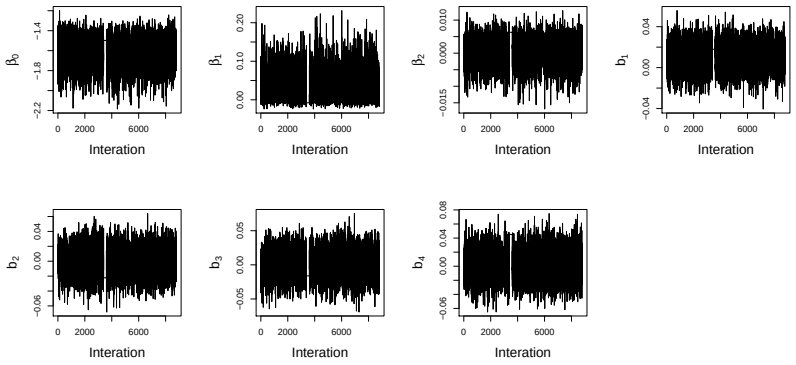
	α	95 % C.I.	P	β	95 % C.I.	P
49	-41.33	(-130.26,1.09)	0.9489	-2.6	(-18.35,9.73)	0.1831
50	7.58	(6.54,8.49)	<0.0001	-1.41	(-4.93,-0.07)	0.0107
51	28.49	(28.3,28.68)	<0.0001	0.35	(-0.02,1)	0.9533
52	23.74	(23.41,24.07)	<0.0001	-0.01	(-0.53,0.48)	0.4853
53	26.58	(26.23,26.96)	<0.0001	-0.42	(-1.35,0.1)	0.0818
54	27.62	(27.31,27.92)	<0.0001	-0.65	(-2.01,0.01)	0.0289
55	29.04	(28.92,29.17)	<0.0001	0.09	(-0.09,0.39)	0.8569
56	12.91	(12.19,13.55)	<0.0001	-0.37	(-1.47,0.32)	0.188
57	26.52	(26.21,26.83)	<0.0001	-0.85	(-2.47,-0.02)	0.0151
58	18.4	(17.95,18.86)	<0.0001	-1.49	(-5.14,-0.07)	0.0076
59	15.38	(14.83,15.85)	<0.0001	-0.44	(-1.43,0.07)	0.0711
60	6.39	(5.21,7.46)	<0.0001	-0.92	(-3.01,0.07)	0.0498
61	29.86	(29.64,30.08)	<0.0001	0.41	(-0.03,1.22)	0.9413
62	30.33	(30.15,30.5)	<0.0001	0.07	(-0.2,0.43)	0.6862
63	23.89	(23.59,24.19)	<0.0001	-0.44	(-1.38,0.04)	0.064
64	-33.75	(-74.38,-0.17)	0.9751	-1.58	(-17.62,12.6)	0.2871
65	27.4	(27.19,27.62)	<0.0001	0.18	(-0.11,0.66)	0.8471
66	17.48	(16.52,18.32)	<0.0001	0.46	(-0.42,1.86)	0.8062
67	15.84	(15.53,16.15)	<0.0001	-0.82	(-2.36,-0.02)	0.0138
68	19.07	(18.81,19.33)	<0.0001	-1.09	(-3.09,-0.05)	0.0053
69	26.48	(26.24,26.72)	<0.0001	-0.67	(-1.88,-0.01)	0.0218
70	26.75	(26.59,26.92)	<0.0001	-0.63	(-1.57,-0.03)	0.0089
71	29.37	(29.23,29.51)	<0.0001	-0.05	(-0.33,0.14)	0.3022
72	29.39	(29.27,29.51)	<0.0001	-0.44	(-1.12,-0.01)	0.0164
73	18.09	(17.73,18.44)	<0.0001	1.12	(0.02,3.46)	0.9853
74	23.92	(23.66,24.18)	<0.0001	1.42	(0.06,4.1)	0.9942
75	27.46	(27.3,27.64)	<0.0001	0.21	(-0.06,0.69)	0.8796
76	25.19	(24.93,25.45)	<0.0001	-0.05	(-0.5,0.36)	0.4187
77	28.6	(28.46,28.74)	<0.0001	-0.17	(-0.59,0.04)	0.1084
78	27.3	(27.11,27.48)	<0.0001	-0.65	(-1.75,-0.02)	0.0116
79	26.91	(26.61,27.21)	<0.0001	-0.35	(-1.22,0.1)	0.1049
80	29.76	(29.58,29.93)	<0.0001	-0.03	(-0.35,0.22)	0.4329
81	21.76	(21.3,22.19)	<0.0001	-0.65	(-2.13,0.07)	0.0627
82	18.62	(18.04,19.15)	<0.0001	0.07	(-0.54,0.76)	0.5964
83	22.34	(21.7,22.95)	<0.0001	0.46	(-0.3,1.58)	0.8587
84	29.33	(29.12,29.55)	<0.0001	-0.05	(-0.41,0.25)	0.368
85	11.68	(10.46,12.79)	<0.0001	0.77	(-0.14,2.61)	0.9156
86	25.49	(25.23,25.75)	<0.0001	-0.01	(-0.42,0.38)	0.4853
87	11.08	(10.49,11.64)	<0.0001	-1.09	(-3.86,0.03)	0.036
88	22.8	(22.61,22.99)	<0.0001	1.02	(0.02,2.73)	0.9907
89	28.55	(28.4,28.7)	<0.0001	-0.13	(-0.5,0.07)	0.1644
90	28.41	(28.26,28.57)	<0.0001	-0.87	(-2.15,-0.04)	0.004
91	18.77	(18.31,19.22)	<0.0001	-1.02	(-3.53,0.05)	0.048
92	25.42	(25.21,25.63)	<0.0001	-0.58	(-1.62,0.02)	0.0418
93	22.97	(22.67,23.26)	<0.0001	-0.78	(-2.33,0)	0.0227
94	25.98	(25.77,26.2)	<0.0001	-1.02	(-2.69,-0.04)	0.0067
95	21.18	(20.88,21.47)	<0.0001	-0.35	(-1.15,0.07)	0.0893
96	17.27	(16.8,17.7)	<0.0001	-0.7	(-2.16,0.03)	0.0422
97	27.94	(27.65,28.22)	<0.0001	-0.34	(-1.06,0.08)	0.0991
98	29.39	(29.28,29.5)	<0.0001	-0.28	(-0.78,0)	0.0338

	α	95 % C.I.	P	β	95 % C.I.	P
99	14.7	(14.31,15.09)	<0.0001	-1.5	(-4.81,-0.08)	0.0031
100	25.06	(24.84,25.27)	<0.0001	-0.41	(-1.18,0.03)	0.0564
101	14.3	(13.81,14.77)	<0.0001	-0.7	(-2.28,0.05)	0.0511
102	30.56	(30.45,30.67)	<0.0001	-0.39	(-0.98,-0.01)	0.0138
103	23.72	(23.48,23.94)	<0.0001	-0.58	(-1.57,-0.02)	0.0124
104	25.68	(25.58,25.78)	<0.0001	-0.73	(-1.73,-0.04)	0.004
105	24.27	(24.04,24.51)	<0.0001	-0.65	(-1.9,0)	0.0258
106	18.54	(18.27,18.8)	<0.0001	0.96	(0.04,2.66)	0.9947
107	13.69	(13.32,14.07)	<0.0001	-0.42	(-1.39,0.11)	0.0996
108	22.99	(22.51,23.34)	<0.0001	-0.12	(-0.59,0.2)	0.2431
109	18.13	(17.49,18.71)	<0.0001	-1.76	(-6.54,-0.15)	0.0044
110	-8.42	(-27.79,5.12)	0.8044	-0.59	(-5.77,3.87)	0.2804
111	-12.32	(-16.87,-8.74)	1	-0.75	(-4.51,2.95)	0.3356
112	24.59	(24.37,24.81)	<0.0001	-0.56	(-1.52,0)	0.0244
113	27.13	(26.84,27.41)	<0.0001	-1.53	(-4.39,-0.07)	0.0018
114	21.31	(20.89,21.7)	<0.0001	-1.07	(-3.33,-0.01)	0.0173
115	20.95	(20.58,21.3)	<0.0001	-0.41	(-1.42,0.12)	0.1031
116	14.1	(13.63,14.53)	<0.0001	-0.49	(-1.59,0.12)	0.0956
117	28.49	(28.2,28.78)	<0.0001	0.27	(-0.16,0.95)	0.8671
118	22.92	(22.54,23.29)	<0.0001	0.07	(-0.45,0.65)	0.5991
119	27.48	(27.33,27.64)	<0.0001	0.64	(0.03,1.65)	0.9907
120	27.12	(26.9,27.35)	<0.0001	0.38	(-0.04,1.13)	0.9427
121	13.31	(12.58,13.97)	<0.0001	-1.04	(-3.81,0.08)	0.0618
122	21.67	(21.29,22.03)	<0.0001	-0.47	(-1.5,0.07)	0.0769
123	17.26	(16.7,17.81)	<0.0001	-0.27	(-1.28,0.42)	0.2436
124	27	(26.69,27.31)	<0.0001	-0.73	(-2.17,0.03)	0.0436
125	25.35	(24.79,25.9)	<0.0001	0.4	(-0.46,1.86)	0.7613
126	31.15	(30.99,31.31)	<0.0001	-0.04	(-0.36,0.19)	0.3969
127	25.19	(24.82,25.55)	<0.0001	-0.03	(-0.65,0.57)	0.4667
128	27.23	(27.13,27.34)	<0.0001	-0.65	(-1.51,-0.03)	0.0022
129	18.82	(18.51,19.12)	<0.0001	-1.36	(-4.03,0)	0.0236
130	29.71	(29.54,29.89)	<0.0001	0.2	(-0.07,0.71)	0.8551
131	30.25	(30.09,30.41)	<0.0001	0.93	(0,2.4)	0.9796
132	28.47	(28.18,28.75)	<0.0001	0.55	(-0.04,1.65)	0.9507
133	28.83	(28.59,29.07)	<0.0001	-0.14	(-0.57,0.16)	0.2036
134	26.64	(26.27,27.01)	<0.0001	-0.92	(-2.87,0.01)	0.0276
135	23.5	(23,23.99)	<0.0001	0.04	(-0.67,0.79)	0.5391
136	18.31	(17.75,18.79)	<0.0001	-0.66	(-1.98,0.03)	0.0422
137	8.73	(8.12,9.31)	<0.0001	-0.73	(-2.51,0.08)	0.0684
138	17.73	(17.12,18.34)	<0.0001	0.02	(-0.89,0.98)	0.5182
139	20.49	(20.01,20.98)	<0.0001	-0.48	(-1.86,0.21)	0.1564
140	28.25	(28.01,28.49)	<0.0001	0.08	(-0.29,0.53)	0.648
141	22.81	(22.33,23.27)	<0.0001	0.01	(-0.64,0.7)	0.5196
142	30.03	(29.86,30.19)	<0.0001	-0.25	(-0.77,0.03)	0.068
143	29.36	(29.11,29.63)	<0.0001	-0.79	(-2.2,-0.02)	0.0133
144	15.6	(15.31,15.9)	<0.0001	-0.42	(-1.29,0.05)	0.0636
145	20.86	(20.53,21.16)	<0.0001	-1.24	(-3.51,-0.05)	0.0058
146	25.62	(25.33,25.91)	<0.0001	-0.59	(-1.68,0.03)	0.0391
147	26.25	(26.03,26.48)	<0.0001	-0.01	(-0.37,0.35)	0.4911
148	18.51	(17.96,19.03)	<0.0001	-1.28	(-4.44,-0.01)	0.0218

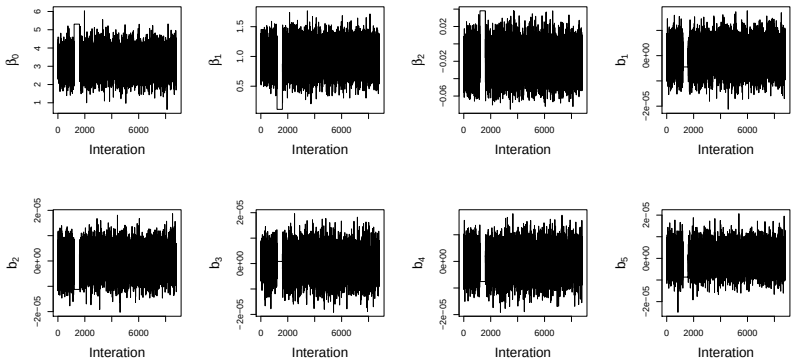
	α	95 % C.I.	P	β	95 % C.I.	P
149	24.16	(23.85,24.48)	<0.0001	-0.82	(-2.36,-0.02)	0.02
150	23.37	(22.86,23.88)	<0.0001	-0.29	(-1.33,0.42)	0.2458
151	30.2	(30.05,30.35)	<0.0001	-0.34	(-0.94,0.01)	0.0404
152	26.88	(26.66,27.1)	<0.0001	-0.77	(-2.02,-0.03)	0.0093
153	26.52	(26.27,26.78)	<0.0001	-0.7	(-2.06,0)	0.0249
154	26.56	(26.29,26.81)	<0.0001	-0.82	(-2.32,-0.02)	0.0133
155	29.1	(28.99,29.2)	<0.0001	-0.65	(-1.6,-0.03)	0.0009
156	28.19	(28.08,28.3)	<0.0001	-0.2	(-0.62,0.01)	0.0529
157	21.67	(21.39,21.97)	<0.0001	-0.25	(-0.88,0.11)	0.1218
158	25.39	(25.15,25.63)	<0.0001	-0.26	(-0.84,0.06)	0.0773
159	26.27	(26.08,26.45)	<0.0001	-0.47	(-1.26,0)	0.0289
160	23.08	(22.7,23.42)	<0.0001	-0.5	(-1.53,0.02)	0.0471
161	27.89	(27.63,28.16)	<0.0001	-0.02	(-0.43,0.35)	0.4564
162	30.02	(29.87,30.16)	<0.0001	-0.15	(-0.52,0.05)	0.1209
163	25.95	(25.67,26.22)	<0.0001	0.21	(-0.15,0.8)	0.816
164	29.02	(28.83,29.2)	<0.0001	-0.3	(-0.92,0.04)	0.0804
165	30.35	(30.16,30.53)	<0.0001	-0.19	(-0.66,0.09)	0.1582
166	25.08	(24.8,25.35)	<0.0001	0.62	(0.01,1.77)	0.9764
167	23.68	(23.32,24)	<0.0001	0.88	(0.02,2.53)	0.9876
168	28.41	(28.26,28.55)	<0.0001	-0.03	(-0.29,0.17)	0.38
169	23.58	(23.38,23.78)	<0.0001	0.02	(-0.3,0.37)	0.5338
170	25.22	(24.92,25.53)	<0.0001	0.34	(-0.12,1.13)	0.8849
171	28.83	(28.62,29.03)	<0.0001	-0.02	(-0.39,0.3)	0.4609
172	12.89	(12.33,13.43)	<0.0001	-0.36	(-1.13,0.09)	0.0858
173	16.27	(15.94,16.58)	<0.0001	-0.34	(-1.08,0.08)	0.084
174	24.82	(24.47,25.16)	<0.0001	-0.76	(-2.43,0.01)	0.0342
175	22.8	(22.4,23.21)	<0.0001	-0.84	(-2.79,0.07)	0.0542
176	25.78	(25.5,26.07)	<0.0001	-0.2	(-0.78,0.19)	0.1898
177	30.36	(30.22,30.49)	<0.0001	-0.15	(-0.52,0.04)	0.1249
178	28.6	(28.42,28.77)	<0.0001	-0.27	(-0.83,0.02)	0.0564
179	24.97	(24.68,25.27)	<0.0001	-0.68	(-2.02,0)	0.0284
180	28.8	(28.62,28.98)	<0.0001	-0.58	(-1.55,-0.01)	0.0147
181	19.18	(18.68,19.62)	<0.0001	-0.61	(-1.94,0.08)	0.0689
182	13.13	(12.5,13.72)	<0.0001	0.76	(-0.05,2.53)	0.948
183	17.29	(16.64,17.86)	<0.0001	0.68	(-0.11,2.1)	0.924
184	27.32	(26.95,27.69)	<0.0001	0.45	(-0.13,1.53)	0.8978
185	30.08	(29.94,30.21)	<0.0001	-0.45	(-1.14,-0.01)	0.0169
186	20.48	(20.17,20.8)	<0.0001	-0.28	(-0.98,0.15)	0.1418
187	23.48	(23.21,23.74)	<0.0001	-0.75	(-2.01,-0.01)	0.0147
188	8.95	(7.36,10.15)	<0.0001	-0.11	(-0.75,0.56)	0.3142
189	14.36	(14.01,14.71)	<0.0001	-0.71	(-2.05,0.02)	0.032
190	23.77	(23.41,24.12)	<0.0001	-0.43	(-1.46,0.11)	0.0969
191	24.8	(24.54,25.08)	<0.0001	-1.17	(-3.47,-0.04)	0.0071
192	30.1	(29.85,30.32)	<0.0001	-0.38	(-1.1,0.04)	0.0547
193	30.41	(30.23,30.6)	<0.0001	-0.4	(-1.12,0.01)	0.0382
194	23.73	(23.45,23.98)	<0.0001	0	(-0.37,0.36)	0.4804

Appendix D

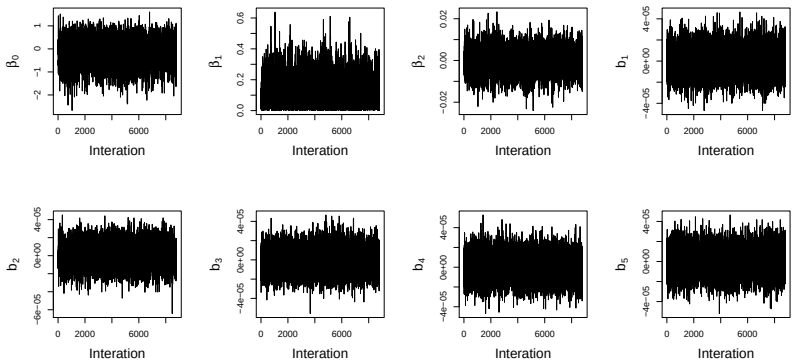
Convergence Plots for Mean Deviation Spline Model



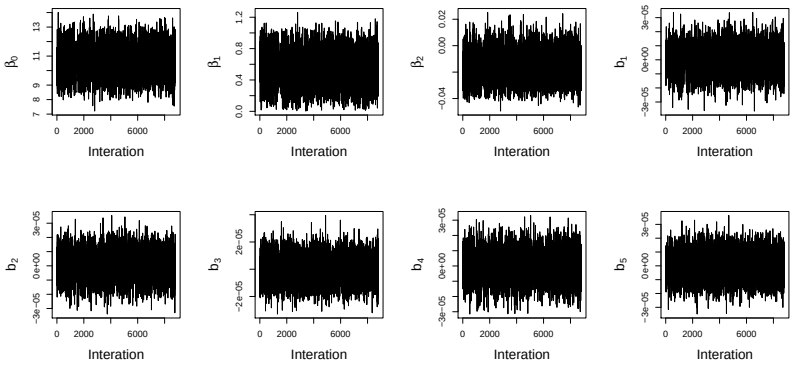
(a) Subject 1, Right Eye



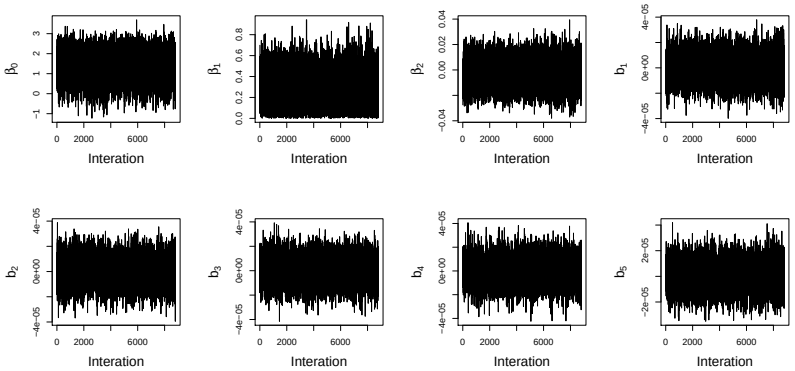
(b) Subject 1, Left Eye



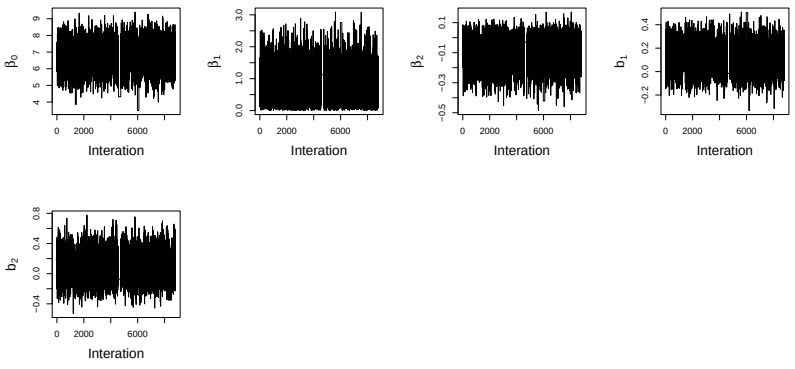
(c) Subject 2, Right Eye



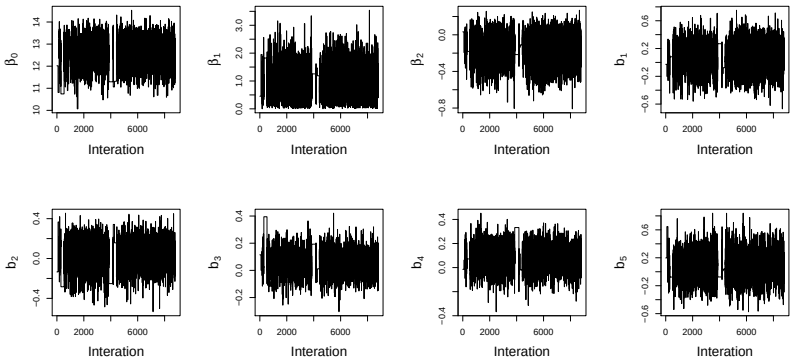
(d) Subject 2, Left Eye



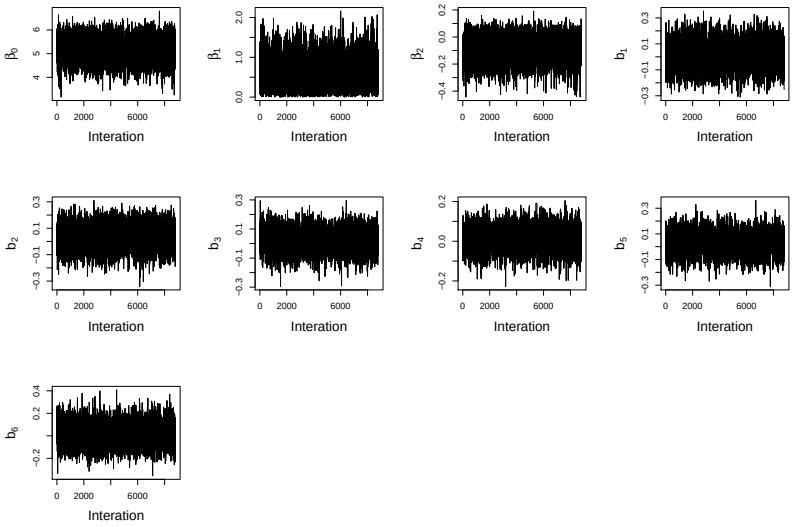
(e) Subject 3, Right Eye



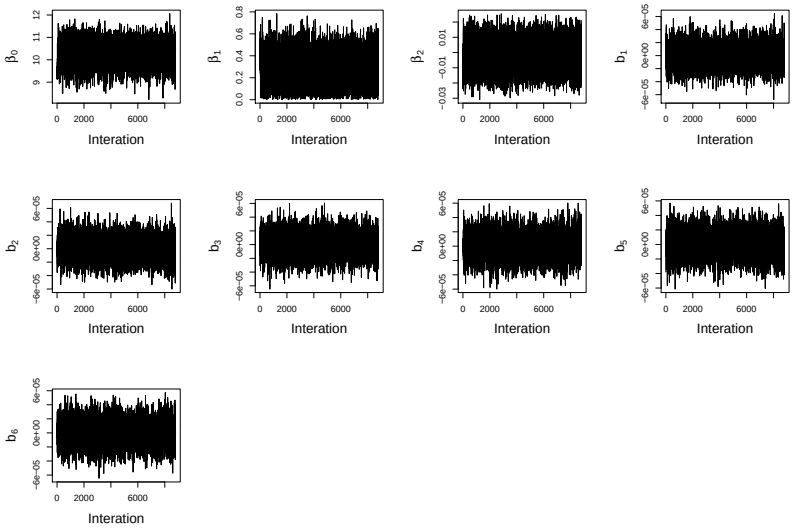
(f) Subject 3, Left Eye



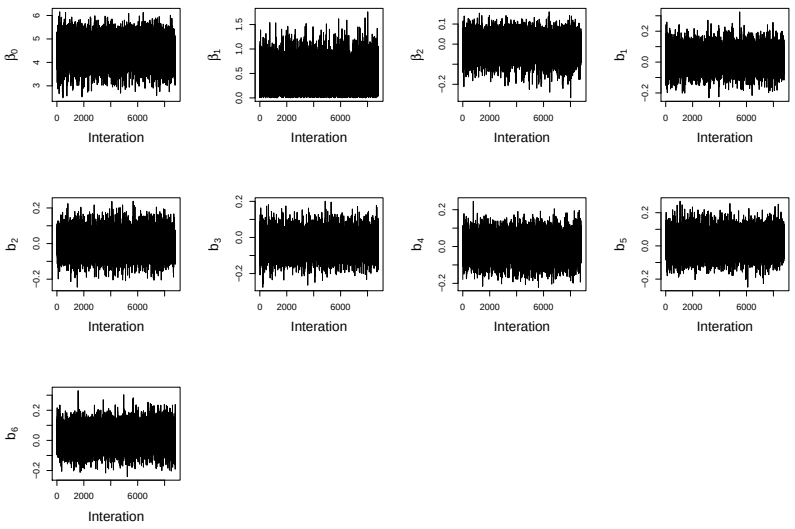
(g) Subject 4, Right Eye



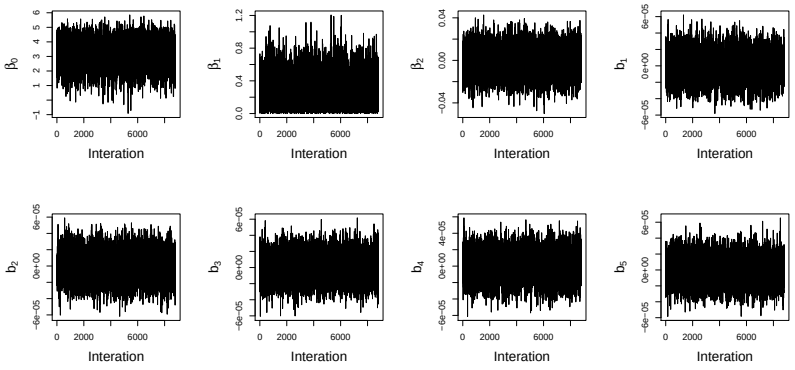
(h) Subject 4, Left Eye



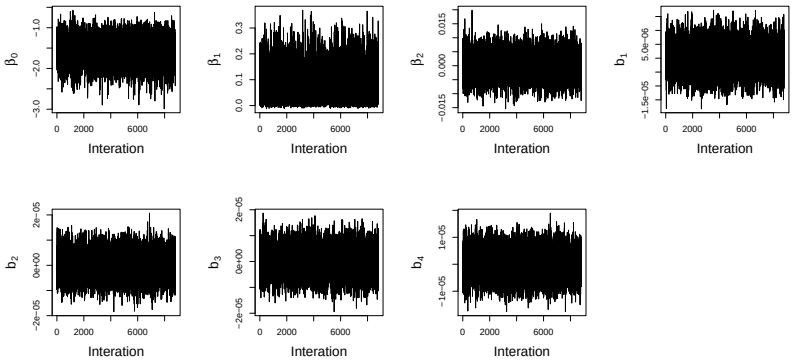
(i) Subject 5, Right Eye



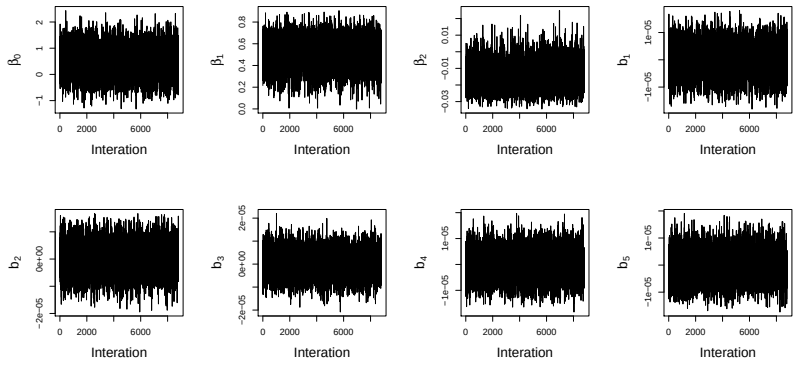
(j) Subject 5, Left Eye



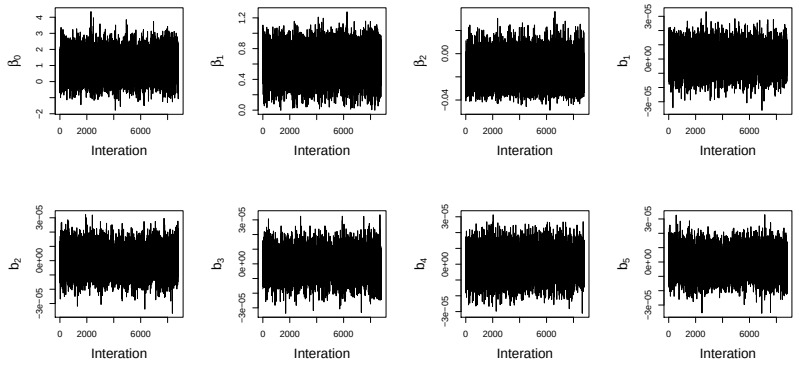
(k) Subject 6, Right Eye



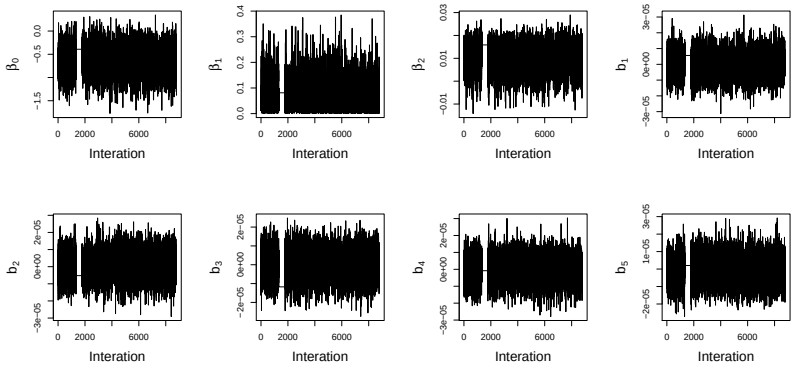
(l) Subject 6, Left Eye



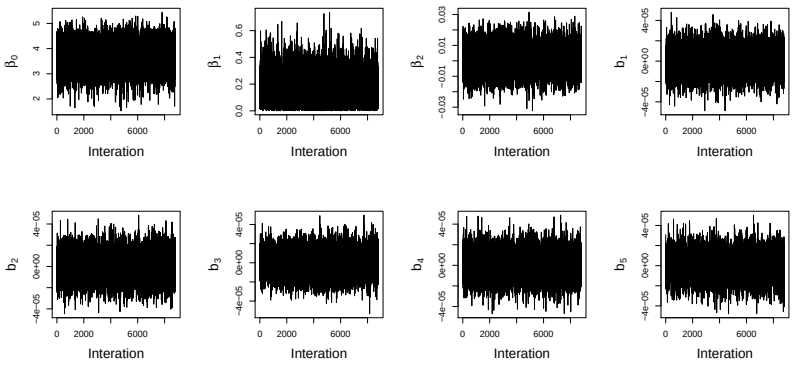
(m) Subject 7, Right Eye



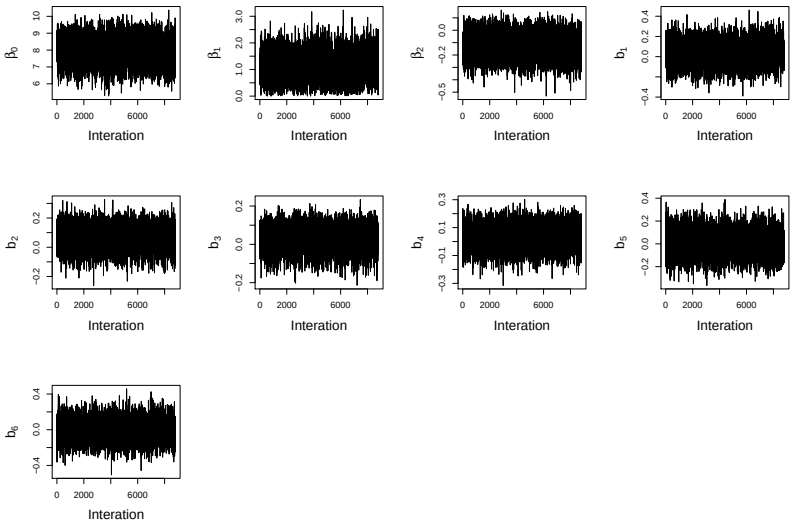
(n) Subject 7, Left Eye



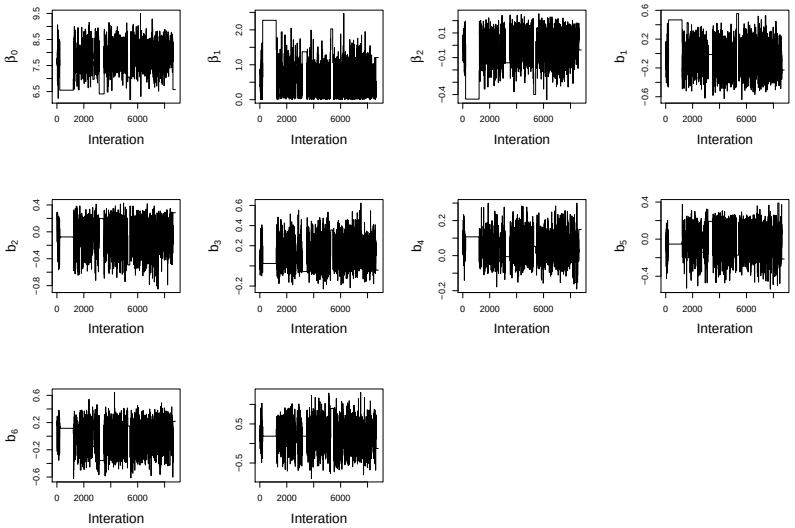
(o) Subject 8, Right Eye



(p) Subject 8, Left Eye



(q) Subject 9, Right Eye



(r) Subject 9, Left Eye

Appendix E

Statement of Contribution to Doctoral Thesis Containing Publications

MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION TO DOCTORAL THESIS CONTAINING PUBLICATIONS

(To appear at the end of each thesis chapter/section/appendix submitted as an article/paper or collected as an appendix at the end of the thesis)

We, the candidate and the candidate's Principal Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of Candidate: Brigid Betz-Stablein

Name/Title of Principal Supervisor: Prof. Martin Hazelton

Name of Published Research Output and full reference:

Betz-Stablein, B.D., Morgan, W.H., House, P.H., Hazelton, M.L. (2013). Spatial modeling of visual field data for assessing glaucoma progression. Investigative Ophthalmology & Visual Science, 54(2), 1544-1553.

In which Chapter is the Published Work: Chapter 4

Please indicate either:

- The percentage of the Published Work that was contributed by the candidate: 80%

  and / or
- Describe the contribution that the candidate has made to the Published Work:

Brigid Betz-Stablein — Digitally signed by Brigid Betz-Stablein DN: cn=Brigid Betz-Stablein, o=Massey University, ou=IFS, email=b.d.betz-stablein@massey.ac.nz, c=NZ Date: 2013.11.11 09:50:56 +13'00'

Candidate's Signature — 11/11/2013 Date

Martin Hazelton — Digitally signed by Martin Hazelton DN: cn=Martin Hazelton, o=Massey University, ou=Institute of Fundamental Sciences, email=m.hazelton@massey.ac.nz, c=NZ Date: 2013.11.11 10:02:03 +13'00'

Principal Supervisor's signature — 11/11/2013 Date