REFERENCE ONLY

STATISTICAL METHODS FOR THE ANALYSIS OF VISUAL FIELD DATA IN GLAUCOMA

STUART KEITH GARDINER

A thesis submitted in partial fulfilment of the requirements of The Nottingham Trent University for the degree of Doctor of Philosophy

August 2003

ProQuest Number: 10183217

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10183217

Published by ProQuest LLC (2017). Copyright of the Dissertation is held by the Author.

All rights reserved.
This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

10369668

THE NOTTINGHAM TRENT
UNIVERSITY LIS

REF | PH.D/LS/03 GAR

S.L.

Declaration

This thesis has been completed solely by the candidate, Stuart Gardiner. The work contained in it was done by the candidate.

It has not been submitted for any other degrees, either now or in the past. Where work contained within it has been previously published, this has been stated in the text.

All sources of information have been acknowledged and references given.

Abstract

Statistical Methods for the Analysis of Visual Field Data in Glaucoma

Stuart K Gardiner (Nottingham Trent University)

Glaucoma is a leading cause of visual disability. Automated static perimetry determines the minimum contrast that a subject can perceive (called the sensitivity) at individual locations within their visual field. This instrumentation currently offers the most reliable strategy for the clinical follow-up of the disease. The accurate detection of glaucomatous change in a series of visual field results is important in the clinical management of a patient, and in the evaluation of which treatments are most effective in arresting progression. The slow, often equivocal rate of sensitivity loss, and the variability that exists between field results, makes this a difficult task.

The purpose of this thesis is to improve the statistical methods used for determining true visual field defects and progression in suspected glaucoma patients.

Pointwise linear regression models for identifying true change in the sensitivity at a point are examined. By simulating visual field data, it is shown that a testing frequency of three tests per year is optimal for a given rate of sensitivity loss. Next, the use of confirmation fields to verify suspected progression is examined, and the comparative specificity and sensitivity of different confirmation techniques assessed; a novel technique where the penultimate field is omitted is found to be the most specific.

A novel spatial filter is developed, based on pair-wise covariances between different points in the fields of patients at a tertiary glaucoma clinic. The filter is found to closely reflect the anatomical structure of the retinal nerve fibre layer. This filter appears to reduce the noise present in visual field tests, as evidenced by an improved fit to pointwise linear regression, fewer false positives and fewer false negatives when looking for progression. It has been programmed into computer software used to identify progressing defects.

The effect of applying the filter to patient data is proposed as a measure of the amount of noise present. This noise is modeled statistically to determine the suitability of assuming that the noise is normally distributed. The assumption is found to be not unreasonable but inaccurate; an improved model is developed which fits the empirical data better and over a wider range of sensitivities.

These new quantitative methods, in particular the new spatial filter, may prove clinically useful in the analysis of visual function deterioration in glaucoma.

Contents

TITLE PAGE i

Declaration ii

ABSTRACT 1

Contents 2
List of Figures and Tables 4

Acknowledgements 8

OUTLINE 9

Chapter 1 – BACKGROUND 11

1.1 Glaucoma 11
1.2 Visual Fields and Perimetry 13
1.2.1 Variability in Perimetry 19
1.3 Progression 20
1.4 Aims of Study 25

Chapter 2 – VIRTUAL EYE SIMULATION 26

2.1 Previous Work 26
2.2 The 'Virtual Eye' Simulation Model 29

Chapter 3 – FREQUENCY OF TESTING 34

3.1 Method 34
3.2 Results 35
3.3 Conclusions 37

Chapter 4 – CONFIRMATION FIELDS METHODS 39

4.1 The PLR Methods Examined 39
4.2 Theoretical Comparison 41
4.2.1 Assumptions 41
4.2.2 Derivation 41
4.3 Virtual Eye Simulation Experiment 46
4.3.1 Method 46
4.3.2 Results 48
4.4 Conclusions 50

Chapter 5 – DERIVATION OF FILTER 52

5.1 Previous Work 52
5.2 Mathematical Derivation 53
5.3 The Data Used 56

5.3.1 Use of Unreliable or Non-glaucomatous Fields 57
5.3.2 Use of All Fields from Each Patient 58
5.4 The Filter 59

Chapter 6 – TESTING THE FILTER 63

6.1 Longitudinal Data 63
6.2 Virtual Eye Simulation 64
6.2.1 The Simulated Visual Fields 64
6.2.2 Results from the Virtual Eye Experiments 67
6.3 Conclusions 70

Chapter 7 – MODELLING NOISE 72

7.1 Using The Filter To Estimate Noise 72
7.2 Estimates Of The Noise 73
7.3 Pearson Densities 75
7.4 Modelling The Noise 76
7.4.1 The sample mean of noise estimates 80
7.5 Examples 82
7.6 Testing The Model 85

Chapter 8 – CONCLUSIONS 87

Appendix 1 – VIRTUAL EYE PROGRAM 89

Appendix 2 – PROGRESSOR CODE 93

REFERENCES 97

List of Figures and Tables

Page:

Figure 1.1: A photograph of the retina of a healthy eye. 12

Figure 1.2: The subset of points in the 30-2 field which comprise the 24-2 visual field. The dashed grey lines show the horizontal and vertical meridians. The nose is on the right; and the large grey area towards the left shows the approximate position of the physiological blind spot (note that two of the test locations coincide with this area). 15

Figure 1.3: A sample printout from a Humphrey Field Analyser incorporating Statpac 2. 17

Figure 1.4: Two sample visual fields, from a HFA. The top field is from a healthy eye; the bottom field shows an eye with an advanced defect. The grid on the left in each case shows the raw sensitivities at each point, in decibels; where a second number is shown underneath the number, it is the second threshold estimate at that point (where applicable). The key at the bottom shows how the greyscales are generated from the sensitivities in the decibel scale (bottom row) or the equivalent luminances (middle row). Note that both of these eyes were tested using the 24-2 visual field algorithm. 18

Figure 1.5: Pointwise 90% test-retest intervals for full-threshold perimetry, reproduced from Artes et al (2003). For each threshold level of the initial test, the circles represent the 5^{th} and 95^{th} percentiles of the distribution of retest values based on patient data. The solid lines indicate the percentiles established from simulations of the testing algorithm. 20

Table 1.1: The differing clinical opinions of two experienced observers on the status of 27 series of 16 visual fields. 21

Figure 1.6: Sample output from the Progressor program. Visual fields can be displayed individually in greyscale form (as with the right eye here); alternatively, colour-coded bars can be displayed for each point (as with the left eye here), with the size of each bar representing the deviation from normal sensitivity for that point at that test, and the colour representing the statistical significance of the regression slope up to and including that test. A series of bars is shown at each point, showing the deviation and significance at that point for each test in the series up to and including the currently selected field. This enables progressing points to be easily identified. Such points can also be automatically labelled on the greyscale diagram; the exact criteria for slope and significance level are adjustable by the user. 24

Figure 2.1: Example series of greyscales. The four fields were taken on (from left to right) 24/4/86, 16/7/87, 9/6/88 and 11/7/89. 26

Figure 2.2: Illustrative series for a stable Virtual Eye, with an age-related decline of 0.1dB/year. 31

Figure 2.3: Illustrative series for a Virtual Eye progressing at 2dB/year. 32

Figure 2.4: Three series of visual fields, demonstrating the effect of the Virtual Eye simulation. The left-hand column shows a noise-free field in which six points (those highlighted in red in the final field of the series) are deteriorating at 2dB/year over six years. The middle column shows the same series when noise has been added by the Virtual Eye; note that now only four of the six progressing points are identified by PLR after six years. The right-hand column shows a stable series after noise has been added by the Virtual Eye; note that five points are still (incorrectly) flagged as progressing by PLR. Comparing the middle and right-hand columns shows the difficulty of separating progressing from stable eyes, especially when based solely on examining the greyscales by eye. 33

Figure 3.1: The effect of changing the frequency of testing on the proportion of points flagged as progressing, when the noise-free behaviour of the eye is Stable (top) and Progressing (bottom). 36

Table 3.1: The effect of changes in the specificity of a single point on the specificity of the entire visual field. 37

Figure 4.1: The effect of omitting a point from the regression on the PLR slope. 40

Figure 4.2: Theoretical relative specificity and sensitivity of the different PLR methods. 45

Figure 4.3: The effect of adding noise to the actual input sensitivity values for each point in a Virtual Eye. Visual fields are represented as 30-2 Humphrey greyscales with darker areas indicating lower sensitivity readings. The three noisy eyes generated are noticeable different both from each other, and also from the true physiological values represented by the top greyscale. 46

Figure 4.4: The progression in our noise-free eye, as used in the Virtual Eye simulation. 47

Figure 4.5: The relative specificity and sensitivity (respectively) of four different PLR methods, as determined by the Virtual Eye simulation. 49

Figure 4.6: The relative specificity and sensitivity (respectively) of the remaining three different PLR methods, plus the Standard Criteria method again for reference, as determined by the Virtual Eye simulation. 49

Figure 5.1: How the filtered sensitivity is calculated by the Gaussian filter. In this method, the raw sensitivity value is replaced by one derived from a linear combination of the sensitivities at the nine points in a square centred on the point of interest. This is repeated for each point in the field in turn, each time looking at the points in a square surrounding the point of interest. 53

Figure 5.2: An illustration of the shape of the filter. The grey area represents the blind spot. 60

Figure 5.3: The effect of the new filter on a visual field exhibiting a scotoma in the superior hemifield, compared with the effect of applying the Gaussian filter to the same data. 61

Figure 5.4: An example of the use of Progressor to apply the filter to a visual field from one of the program's demonstration patients. The same field is shown twice, with the raw field on the left and the filtered field on the right. 62

Figure 6.1: Frequency Distribution of R-Squared values for series of Raw sensitivities (dashed line); Gaussian-filtered sensitivities (dotted line) and Filtered sensitivities (solid line). 64

Figure 6.2: The optic nerve head location, in degrees, for each non-blind spot visual field test point in a right eye. The temporal meridian (9-o'clock position) was designated as 0°, and degrees counted clockwise from there. Reproduced from Garway-Heath et al (2000). 65

Figure 6.3: The shapes of some of the 52 generated defects tested, superimposed onto a Humphrey 24-2 visual field. The black square in each case is the Central Point. Dark grey squares represent the *Defective* points (within five degrees of the Central Point in the ONH); light grey squares represent the *Border* points (between six and ten degrees away from the Central Point in the ONH). The large grey circle towards the left of each field represents the location of the physiological blind spot. 66

Figure 6.4: The shape of the non-glaucomatous defect used, superimposed onto a Humphrey 24-2 visual field. As before, the grey circle represents the physiological blind spot. 66

Figure 6.5: Percentage of Central Points correctly flagged as progressing for each of the tested defects. Each column contains three symbols, representing the detection rates for the raw data, the data after application of the Gaussian filter, and the rate after application of our new filter; there is one column per point tested. The points have been split into the two hemifields. The points are in each case shown in order of increasing detection rate after application of the Gaussian filter (rather than in order of location), to make the figure clearer to understand. No relationship was found between the area of the field which a point was in and the effect on the detection rate of filtering. 68

Figure 6.6: The five generated localised defects whose detection rate appears to be reduced by filtering in Figure 6.5, using the same symbols as in Figure 6.3. The lower line shows the points involved in the filtering algorithm for that particular Central Point; the thicker the line, the more effect the point has on the filtered value for the Central Point, as in Figure 5.2. 69

Figure 7.1: The empirical distribution of the noise estimates for different filtered sensitivity values. The red line shows the pdf for the noise estimates when the filtered sensitivity is between 32.25 and 32.75dB; the green line is for filtered sensitivities between 26.75 and 27.25dB; and the blue line is for filtered sensitivities between 16.75 and 17.25dB. 74

Figure 7.2: Graphs of the four distribution measures. 77

Figure 7.3: Empirical pdf of the raw sensitivity when the filtered sensitivity is between 9.25 and 9.75dB, showing the peak corresponding with the censoring effect at zero. 77

Figure 7.4: Comparing the empirical estimates of the mean noise (shown by crosses) with estimates calculated from the model (shown by dots) at different sensitivities. 79

Table 7.1: Estimated distribution characteristics for the three example sensitivities of 30dB, 20dB and 10dB. 82

Table 7.2: Estimated distribution moments for the three example sensitivities. 82

Table 7.3: Estimated Pearson coefficients for the three example sensitivities. 82

Table 7.4: Estimated distributional constants obtained from numerical integration for the three example sensitivities. 83

Figure 7.5: Comparing the pdf of the modelled distribution of noisy sensitivities (blue line) to the empirical pdf (red line), with the normal distribution (yellow line) for reference. The top graph shows this for an input sensitivity of 30dB; the middle graph for 20dB; and the bottom graph for 10dB. 84

Figure 7.6: Comparing the test statistics (as defined in Equation 7.6.1) for the normally-distributed noise model (crosses) and the model derived in this chapter (circles). 86

Acknowledgments

The author would like to acknowledge the assistance of his supervisory team in the work leading up to and preparation of this thesis; Dr David Crabb PhD (Nottingham Trent University) and Prof Fred Fitzke PhD (Institute of Ophthalmology, London).

The author would also like to thank Prof Roger Hitchings FRCS FRCOphth (Moorfields Eye Hospital, London) for providing the data used in this thesis; and Mr Ananth Viswanathan BSc MBBS FRCOphth MD (Moorfields Eye Hospital) and Mr Ted Garway-Heath MD FRCOphth (Moorfields Eye Hospital) for their useful advice and encouragement during the work leading up to this thesis.

Further, the author would like to thank Dr Mike Baxter PhD and Dr John Marriott PhD for going through most of the statistical content of this thesis.

Finally, the author would like to thank all those who have provided friendship and support during the last three years.

Outline

Chapter 1 discusses the background behind this thesis. It introduces glaucoma as a severe and irreversible cause of damage to visual function, particularly prevalent in the older population. It then goes on to discuss the testing methods for glaucoma, in particular the use of perimetry to produce a map of the sensitivity to light of different parts of the retina. The Full Threshold algorithm for determining the sensitivity of each point in the visual field is described. The amount of variability in sensitivity estimates from perimetry is discussed, as are the reasons for this variability. Next, different methods for determining whether or not a field is progressing are described, including pointwise linear regression.

Chapter 2 describes a Virtual Eye simulation developed to address questions about the progression of visual fields. The chapter starts by summarising the reasons for using a simulation approach to visual fields, and discusses previously published simulations. It then goes on to describe the Virtual Eye simulation in detail, giving examples of the results it produces.

Chapter 3 describes an experiment which uses the Virtual Eye to determine the ideal number of field tests per year which should be carried out to optimise the specificity and sensitivity of pointwise linear regression. It concludes that performing three tests per year provides a significant improvement in sensitivity when compared with using fewer tests per year; yet increasing the frequency of testing beyond this level leads to unacceptable levels of false positives.

Chapter 4 discusses the use of confirmation fields when progression is suspected. Seven different methods for using confirmation fields are examined. Theoretical comparisons and an experiment using the Virtual Eye both showed that for maximum specificity, methods introduced in this chapter which omit the penultimate point when carrying out the confirmation test should be used.

Chapter 5 describes the derivation of a new spatial filter which aims to reduce the noise in visual fields, by comparing the sensitivities at each point with the sensitivities at points which are anatomically related to it in the retinal nerve fibre layer. The filter is based on the pair-wise covariances between points in the field, as derived from a large retrospective patient database, which is described in detail. The shape of the filter is found to resemble the pattern of nerves on the retina.

Chapter 6 describes two methods for determining the efficacy of the filter. First, series of visual fields from 303 eyes are examined, and it is found that filtering improves the fit of pointwise linear regression, indicating a reduction in the amount of noise. Next, the Virtual Eye is used to simulate stable and defective eyes. It is found that filtering reduces the proportion of false positives amongst stable eyes, whilst increasing the detection rate of points which are truly progressing in realistic glaucomatous defects, as derived from an anatomical structural map of the retina. Further, it is found that non-glaucomatous localised defects become less likely to be flagged as progressing after filtering. This indicates that

filtering significantly improves the process of distinguishing between healthy or stable eyes and progressing glaucomatous eyes.

Chapter 7 describes and justifies a novel way of estimating the noise present in visual fields by comparing the filtered sensitivity with the raw sensitivity at each point. Empirical estimates of the noise at each sensitivity level are then modelled by a Pearson Type IV distribution, and examples of the fit of the model are given. The model is found to fit the empirical data significantly better than a normal distribution.

Chapter 8 sums up the work in the thesis, noting the novel contributions to the field of work; and gives suggestions for future work.

This is an applied thesis, written primarily with the statistically literate researcher in visual fields in mind. It is also to be hoped that a clinical reader should gain useful information from it; and further that the thesis provides a useful example of the application of a selection of applied statistical methods.

1. Background

This chapter gives an introduction to glaucoma and visual fields, and the background behind the current work. It also sets out the aims of this study. Background to the more specialised work underlying different elements of the study is included in later chapters when they become relevant.

1.1 Glaucoma

Glaucoma is estimated to be the second commonest cause of blindness in the world. Although there is no universally accepted definition, it can be considered as an eye disease characterised by damage to the optic nerve head. This can result in partial or full loss of vision from the eye in question in certain directions, determined by the part of the optic nerve head which has been damaged. If untreated, the disease will normally progress, resulting in the vision loss in affected areas worsening and/or spreading to other areas; eventually this may result in complete loss of the vision in that eye. Once vision loss has occurred, it is irreversible; although the decline can be slowed or even halted by appropriate treatment in most cases. Since adequate therapy significantly slows progression of the disease (Hitchings et al 1995; Bhandari et al 1997; Koseki 1997; AGIS 2000; Wormald 2003), the early identification of individuals at risk is considered important.

The commonest form of glaucoma is primary open-angle glaucoma (POAG); this is when the disease is unrelated to other underlying ocular conditions, and the aqueous drainage route is not closed by some known other normal or pathological structure. Frequently, glaucoma is associated with a higher than normal pressure inside the eyeball (the intraocular pressure or IOP) (Coleman 1999; Henson 2000, p.92), in which case the lowering of IOP, by either increasing the rate of outflow from the eye or reducing the rate of production, is an important clinical tool; however there is a subset of patients whose IOP is considered normal, known as Normal Tension Glaucoma (NTG) patients. Patients whose IOP is abnormal but who do not, as yet, show other symptoms of glaucoma are described as having Ocular Hypertension (OHT), and are considered as glaucoma suspects. Results from recent clinical trials have confirmed the long-held belief that reducing IOP reduces the risk of glaucoma developing and progressing (Kass et al 2002; Heijl et al 2002; Wormald 2003).

Glaucoma initially results in partial or full loss of vision in small areas of the visual field. These areas (called localised defects) follow the shape of the nerves along the retina (the retinal nerve fibre layer, or RNFL) which conduct messages from the light sensitive cells (called rods and cones) spread across the retina into the optic nerve, and hence to the brain. In Figure 1.1, the optic nerve enters the retina towards the left of the picture; at this point (and only at this point) there are no light-sensitive cells, causing the phenomenon known as the physiological blind spot. This occurs just above the horizontal meridian, approximately 10 to 15 degrees from the centre towards the nose; resulting in a blind spot in vision a similar distance temporal of the centre just below the meridian. The nerves in the RNFL then spread out

from the optic nerve head in an arcuate pattern, as seen in the photograph; the nerve fibres (which appear as light grey lines on the darker background) leave the optic nerve head in all directions, but tend to bend to the right as they get further away, eventually heading back towards the horizontal meridian towards the right hand side of the photograph. Therefore a glaucomatous localised defect will normally also follow an arcuate pattern mimicking the shape of the RNFL.

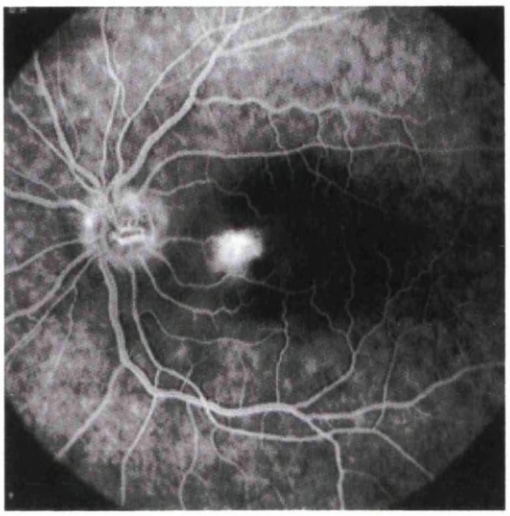


Figure 1.1: A photograph of the retina of a healthy eye.

The literature on the underlying mechanisms of glaucoma is both extensive and subject to ongoing discussion; some papers have failed to find significant correlation between IOP and loss of visual function (Holmin & Krakau 1982; O'Brien et al 1991; Chauhan & Drance 1992). Since this discussion is tangential to the work described in this document, it shall not be considered in detail; a comprehensive review is given by Hayreh (1994).

Prevalence is higher amongst people of Afro-Caribbean origin and other black populations (Tielsch et al 1991; Leske et al 1994; Buhrmann et al 2000). In a white population, the prevalence of glaucomatous visual field defects has been reported to be 1.5% (Wolfs et al 2000). Family history of the disease is amongst the other risk factors associated with an increased susceptibility to POAG (Leske 1983; Tielsch et al 1994; Wolfs et al 1998). Work on identifying gene loci associated with an increased risk of glaucoma is continuing (Stoilova et al 1996; Stone et al 1997; Wirtz et al 1997; Trifan et al 1998; Sarfarazi et al 1998; Wirtz et al 1999; Aung et al 2002a; Aung et al 2002b; Pang et al 2002; Aung et al 2003).

Coffey et al (1993) suggested that at the time of the study there were approximately 250,000 known sufferers in the UK alone, with probably an equal number of undiagnosed cases; since glaucoma is an age-related condition, this number would be expected to rise in line with the country's aging population. It has been estimated that approximately 5.2 million people worldwide are bilaterally blind from glaucoma, representing 15% of the total number of blind people (Thylefors & Negrel 1994). Although exact definitions of POAG differ from one study to another (most importantly the cut-off point at which the eye is said to be either normal or diseased), the prevalence of POAG has been estimated to be around 2% of the middle-aged and old-aged population (Coffey et al 1993; Tielsch et al 1991; Klien et al 1992).

This has been reported to range from 0.9% for subjects from 43 to 54 years of age, up to 4.7% for patients aged 75 or over (Klein et al 1992). Whilst glaucoma has a lower prevalence than cataract, the fact that it is irreversible (unlike cataracts which can be removed with surgery) contributes to its economic burden being of the same magnitude (Guzman et al 1992).

A thorough review of the definition of glaucoma and methods its treatment, and their side-effects, is given by Coleman (1999).

1.2 Visual Fields and Perimetry

The visual field is defined as being the area from which light can enter the eye and reach the retina, thus stimulating the light-sensitive cells on the retina (called rods and cones) into sending messages to the brain (Tate & Lynn 1977; Werner 1991; Henson 2000, p.1). The ganglion cells in the inner retina process the signals from these rods and cones prior to relaying them to the brain via their axons, which pass across the inner layer of the retina and exit the eye via the optic nerve head. For a bright stimulus, the normal extent of the visual field is 60° up, 75° down, 100° temporal and 60° nasal from the line of sight (Henson 2000, p.2). The visual field is said to be damaged if the patient has reduced or non-existent vision in a certain direction with the eye in question; that is, if the patient looks directly at a given point with that eye (called the point of fixation), they will not be able to see an object presented in another direction. The centre of the retina in a healthy eye has a higher concentration of ganglions than areas further away from fixation, and so the resolution is typically better there.

Traditionally, the visual field has been measured by perimetry (Lachenmayr & Vivell 1993). This is a technique which measures the light-difference sensitivity across the visual field. To do this, a visual stimulus is presented at one or more points, and the patient indicates whether or not the stimulus has been seen. Naturally, this must be carried out for one eye at a time (with the other eye being covered or closed), with the patient looking directly forward at a pre-determined point of fixation.

Recently, new technology has enabled imaging techniques to be developed, whereby the retina is 'photographed' so that damage can be looked for and assessed. While this is a promising (and very broad) area of research, perimetry is still far more widespread. Further, the relationship between structural damage (as shown by imaging techniques) and functional damage (as evidenced by perimetry) is far from straightforward (Shareef et al 2002); the two types of testing do not necessarily produce diagnoses concurrently. It is possible that in the future, imaging techniques will overtake perimetry and become the accepted diagnostic tool for glaucoma; however, that time has not yet come, and so perimetric research is still as relevant now as at any time in the past. It is still the only way of determining what a patient can actually see, and so addressing such important issues as quality-of-life and fitness-to-drive.

Several different types of perimetry have been used. Originally, manual perimeters were used. Manual static perimetry was considered to be a highly sensitive and exact, but very time-consuming, technique, and so was mainly used in research rather than routine clinical settings (Haley 1997, p.1-4). It required

the full attention of either a doctor or well-trained and experienced technician. Since the advent of computers, automated perimetry has become available; by following set rules and algorithms, standardised repeatable testing can be carried out. These tests do not have to be administered by a specialised technician. Further, visual stimuli can now be presented in random order, with their duration and intensity strictly controlled. These advantages have led to automated perimetry becoming the accepted tool in a clinical setting.

Most automated perimeters are based on white-on-white perimetry; that is, the patient looks at a white screen (actually normally a bowl) with one eye at a time, fixating on a point in the centre, and then white visual stimuli of varying brightness are presented at set points in the screen, designed to ascertain the patient's visual sensitivity at the corresponding point in their visual field. Alternative developments include blue-on-yellow perimetry (where blue stimuli are presented on a yellow screen) and kinetic perimetry (where the stimuli are in motion), but they are less widespread. Short-Wavelength Automated Perimetry (SWAP), based on blue-on-yellow perimetry, has been of much interest recently, even though it has been shown to have a higher variability than white-on-white perimetry (Sample et al 1993; Wild et al 1995; Kwon et al 1998; Wild et al 1998). Frequency-Doubling (FDT) Perimetry is designed primarily for the screening of possible glaucoma patients (Johnson & Samuels 1997). This study will be concerned solely with white-on-white perimetry; comprehensive reviews of the principles of perimetry, and the different instruments used, are given by Lachenmayr & Vivell (1993) and Henson (2000, p.23-49).

The most widely used automated perimeters in both clinical (especially in secondary or tertiary care) and research settings are the Octopus perimeter (Interzeag AG, Schlieren-Zurich, Switzerland) and the Humphrey Field Analyzer (or HFA) (Carl Zeiss Meditec AG, Germany). This study is based on results from the HFA, which is the more commonly used instrument in the UK. The methods developed could easily be transferred to data from other types of perimeter if desired. For full information on the HFA, see Haley (1987) or Werner (1991, p.67-89). Each stimulus can be presented at varying brightness, from 0.08 to 10,000 apostilbs luminance; the lowest brightness which can be seen at the point in question is known as the sensitivity threshold of that point, reported on a decibel (dB) scale which is based on the log luminance. If the stimulus intensity is measured in Lamberts (where one Lambert = 10,000 apostilbs), then $1dB = 10 * \log(1/L)$. A sensitivity of 30-40dB is considered as normal (depending on, for example, the age of the patient); a sensitivity of 0dB means that the patient could not detect the brightest stimulus presented by the machine at that point. It has recently been suggested that switching to a Lambert scale rather than a decibel scale could improve diagnostic techniques (Garway-Heath et al 2000a), but this is not yet proven (Racette et al 2003).

Different testing strategies have been proposed. Full Threshold testing (described below) was the standard for many years, and is still the benchmark to which other strategies are compared; the data used in Chapters 5-7 was obtained using the Full Threshold strategy. More recently, the Swedish interactive threshold algorithm (SITA) (Bengtsson et al 1997) has been gaining acceptance; it is designed to provide sensitivity values of an equal reliability in a shorter testing time. It does this by estimating in a Bayesian manner the expected threshold value at each point, based on prior knowledge of the point's neighbours and the distribution of threshold values from other patients. Whether the variability of SITA results is in

fact the same as those from the Full Threshold strategy is currently subject to much debate (Bengtsson & Heijl 1999; Shirato et al 1999; Wild et al 1999a; Sharma et al 2000; Artes et al 2002); whilst it is to be hoped that the results presented in this study would be equally applicable to SITA data (as they would be if the variability was truly the same, both in magnitude and in other distributional characteristics), this has not yet been tested. SITA Fast is a similar algorithm with an even further reduced testing time, although with an associated higher test-retest variability (Bengtsson & Heijl 1998).

In the Full Threshold algorithm, the HFA measures the threshold sensitivity at each point by a so called "repetitive staircase" procedure. An initial stimulus is presented slightly brighter than the patient's expected sensitivity at the point being measured, based on neighbouring points. Every time a stimulus is seen, the patient presses a button. Each time a stimulus is seen, the next stimulus presented by the machine at that point will be 4dB dimmer (i.e. 4dB higher on the decibel scale for sensitivity); this is repeated until the stimulus is too dim to be seen. At this point, the intensity is increased again in steps of 2dB until it is seen again. This then gives the measurement of the threshold at this point. If the initial stimulus is not seen, the process runs in reverse; stimuli are presented in steps 4dB brighter (4dB lower) until one is seen, at which point the intensity is reduced by 2dB at a time until it is no longer visible. For example, if the sensitivity of a point was 16dB (that is, stimuli of 16dB or below will be seen) then a typical sequence of stimuli could be 30dB, 26dB, 22dB, 18dB, 14dB, 16dB, 18dB, where only the 14dB and 16dB stimuli were seen. The strategy begins by twice thresholding four initial points, one in each quarter of the visual field; it subsequently moves onto the other locations in the field. Some points are thresholded twice, partly to gain an estimate of the reliability and variability of the test, and also to confirm the threshold at any locations whose estimate is more than 4dB away from the expected value (based on neighbouring values). The stimuli for different points are presented in random order, rather than determining each threshold individually one at a time.

The locations tested are normally determined by either the 30-2 or 24-2 patterns. The 30-2 field gives the threshold estimates of locations in a grid covering the central 30° of the visual field; the locations are 6° apart, with none falling on the horizontal or vertical meridians. This gives a total of 76 test locations, as shown in Figure 1.2.

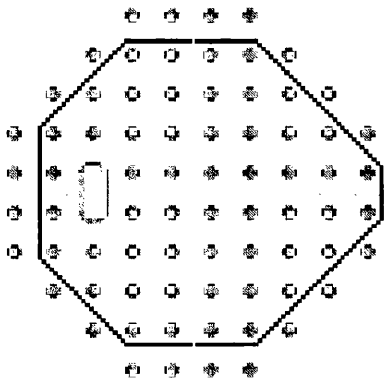


Figure 1.2: The subset of points in the 30-2 field which comprise the 24-2 visual field. The dashed grey lines show the horizontal and vertical meridians. The nose is on the right; and the large grey area towards the left shows the approximate position of the physiological blind spot (note that two of the test locations coincide with this area).

The 24-2 field consists of the subset of the 30-2 locations which fall within the central 24° of the visual field, plus the two most peripheral locations (nearest the nose, and hence furthest from the physiological blind spot caused by the optic nerve entering the retina; one immediately either side of the horizontal meridian); giving a total of 54 test locations, a shown in Figure 1.2.

Although defects further from the centre of the field (further from fixation) do occur in glaucoma and retinal disease, they are relatively rare, and have much less effect on visual function (and hence important factors such as quality-of-life); this, coupled with the increased variability further from fixation (Heijl et al 1989), and the reduced testing time required, has led to the 24-2 field becoming the standard in recent years.

Additionally, the HFA at certain times during the test will move as if to present a stimulus (causing the noise associated with the machine's movement), but not present the stimulus; if the patient presses the button at this point, a False Positive is recorded. Similarly, stimuli will be presented at intensities well above the patient's threshold at certain points; if the patient fails to respond, a False Negative is recorded. Thirdly, stimuli are presented at the physiological blind spot; if these are seen by the patient, it is evidence that they are failing to fixate properly (recall that the patient is looking directly forward all the time the test is happening; a fixation loss occurs when their eye 'wanders'). Tests with high levels of False Positives, False Negatives or Fixation Losses are automatically labelled as unreliable fields. The SITA algorithm (Bengtsson et al 1997) uses an alternative method for estimating the prevalence of false positives based on the number of responses which occur before the minimum possible reaction time after a stimulus has been presented.

Perimetry results in a grid of sensitivity values, on the decibel scale as described earlier. This can then be represented as a greyscale, to give a visual impression of the state of the visual field; this is then much easier to understand and interpret, both for the clinician and for the patient. The standard output from the HFA, including the Statpac2 analysis program (Heijl et al 1991a), is a printout as shown in Figure 1.3 (although it can be stored electronically as well) which includes:

- The patient's name and date of birth
- Details of the test carried out (including the testing strategy and the current date)
- The grid of the raw sensitivity values
- A greyscale representing the sensitivities
- The numbers of False Positives, False Negative and Fixation Losses
- The testing time, and the total number of stimuli presented
- Probability maps of the total deviation (comparing the sensitivity of each point with the HFA normal database) and pattern deviation (the same after mathematically adjusting to take account of a possible overall reduction in sensitivity across the field, making localised defects more evident)
- Global indices giving a measure of whether the visual field has a lower sensitivity than normal globally and/or in the form of a localised defect.

The probability maps and global indices are added to the output by the Statpac2 program.

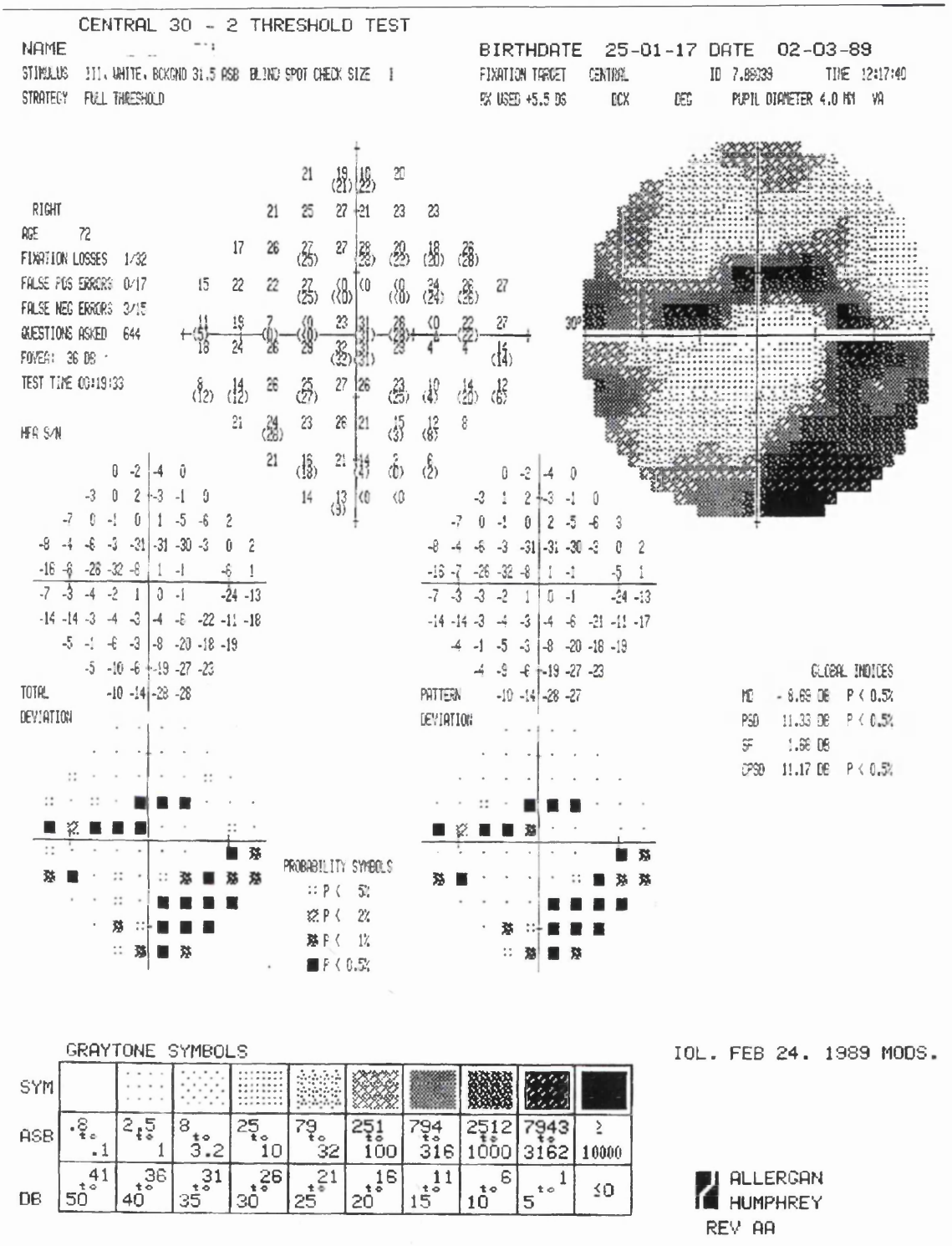


Figure 1.3: A sample printout from a Humphrey Field Analyser incorporating Statpac 2.

This study considers only the actual sensitivity values and the accompanying greyscale. Two examples are given in Figure 1.4, along with the key showing how the sensitivity values are converted into the greyscale.

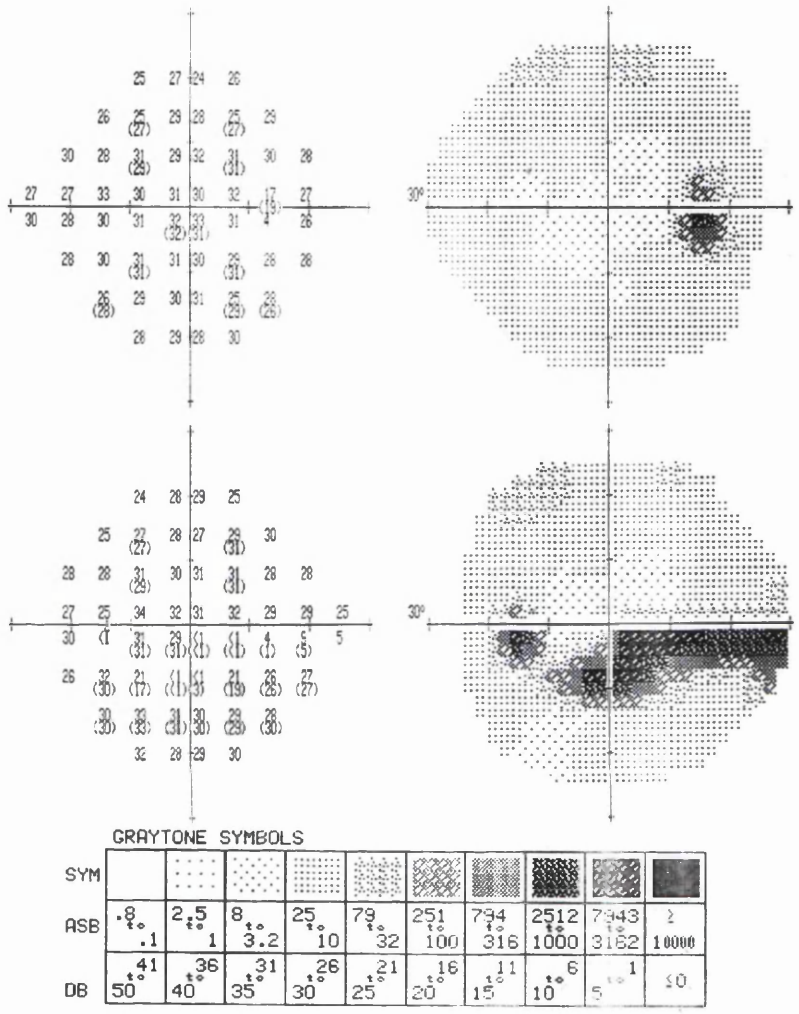


Figure 1.4: Two sample visual fields, from a HFA. The top field is from a healthy left eye; the bottom field shows a right eye with an advanced defect. The grid on the left in each case shows the raw sensitivities at each point, in decibels; where a second number is shown underneath the number, it is the second threshold estimate at that point (where applicable). The key at the bottom shows how the greyscales are generated from the sensitivities in the decibel scale (bottom row) or the equivalent luminances (middle row). Note that both of these eyes were tested using the 24-2 visual field algorithm.

In a healthy eye, contrast sensitivity decreases further from the fixation. This results in the so-called 'hill of vision' (Traquair 1931), whereby sensitivity decreases as eccentricity increases. The height of the hill represents the sensitivity at that point; stimuli equivalent to a sensitivity which would put them above this hill cannot be seen. Beyond the extent of the hill are points at which no stimuli can be seen, no matter how bright; this corresponds with points outside of the extent of the visual field (e.g. behind the patient's head).

Sensitivities in a perfectly healthy eye will still decline over time as a result of aging. This decay is slow, and would not in itself cause significant loss of vision; however it should still be taken into account when considering progression of visual fields, as in Section 1.3. This effect has been estimated as a loss of 0.1dB/year (Heijl et al 1991b).

1.2.1 Variability in Perimetry

The major limitation of perimetry is that the threshold estimates fluctuate, due to both measurement error and other factors such as the nature of the psychophysical test, whereby the results depend on a patient's response. The sensitivity of each test location can vary physiologically and become more or less sensitive over time. The variability between tests is commonly known as long-term fluctuation, and becomes an issue when examining a series of visual fields, especially when looking for evidence of progression, as discussed in Section 1.3 (Hutchings et al 2000). Also problematic is variability caused by measurement errors; repeated measurements at the same point during the same test will often produce different sensitivities. Reducing this variability, commonly referred to as short-term fluctuation (Flammer et al 1984; Lachenmayr & Vivell 1993), is a key aim of work in this field.

Over a short time, such as the duration of a test, the threshold above which a stimulus will be detected is in fact probabilistic, not deterministic. Brighter stimuli merely have a higher probability of being detected. By presenting many stimuli of the same brightness, known as the method of constant stimuli (Schwartz 1999), a graph can be produced plotting stimulus intensity versus the percentage of presentations seen; this is known as a frequency-of-seeing curve (Henson 2000, p.19-21). The point on this curve at which a stimulus has a 50% chance of being seen provides the best estimate for the threshold. With a steep slope, there is little difference in stimulus intensity between an always-seen and a never-seen response; this indicates a low short-term fluctuation. If the slope is flatter, there is greater variability. In a normal, healthy individual, the inter-quartile range of the frequency-of-seeing curve will usually be 1.5-2dB for standard automated perimetry (Chauhan et al 1993).

Many factors contribute towards the variability, including:

- Learning effects: the first time (and possibly more) a patient has been tested, they will respond less well, typically causing artificially low threshold estimates (Wood et al 1987).
- Fatigue: tests can frequently last for more than 15 minutes per eye for the Full Threshold algorithm. Patients may well lose concentration during the test, and so fail to respond to stimuli (Hudson et al 1994; Wall et al 2001).
- Alertness: similarly, concentration may be lower if the patient is for example more tired that usual on the day in question.
- Button pushing: the patient responds to a seen stimulus by pressing a button. This is open to errors due to the button being pushed too slowly, or not fully depressed, or pushed accidentally.
- Fixation losses: during the test, the patient is asked to look directly forward at a set point. If their eye wanders, the visual stimuli will be presented into the wrong part of the retina. Similar effects can be caused by the head tilting during testing; usually the chin is rested on a stand to reduce head movements, but this system is not perfect. However, this is now thought to be unlikely to be a major cause of variability (Henson et al 1995; Henson et al 1996).
- Ocular media opacities: the presence of cataracts causes a general diffuse loss of sensitivity in the visual field (Smith et al 1997; Chen & Budenz 1998).

Some of these factors are more controllable than others; furthermore, some are more quantifiable than others. More in depth discussions of each of these, and other, factors are given by Lachenmayr & Vivell (1993) and Henson (2000, p. 50-64). Not all variability can be explained by these factors; the cause of much of the noise is unknown (Blumenthal et al 2000). Some patients will produce very reliable results, with little variation over time; others will produce very different threshold estimates at the same point during the same test, and also from one test to the next. A point whose threshold is estimated at 20dB one day could be repeated the following day, and the estimate be anywhere from 0dB to 30dB (Heijl et al 1989; Chauhan & Johnson 1999b). Figure 1.5, reproduced from Artes et al (2003), shows the 5th and 95th percentiles of the distribution of the sensitivity at a second test for each value of the sensitivity at an initial test, showing the alarmingly large test-retest variability of full-threshold perimetry. This inevitably makes identification of defects, and their follow-up over time, very difficult.

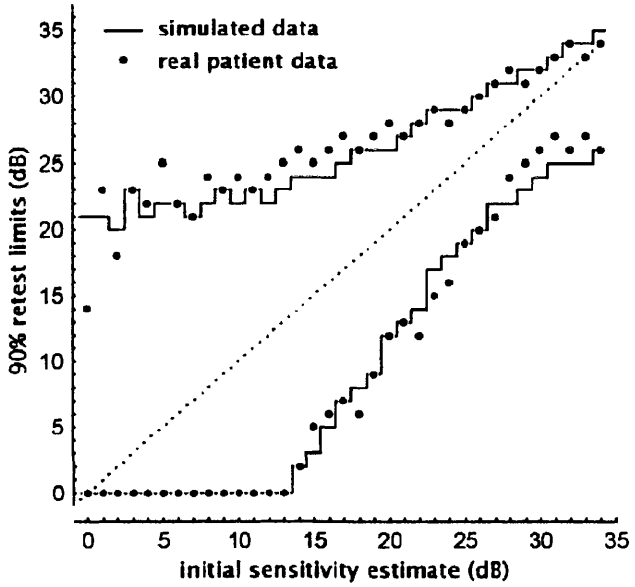


Figure 1.5: Pointwise 90% test-retest intervals for full-threshold perimetry, reproduced from Artes et al (2003). For each threshold level of the initial test, the circles represent the 5th and 95th percentiles of the distribution of retest values based on patient data. The solid lines indicate the percentiles established from simulations of the testing algorithm.

It is widely agreed that variability increases as the sensitivity of the point in question decreases (Heijl et al 1989; Heijl et al 1991b; Chauhan & Johnson 1999b; Henson et al 2000; Artes et al 2002). This means that the variability will naturally increase when a patient has a glaucomatous defect, and with age, and with eccentricity.

A more in-depth discussion of the causes of variability in visual fields, and methods for testing it, is given by Spry & Johnson (2002).

1.3 Progression

As yet, there is no accepted mathematical model or universally recognised quantitative technique for following how glaucomatous visual fields change over time. Because of ethical considerations, no data is available on how glaucomatous eyes would change if left untreated, except in special cases such as the

subset of patients who exhibit normal tension glaucoma (McNaught et al 1995), or in very early stages of the disease (Heijl et al 2002). It seems reasonable to assume that defects occurring in treated eyes probably exhibit the same patterns of change as untreated eyes, although at lower rates (Spry & Johnson 2002). However there are still few longitudinal studies providing information about the evolution of progressive glaucomatous visual field loss. Mikelberg & Drance (1984) studied loss in 42 glaucomatous eyes with progressive defects, over a mean follow-up period of 8.2 years; they found that defects became deeper in 79% of the eyes, larger in 52% of the eyes, and new defects occurred in 50% of the eyes. However, this was a small study based on manual perimetry, which (as explained in Section 1.2) is less reliable than automated perimetry.

What is known from clinical observations is that change is very slow, measured over several years rather than weeks. This means that it is necessary to follow patients over a period of several years to answer the two key questions: does the eye have a glaucomatous defect, and if so is it getting worse? If the defect is worsening – called a progressing defect – then treatment can be applied to slow or halt this deterioration (Hitchings et al 1995; Bhandari et al 1997; Koseki et al 1997; AGIS 2000). If the defect is stable, then unnecessary treatment can be both painful and costly; many treatments for glaucoma have associated side-effects (Coleman 1999). Furthermore, treatment has recently been shown to increase the risk of nuclear cataract (Heijl et al 2002; Leske et al 2002; Wormald 2003). Also, in clinical trials it is important to have some means of determining whether one group of glaucomatous eyes are deteriorating faster than another group. This raises the question of how to best detect true change in visual fields over time.

Clinical judgment, consisting of a simple subjective observation of sequential visual field test results (normally in greyscale form), and other information on the perimeter output, is probably the typical method for identification of progressing defects; and is still commonly used in glaucoma clinics. It is quicker, since it involves no computer use; it is easy to perform; and it is highly flexible, since an experienced clinician can take account of other factors such as the appearance of the optic disc, and other clinical and non-clinical factors. However, this subjectivity is also the method's downfall; the same series of fields can be interpreted differently by different people, and agreement between even experienced clinicians has been shown to be very poor (Werner et al 1988; Viswanathan et al 1997b; Viswanathan et al 2003). Table 1.1, reproduced from Viswanathan et al (2003), shows how poor the agreement between two experienced clinicians is, even when based on an exceptionally long series of sixteen fields.

		Observer B		
Observer A	Definitely Stable	Probably Stable	Probably Progressing	Definitely Progressing
Definitely Stable	**5**	4	1	0
Probably Stable	4	**1**	2	1
Probably Progressing	0	3	**1**	3
Definitely Progressing	0	1	0	**1**

Table 1.1: The differing clinical opinions of two experienced observers on the status of 27 series of 16 visual fields.

Early attempts at more formal trend analysis were still largely subjective; based on either an estimate of the area of the field where the perimetric stimulus was seen (Smith 1972; Sponsel et al 1983; Smith

1986), or on classification into (essentially arbitrary) stages based on the size and depth of the defect (Jay & Allan 1989), all based on manual perimetry. More objective analysis became possible with the advent of repeatable automated perimetry. Initially, trend analyses were carried out on global indices summarizing the field. The first version of the Statpac analysis package (Heijl et al 1987), developed for the HFA, tests whether the mean deviation (i.e. the mean difference between the sensitivities of each point and the age-corrected normal values for those points) is increasing over time, using a simple linear regression (subject to there being enough fields in the series). In Statpac2 (Heijl et al 1991a), this is adjusted so that the first test result is discarded if it deviates substantially from the trend, to take account of learning effects. Other global indices (such as the pattern standard deviation) have been used similarly. However, clinical opinion (Hitchings 1993; Johnson 1993; Wild et al 1993) and research findings (Chauhan et al 1989; Chauhan et al 1990; Katz et al 1997) suggested that following global indices over time was insufficiently sensitive to glaucomatous loss; the reduction of the data to just one measure removes much useful information on the distribution of loss through the field, ignoring the important spatial nature of visual field loss in progressive glaucoma.

A similar technique used is to give each visual field a score, giving an objective indication of the amount of glaucomatous loss. One example of this is the AGIS scoring system (AGIS 1994), where the field is given a score based on the depth and size of defects in each of three areas – the upper and lower hemifields, and the six most nasal points, in the 24-2 field – as shown by the total deviation map on the HFA printout. This results in a score from 0 to 20, where a score of 20 indicates that all test sites are deeply depressed. The Collaborative Initial Glaucoma Treatment Study (Musch et al 1999) produces a similar score from 0 to 20, this time based on the total deviation probability plot provided by Statpac 2 (Heijl et al 1991a), described below. While this technique is attractive for producing objective results for clinical trials such as these, it is of less use for clinically examining a patient, since it give no indication of the spatial profile of any defect. It also exhibits a large variability between tests; the AGIS investigators found that the scores of 16% of the 756 eyes in the study varied by four or more points when retested after a short interval (median follow-up time one week).

This led to the consideration of so called pointwise methods for identifying progression. In a 'pure' pointwise method, each point in the field is examined separately to determine whether the sensitivity at that point appears to be deteriorating over time. Clinically, it is common to look for clusters of two or more neighbouring points in expected locations (i.e. adjacent along one of the arcs which localised glaucomatous defects follow, as described in Section 1.1) which all appear to be progressing. While such methods are not technically 'pointwise' in the usual definition of the term, the methods for determining which individual points are progressing are still the same as if a pure pointwise technique was being used; and so these too are referred to as pointwise methods. It has been shown that such methods identify progressive field loss which would have been masked by looking at changes in global indices (Wild et al 1994; Nouri-Mahdavi 1997).

One commonly used package for analysing pointwise changes in glaucomatous visual field series is the Statpac 2 program (Heijl et al 1991). The 'Glaucoma change probability analysis' evaluates the amount of change with respect to empirical results of repeated tests, from a population of patients with stable

glaucomatous field loss. Two of the first three fields in the series are selected and averaged to give a merged baseline field. The change from this baseline, in decibels, is calculated and displayed for each test location; this is to highlight locations where the sensitivity has changed by more than would typically be observed were the field stable, based on the location of the point in question, the initial sensitivity loss and the mean deviation of the field as a whole. It results in a map of symbols; for example a solid triangle at a point indicates a deterioration which would be found less than 5% of the time in a stable field (so the principle is similar to, although not identical to, a statistical p-value). Additionally, the program evaluates change in the mean deviation of the whole field, in the same manner.

One limitation of the Statpac2 program is that its results are based solely on comparing the field with the baseline field(s); intermediate tests are ignored. It is really an event analysis rather than a trend analysis. One result of this is that changes detected always need to be confirmed by another test; another result is that the test is highly dependent on the accuracy of the baseline field. Further, the normal database which the probability values are based on is taken solely from patients with stable, medically treated POAG; results from patients whose circumstances are different need to be treated with caution (Morgan et al 1991).

To use all the available data, regression analysis is necessary. Here, each point is considered as a separate time series, and a line found which predicts the values at each point in time, whilst minimising the errors at the times when the sensitivity was measured. Previously, curve-fitting software has been used to identify a best-fit model for glaucomatous sensitivity deterioration (McNaught et al 1995). Complex polynomials produced the best fit, but it was suggested that these polynomials were in fact modelling the noise more than the actual deterioration. The predictive power for the next test in the series was seen to be much better using a simpler linear model; that is, when the line of best fit was calculated based on only the first five sensitivity measurements, and then the sensitivities at future time points estimated based on this line, the linear model produced the smallest errors. Thus the investigation concluded that linear modelling was the best (out of those tested) for determining progression. Linear approaches have been used in other longitudinal visual field models (Wild et al 1993; O'Brien & Schwartz 1993; Spry 2000); and while there is no way of saying that such a model is optimal, it has gained pre-eminence as the most accepted model of those examined so far.

Pointwise Linear Regression (PLR) has been implemented into a computer program called Progressor (Institute of Ophthalmology & Moorfields Eye Hospital, London, UK). This accepts data from a computer disc or directly from the perimeter, and calculates the relationship between threshold sensitivity and test date, displaying the degree of change (and its statistical significance) in a grid form so that the important spatial relationships are not lost (Fitzke et al 1996); an example output is shown in Figure 1.6. It has been shown that the use of this program significantly improves the agreement between experienced clinicians on the matter of whether a field is progressing or not (Viswanathan 1997b; Viswanathan 2003); and that the results obtained compare favourably with those from the Statpac 2 program (Birch et al 1995b; McNaught et al 1996; Viswanathan et al 1997a).

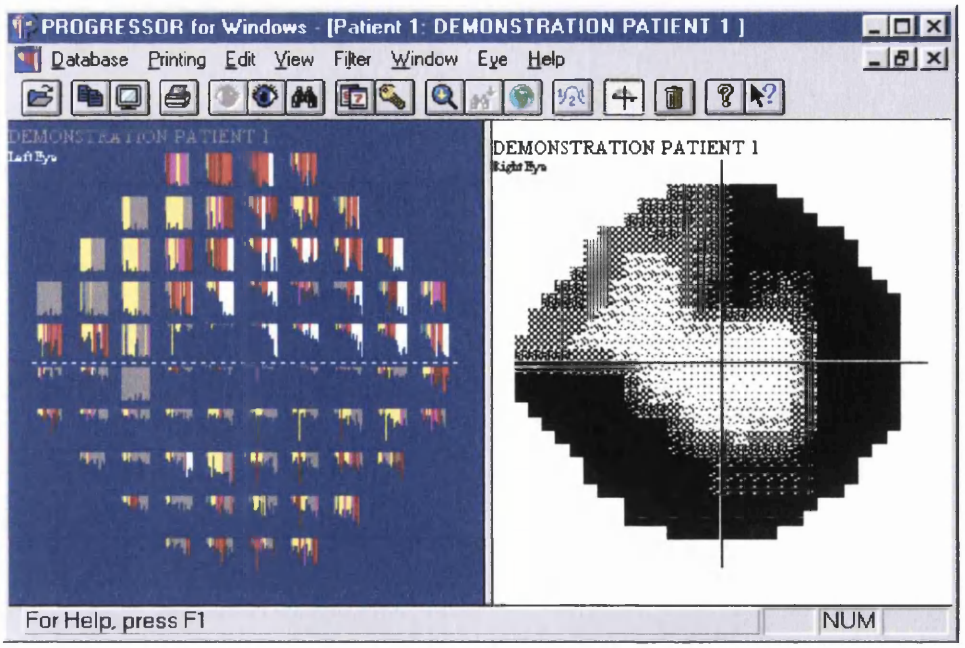


Figure 1.6: Sample output from the Progressor program. Visual fields can be displayed individually in greyscale form (as with the right eye here); alternatively, colour-coded bars can be displayed for each point (as with the left eye here), with the size of each bar representing the deviation from normal sensitivity for that point at that test, and the colour representing the statistical significance of the regression slope up to and including that test. A series of bars is shown at each point, showing the deviation and significance at that point for each test in the series up to and including the currently selected field. This enables progressing points to be easily identified. Such points can also be automatically labelled on the greyscale diagram; the exact criteria for slope and significance level are adjustable by the user.

Pointwise Linear Regression (PLR) is not universally accepted. In its pure form, it makes the possibly unwarranted assumption that different locations progress independently; it takes account of localised defects deepening over time, but not of defects spreading into neighbouring locations. It implicitly assumes that test locations are independent, which is not strictly true (Hoffman et al 2003). This problem is lessened, though not eliminated, by methods which look for clusters of adjacent individually progressing points, or by methods utilising spatial filtering (of which more in Chapter 5). Specificity is also sacrificed (O'Brien & Schwartz 1992); points will be falsely labelled as progressing even in a stable eye because of the random nature of the noise. Again, this effect is reduced when clusters of progressing points are sought, or when further confirmatory tests are performed. It is non-contentious to say that PLR detects more cases of progression than global indices, suggesting that it has a higher level of sensitivity, with an accompanying lower level of specificity (Smith et al 1996; Katz et al 1997). Because there is no gold standard definition of glaucomatous field loss, it is currently impossible to say which has the better discriminatory power. The choice will inevitably vary according to the purpose of the testing; sometimes it is better to be conservative with diagnoses, at other times (for example when examining a patient whose other eye already has partially or completely reduced vision) any change, no matter how small, is crucial.

An additional difficulty associated with using this technique is that the amount of deterioration required before a point is flagged as progressing depends on the exact criteria used. Some studies have defined progressive loss as any statistically significant negative slope, irrespective of gradient, which has a p-value of below 0.001 (Smith et al 1996; Katz et al 1997). Other common criteria are for slopes to be steeper than –1dB/year accompanied by a p-value less than either 0.05 or 0.1 (Wild et al 1997; Birch et al

1995a; Viswanathan et al 1997a; Viswanathan et al 1997c; Membrey et al 2000; Membrey et al 2001). In the Progressor program, the criteria can be altered by the user; the default is a slope of steeper than –1dB/year with a p-value less than 0.1 for inner points, and a slope of steeper than –2dB/year with a p-value less than 0.1 for edge points (points at the periphery of the Humphrey 30-2 field, i.e. those points appearing in the 30-2 field but not in the 24-2 field). Stricter criteria will inevitably improve the specificity whilst reducing the sensitivity.

There is also no universally accepted criterion for how many points in the field must be flagged as progressing in order for the eye to be said to be progressing, or whether those points should be adjacent. Nor is there any agreement on how often visual fields should be tested (Viswanathan 1997c; Gardiner & Crabb 2002a), which will inevitably have an effect on the significance levels of slopes; this question forms the motivation for the work done in Chapter 3. Further, there is disagreement over whether an apparently progressing point must be confirmed by a further test or tests; different criteria have been used in different studies (Hitchings et al 1994; Birch et al 1995a; McNaught et al 1996; Smith et al 1996; Katz et al 1997; Nouri-Mahdavi et al 1997; Viswanathan et al 1997a; Viswanathan et al 1997c; Wild et al 1997; Kamal & Hitchings 1998; Leske et al 1999; Viswanathan et al 1999; Membrey et al 2000; Spry et al 2000; Hofmann et al 2001; Membrey et al 2001; Westcott et al 2001; Gardiner & Crabb 2002b; Heijl et al 2002; Kim et al 2003). This question forms the motivation for the work done in Chapter 4.

Other methods for identifying progression have been postulated. These include using neural networks, which do not rely on the assumption of independence between different points (Goldbaum et al 1994; Spenceley et al 1994; Lietman et al 1999); mixed topographical and longitudinal models (Wild et al 1993); and pointwise multivariate regression analysis with fixed effects on panel data (Nouri-Mahdavi et al 1997). Although some of these techniques are still under investigation, they have addressed many of the drawbacks of the current progression algorithms, potentially leading to a better understanding of glaucoma progression. However, to date no conclusive evidence has been presented showing these methods to be definitively better than PLR.

1.4 Aims of Study

This study aims to improve the analysis of visual field results from HFA perimetry. Given a series of test results over time, how is it possible to distinguish better between progressing and stable eyes? In particular, the study sets out:

- To evaluate the current PLR criteria and methodologies.
- To suggest improvements for those criteria and methodologies, where possible.
- To reduce the variability of threshold estimates that limits these methodologies.
- To implement resulting techniques for variability reduction into clinically useful software.

2. Virtual Eye Simulation

The major aim of this work is to evaluate and improve the current methodologies for testing visual fields for glaucoma. Specifically, given one or more visual fields output from an automated perimeter, it is sought to improve the way these results are used to determine the presence or absence of glaucoma, and whether or not the disease is progressing (note that this is different from trying to improve the methods used to obtain visual fields using an automated perimeter). To achieve this, it is essential to have a fair, reliable, unbiased testing procedure for comparing methods.

2.1 Previous Work

There is currently no widely agreed 'gold standard' for comparing procedures for testing for progression. It is not as simple as finding the detection rates for glaucomatous eyes, since there is no true cut-off point distinguishing a glaucomatous eye from a healthy eye; even a perfectly healthy eye will show a deterioration in sensitivity over time due to aging. There are cases when a glaucomatous defect can clearly be seen on a visual field test result, such as in Figure 1.4 in the previous chapter; but there are many more cases where the diagnosis is less clear-cut, especially during the crucial early stages of the disease. In Figure 2.1, for example, it is very unclear whether the eye in question (taken from a demonstration eye provided with the Progressor software described in Chapter 1.3) is progressing or not over the three year time period.

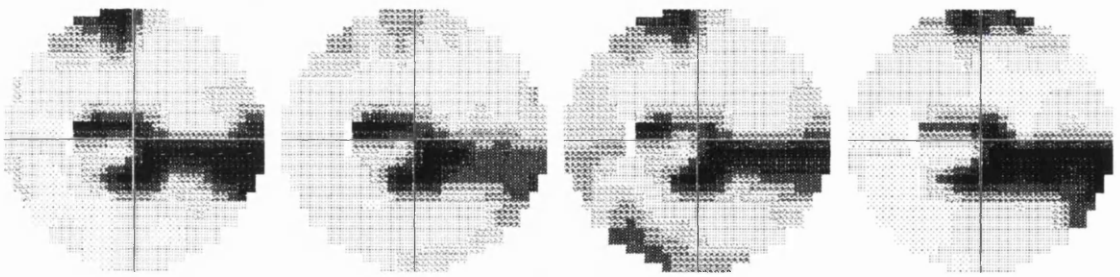


Figure 2.1: Example series of greyscales. The four fields were taken on (from left to right) 24/4/86, 16/7/87, 9/6/88 and 11/7/89.

Given the irreversible nature of glaucomatous damage, the key question is not whether or not a defect is present; more important is whether the defect is progressing (and hence requiring treatment) or remaining stable. Early detection of defects is keenly sought, yet this relies on the same principle of looking for change. The variability described in Section 1.2.1 means that noise can easily be confused with progression. Assessing series of visual fields by hand leads to very poor agreement between experienced clinicians (Werner et al 1988; Viswanathan et al 1997b; Viswanathan et al 2003), as shown in Table 1.1; hence the development of quantification methods, including PLR, as described in Section 1.3. It is these PLR methods which we wish to test. So, we require series of visual fields where the true state of each eye is known, so that it can be determined whether a diagnosis of that eye as deteriorating is due to true progression or simply due to noise. Ideally, we would like to base all testing on real patient data;

however, the true (noise-free) status of these eyes is unknown, and so comparing test results from different methodologies with the actual status is impossible.

Therefore, it is natural to look for ways to specify the noise-free behaviour of an eye over time; then when noise is added in, the series can be tested for progression, and the test results compared with this specified behaviour. There are two ways to do this:

- Theoretically. When a simple form of noise is added in, it may be possible to determine the probability of a correct test result. This is mathematically appealing; but it does require assumptions to be made, and even then many problems posed will be intractable.
- Numerically. Noise can be added in by randomly sampling from a given distribution, or probability model, to simulate noisy visual fields; and then the test results evaluated in each case. This requires many simulated series to be determined. This will result in sample probabilities of a correct test result being generated; which, given sufficient runs of the simulation model (the more the better), will approximate the exact probability. This method is more flexible, as it enables approximations of arbitrary precision to be made, whether or not the theoretical problems are tractable.

A key question in either approach is the form which the noise takes. The biggest assumption made in both cases is that the noise will follow a certain distribution. Test-retest data, where the same point in the field has been tested more than once under the same conditions within a short time period (typically less than a week, so that it is too short a time period for any true change to have occurred), gives one way of estimating the noise. While there are problems with this method of noise estimation – most notably the fact that because both the initial test and the retest are noisy, the estimate is actually double the noise present in one test – this does give a good indication of the shape of the distribution of the noise. This has been done in different sets of conditions, producing similar results (Flammer et al 1984; Heijl et al 1989; Heijl et al 1991; Chauhan & House 1991; Chauhan & Johnson 1999b; Henson et al 2000; Artes et al 2002; Artes et al 2003). It is seen that at sensitivities of around 30dB (a typical reading for a healthy eye), the noise is approximately normally distributed, albeit with slightly fatter tails. As the sensitivity decreases, the noise becomes progressively more negatively skewed, with the left tail being fatter than the right tail. For further details, see Chapter 7 (and in particular Figure 7.1).

The amount of noise also increases as the sensitivity decreases; Henson et al (2000) estimate that the standard deviation *SD* when the sensitivity is *S* is given by:

$$log(SD) = -0.162S + 6.54$$

Equation 2.1

This equation was found by looking for a line of best-fit for *log(SD)*, based on frequency-of-seeing data (as described in Chapter 1.2.1) collected in both patients and normal subjects at different visual field locations. This means that it gives a prescription for the level of noise which a perfect testing strategy would attain; the testing strategy chosen will increase this amount. It should also be noted that the noise is censored; that is, since sensitivities cannot be measured below 0dB, the distribution of the noise reflects this. However, the consensus is that noise is best modelled by a normal distribution, since it is eyes which are healthy or which exhibit early glaucomatous defects, and hence which have sensitivities well above zero, which are of most interest clinically. Although this assumption of normality is clearly not wholly

accurate, it is used in all the current literature on using simulation for visual field progression (Crabb et al 1999; Spry et al 2000; Vesti et al 2002). In Chapter 7, empirical results will be used to examine the assumption, and to develop a better estimate of the distribution of the noise.

There are currently two independently developed simulation models for progression in the literature. The first, called the 'FieldSim' model (Crabb et al 1999) followed on from work on PLR. This model assigns either gradual loss (at a user-defined rate) or episodic loss (by a user-defined amount) to a point over time, and then adds noise by sampling randomly from a normal distribution (with a fixed, user-defined standard deviation). The program is run in SPlus (StatSci Europe, MathSoft Inc., Oxford, UK). This is the program upon which the Virtual Eye model described below was based.

This 'FieldSim' simulation model was used to demonstrate that PLR detected episodic change (a one-off loss of $3y$ dB at a random point over a three year period) just as well as it detected gradual loss (deterioration by y dB per year over three years) for different values of y, when the noise was normally distributed with a standard deviation of 3dB, and tests were carried out every six months over that period. This experiment addressed a possible problem with using PLR to detect progression, namely that it assumes loss is gradual rather than episodic, an as-yet unproven assumption.

The second simulation model was developed by Spry et al (2000). This differs in four key ways:

1. Two forms of noise are used. First, Short-Term Fluctuation is added by randomly sampling from a normal distribution for each point in turn, as in the 'FieldSim' model. Additionally, Long-Term Fluctuation is added by taking another random sample from a normal distribution, and adding this value to all the points in the field. This has no effect when only one point is being considered, since the sum of two normal variables is another normally distributed variable; it only has an effect when neighbouring points are being considered together.
2. The initial and final fields of the series are taken from real patient data. Then, the intermediate fields are derived by interpolating between these two fields and adding noise. The effect of this is that the initial and final fields are effectively noise-free; no account is taken of noise present in these readings. Given that the first and last data points have a much larger effect on the slope of a regression line than the intermediate points, this has major consequences for PLR and any other regression-based techniques.
3. A function is included which can, when desired, produce an increase in the short-term fluctuation with increasing eccentricity. Although this phenomenon has been noted in the literature (Heijl et al 1987; Katz & Sommer 1987), it has been suggested that the increased average variability further from fixation is due purely to the sensitivities decreasing (Flammer & Zulauf 1985), and hence would also be accounted for by using Equation 2.1.
4. The short-term fluctuation can (if desired) be set to increase its standard deviation by 0.4dB for every 5dB of sensitivity loss. This is a simplified form of the principles accounted for by using Equation 2.1.

This simulation model was used to quantify the effects of different levels of noise on the detection rates for progression using PLR (Spry et al 2000); and in a study looking at the effect of applying the Gaussian

filter to visual field data (Spry et al 2002), which will be discussed further in Chapter 5. It has also been validated by examining the simulation's predictive power (Vesti et al 2002); by extrapolating the linear fit and adding noise, how well do the simulated noisy fields match the actual field from the patient in question? Unsurprisingly, it was found in this study that when no noise (in the form of short-term fluctuation) was added in, the predictions were more accurate than when noise was added to the predictions. The study also drew the conclusion that short-term fluctuation is a far more important factor for determining total variability than long-term fluctuation, conforming to the empirical findings of Flammer et al (1984). This gives further support for the principle of simply adding one form of noise to simulated fields, in the form of short-term fluctuation affecting each point in the field individually and independently rather than affecting all the points by the same amount, especially when the noise has been assumed to be normally distributed.

Computer simulation of visual fields has also recently been used to examine testing strategies for a single field test (Artes et al 2003). This simulation sets the sensitivity at intermediate points in the series by linear interpolation as before, and then for each stimulus which would be presented during the field test (by whichever testing strategy is being considered at the time), determines probabilistically whether or not that stimulus was seen based on the sensitivity of the point and the variability (taken as being the slope of the frequency-of-seeing curve, as described in Section 1.2.1). The non-standard noise distribution which results, whilst possibly being more realistic than a normal distribution, makes theoretical comparisons between methods of the sort employed in Section 4.2 impossible. However it does appear promising in terms of the resulting shape of the noise distribution (something which will be examined further in Chapter 7), and takes account for the differences in the shape of the noise distribution caused by the choice of testing strategy. It is reasonable to hope that features from this simulation model could in the future be incorporated into the Virtual Eye model described in the next section, to create an improved model which could be used to examine field progression.

2.2 The 'Virtual Eye' Simulation Model

In order to address the problems being looked at in this work, an improved simulation model was required. Any test results are only as good as the methods used to obtain them. So, an advanced simulation, hereafter referred to as the Virtual Eye, was developed; using the 'FieldSim' as a starting point, but taking into account new developments in the theory of visual fields as well as learning from the model developed by Spry et al (2000). The Virtual Eye was developed using SPlus computer software. The user inputs:

1. The length of the series desired;
2. The number of tests per year to be simulated (tests are assumed to be equally spread in time);
3. The noise-free initial and final fields in the series (or equivalently the noise-free initial field and the rate and pattern of loss over time).

The model then assumes that the sensitivities at each point deteriorate linearly over time; and so the noise-free sensitivities at any point in time can be found by interpolating between the initial and final fields in the series. This assumption of linear deterioration is in keeping with the simulations by Spry et al

(2000) and Artes et al (2003), and uses the results from McNaught et al (1995) showing that a linear model was preferable to a polynomial or exponential model in terms of its predictive power for future field results (since more complicated models such as higher-order polynomials tend to model the noise rather than the underlying trend).

Next, the Virtual Eye adds noise to each point of each generated noise-free field in turn. It does this by replacing the sensitivity value at the point in question with a Monte-Carlo randomly generated value, sampled from the normal distribution whose mean is equal to the noise-free sensitivity, with a standard deviation as given by Equation 2.1. This means that the noise increases as the sensitivity decreases in the manner described by Henson et al (2000), and hence in a more accurate manner than that used in the Spry et al simulation model. So, for example, if the true sensitivity at a point is 28dB, then the simulated sensitivity is drawn randomly from a normal distribution with mean 28dB and standard deviation 2.72dB. Only one estimate of the noise is used, in effect the sum of the short-term and long-term variability components. Note that the use of Equation 2.1 to describe the noise means that the level of noise attained by a perfect testing strategy (i.e. one with an unlimited number of stimulus presentations per point) is being considered; different testing strategies will add different amounts of noise to this figure.

The Virtual Eye makes the simplifying assumption that the noise at neighbouring points is independent, given the noise-free sensitivity at those points. We are not assuming that actual (noise-free) progression is independent between points – it will not be. However because we have specified this progression exactly, the noise-free behaviour of neighbouring points is irrelevant. This means that we are effectively considering only the short-term fluctuation rather than the long-term fluctuation, something which, as explained above, is reasonable. This has two benefits. Firstly, the programming is much simpler, because points can be considered individually, which also puts much less pressure on the computer's memory. Secondly, conclusions can be drawn based on the proportion of points, for example, being flagged as progressing; if the whole eye needed to be considered as one entity then presentation and interpretation of the results would be far more complicated. However, there are times when it becomes necessary to simulate an entire visual field, rather than just an individual point (as in Chapters 4 and 6). The method used is unchanged. The actual (noise-free) field at each point in time is specified, and then noise added to each point in the field in turn. Although the simulation model is no more complicated in this case, more care must be taken with the extraction and interpretation of results. At the moment we have no model of how defects spread to neighbouring points; but the method used here negates this potential problem.

Another feature of note for the Virtual Eye is a simple correction for the censored nature of the decibel values caused by the instrument used for clinical testing. That is, if a generated value is below 0dB or above 36dB then it is assigned to be precisely 0dB or 36dB respectively. Also, the dB values are all rounded to the nearest integer (as they would appear from a Humphrey perimeter). The 'visual fields' are represented, stored and output as numerical arrays on which further analysis can be carried out.

The principles are best illustrated by an example. The results from a simulation can be displayed as individual plots of sensitivity values against time of follow up. In Figure 2.2, the top row shows the actual sensitivity values as specified for a stable eye. This is defined to be one which shows only the expected

age-related deterioration of 0.1dB/year (Heijl et al 1991b). The lines are those produced by linear regression based on the first three points of the series, then the first four points, and so on; so the graphs show how the results would change as more tests were carried out in subsequent years and added on to the end of the series.

Each point on the graph can be thought of as a threshold deviation (dB) from baseline labelled as 0 dB on the vertical axes. No assumption is made about the actual starting sensitivity. For this illustration it is assumed that the visual field tests are taken at yearly intervals. The criteria for a point being flagged as progressing are specified as a regression slope worse than –1dB/year which is also statistically significant at the 1% level.

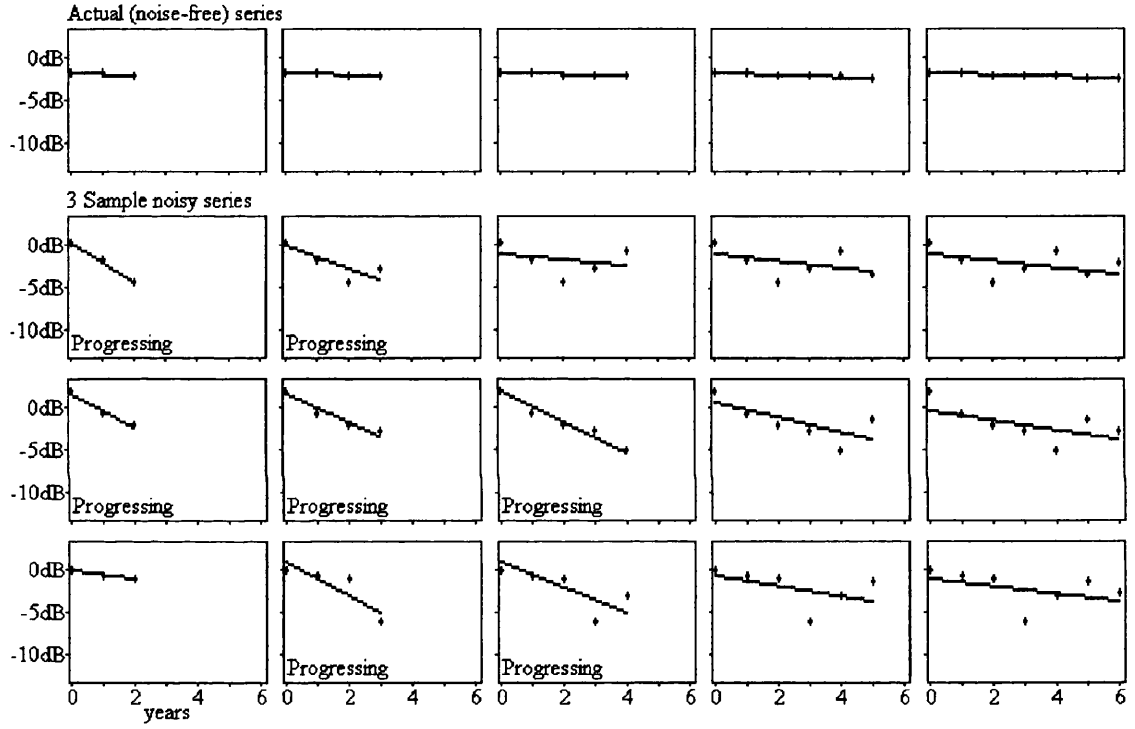


Figure 2.2: Illustrative series for a stable Virtual Eye, with an age-related decline of 0.1dB/year.

Subsequent rows illustrate three different examples of possible artificial series; noise has been added to the actual noise-free values in the top row. It is now seen that due to the effect of the noise, it is possible for the series to be labelled as progressing according to the specified criteria even though we know that the point is actually stable. As time goes on, the diagnosis will get more accurate, with fewer stable series being falsely labelled as progressing.

In Figure 2.3, the actual sensitivity of the point (as seen in the top row) is deteriorating at a rate of 2dB/year. This represents approximately twenty times the normal rate of decay (Heijl et al 1991b), and double the criteria for progression of 1dB/year which we are using for PLR (see above). When noise is added, the point is regularly labelled as being stable, even though we know this not to be the case. This is either because the slope is not sufficiently steep, or because the series is too noisy for the slope to be statistically significant.

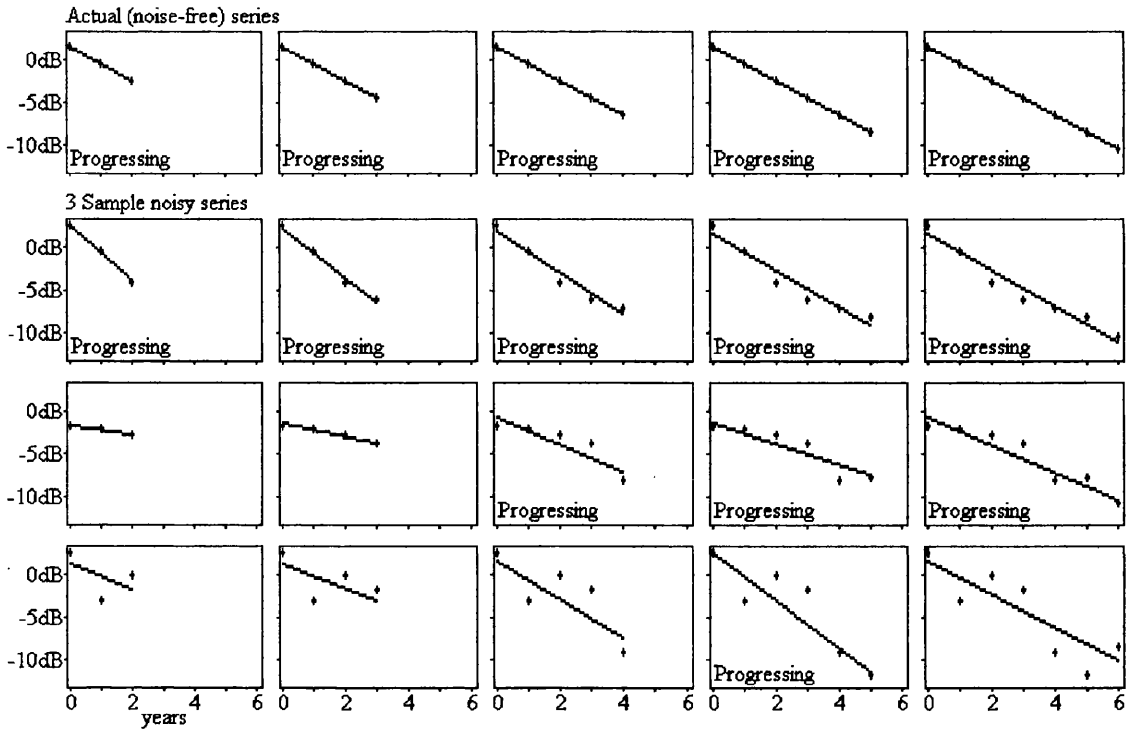


Figure 2.3: Illustrative series for a Virtual Eye progressing at 2dB/year.

Naturally, in both stable and progressing cases, as the series gets longer a higher proportion of series will be correctly flagged.

The Virtual Eye can be easily incorporated into other purpose-written SPlus programs. These programs can then calculate the linear regression slope and p-value based on the noisy series; thus the result of applying PLR to the series can be generated, to determine which points would be flagged as progressing according to different (user-defined) criteria. This forms the basis for all the simulation experiments carried out in later chapters. Experiment-specific details, such as the initial field and the exact PLR criteria used to flag points as progressing, are included in the description of each experiment.

An example of the use of the Virtual Eye is given in Appendix 1; this shows the SPlus program used to implement the Virtual Eye to carry out the experiment described in Chapter 4.

Naturally, the Virtual Eye can be used to simulate an entire visual field rather than just one point. Figure 2.4 shows three series of greyscales, representing annual field testing over a six year period. The first column is a noise-free series with six points progressing at 2dB/year, and all other points remaining stable. The second and third columns are produced by using the Virtual Eye simulation to add noise to each point in each field; it becomes very difficult to distinguish the series based on the eye with six progressing points (the middle column) from the series based on a completely stable eye (the right-hand column).

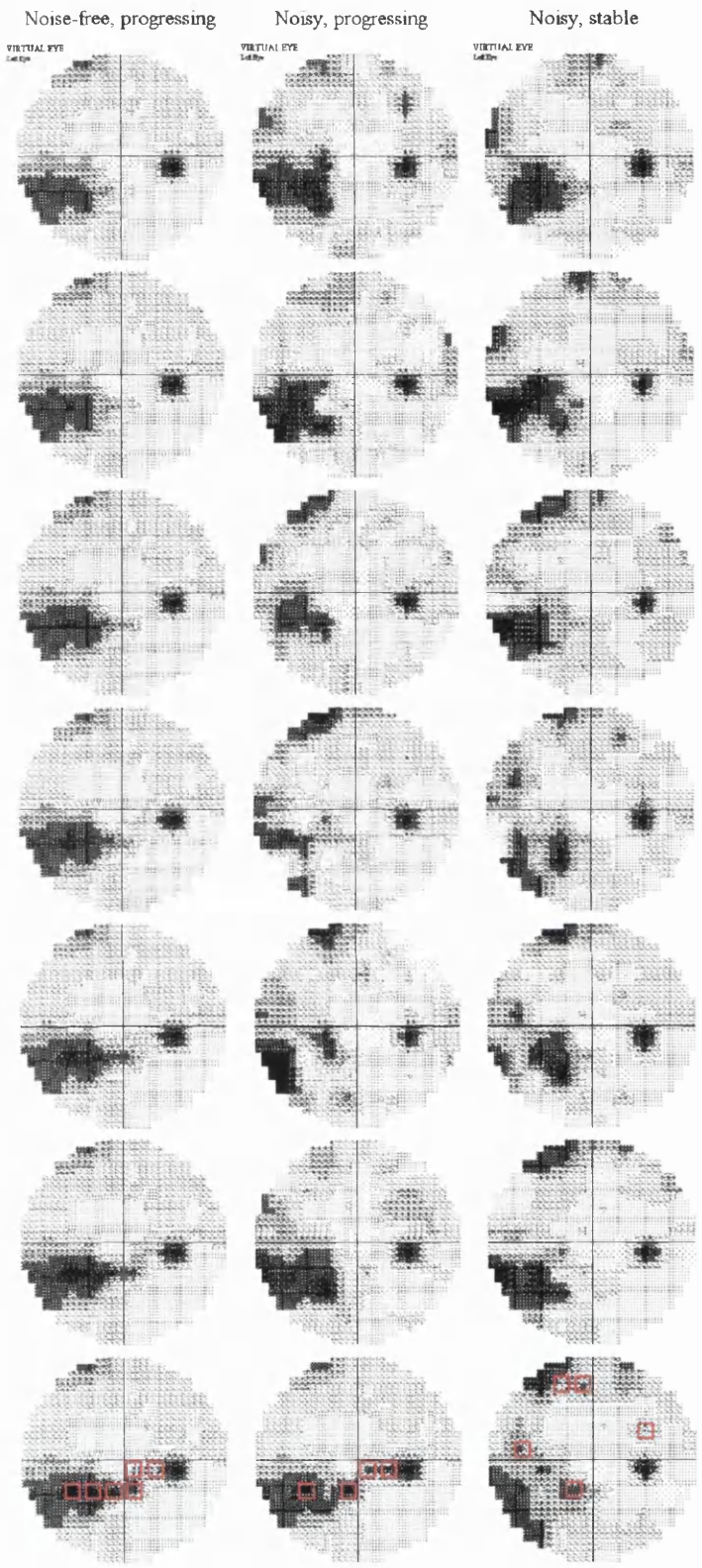


Figure 2.4: Three series of visual fields, demonstrating the effect of the Virtual Eye simulation. The left-hand column shows a noise-free field in which six points (those highlighted in red in the final field of the series) are deteriorating at 2dB/year over six years. The middle column shows the same series when noise has been added by the Virtual Eye; note that now only four of the six progressing points are identified by PLR after six years. The right-hand column shows a stable series after noise has been added by the Virtual Eye; note that five points are still (incorrectly) flagged as progressing by PLR. Comparing the middle and right-hand columns shows the difficulty of separating progressing from stable eyes, especially when based solely on examining the greyscales by hand.

3. Frequency Of Testing

In this chapter, the Virtual Eye described in Section 2.2 is used to determine the optimum number of visual field tests which should be carried out per year to detect glaucomatous progression. Too few tests may result in progression being missed, or alternatively falsely identified; whilst too many tests per year places an undue burden on the patient and clinic in terms of time, inconvenience and resources. Previously, Viswanathan et al (1997c) concluded, based on patient data, that an increase in the frequency of visual field testing from one examination per year to three per year provides more rapid, more sensitive diagnosis of progressive glaucoma. This was done by analysing series of visual fields, with three tests carried out per year; and then repeating the analysis on 'thinned' series, with only one of the tests per year included. One hundred and nine retinal locations which were deteriorating significantly by at least 1dB/year were studied. The results of the analyses on the two series were then compared, and it was found that over a four year period, the 'thinned' tests identified only 45.4% of the deteriorating points flagged by the full series, and there was a mean delay of 1.1 years in initial detection. The work presented in this chapter seeks to replicate and extend those results, based on computer-simulated data. This work formed a paper published in the British Journal of Ophthalmology in May 2002 (Gardiner & Crabb 2002a).

3.1 Method

Series of noise-free individual sensitivity values over a period of 6 years were generated using the Virtual Eye described in the previous chapter. These series were either designated as 'Stable', having an age-related deterioration of 0.1dB/year (according to Heijl et al 1991b); or 'Progressing', deteriorating at 2dB/year. Next, at each test date (with these dates being determined by the number of tests being carried out each year, which were assumed to be equally spaced through the year), a noisy sensitivity reading was generated by using the Virtual Eye to randomly sample from a Normal distribution, with a mean value equal to the calculated actual sensitivity at that time. As with all Virtual Eye experiments, we utilise the simplifying assumption that the amount of noise present in each reading is as determined from Equation 2.1 throughout the series. In fact, some patients, and some points within each patient's visual field, will be noisier than others, as explained in Chapter 1.2.1.

The criteria for flagging a point as progressing were defined as being a slope of at least –1dB/year, statistically significant at the 1% level. These criteria have been used in several published studies using PLR (Wild et al 1997; Viswanathan et al 1997c; Membrey et al 2000; Membrey et al 2001). The minimum slope guards against significant age-related decline, and has also been shown to be related to other methods of pointwise change detection (Fitzke et al 1996; McNaught et al 1996). PLR was then carried out on the first 3 readings to see if they would be flagged as progressing or not. This was then repeated for the first 4 readings, then 5 readings and so on. This way it can be seen how the diagnosis (progressing or stable) would change over time as more readings were added to the series. Then, the percentage of correctly diagnosed series (out

of 1000) at each point in time was calculated. This experiment was repeated for different numbers r of readings per year.

3.2 Results

We compare the number of series of points which would be flagged as progressing at each point in time, when simulated as above, for different numbers r of readings per year. The upper graph in Figure 3.1 is for stable points, for which the correct diagnosis is given by the points not being flagged as progressing (hence the lower the percentage being flagged as progressing the better). The lower graph is for points with an actual deterioration of 2dB/year; here the correct diagnosis is given by the points being flagged as progressing (hence the higher the percentage the better).

For example, if a point is actually progressing, and is being tested twice per year, then after three years the point has a 71.8% chance of being flagged as progressing. At the same time, a stable point will be flagged as progressing 5.4% of the time. Therefore if there is a small local defect of say five points that are actually progressing, then in these circumstances, PLR will flag as progressing (according to these criteria) on average three or four of these truly progressing points; it will also flag two or three of the other 49 points elsewhere in the eye which are actually stable.

Clearly, as the number r of tests per year increases, so does the proportion of points flagged as progressing. This confirms the intuitive notion that a progressing eye will be detected quicker with more frequent testing. However, there are also an increased number of early false positives; that is, more stable points are being flagged as progressing. More frequent testing will result in a higher number of incorrect decisions being made for non-deteriorating patients when the follow-up is relatively short. Over the first two years of testing, doubling the frequency of testing also doubles the proportion of false positive results, that is stable points being incorrectly flagged as progressing by PLR. More than three years have passed before the lines converge, and this higher error rate disappears; but by this time the sensitivities are also converging, and so the benefits of carrying out more tests per year have vanished. Under the conditions of this experiment, it is never worthwhile carrying out more than three tests per year; the extra expense and inconvenience to the patient of doing so brings no reward.

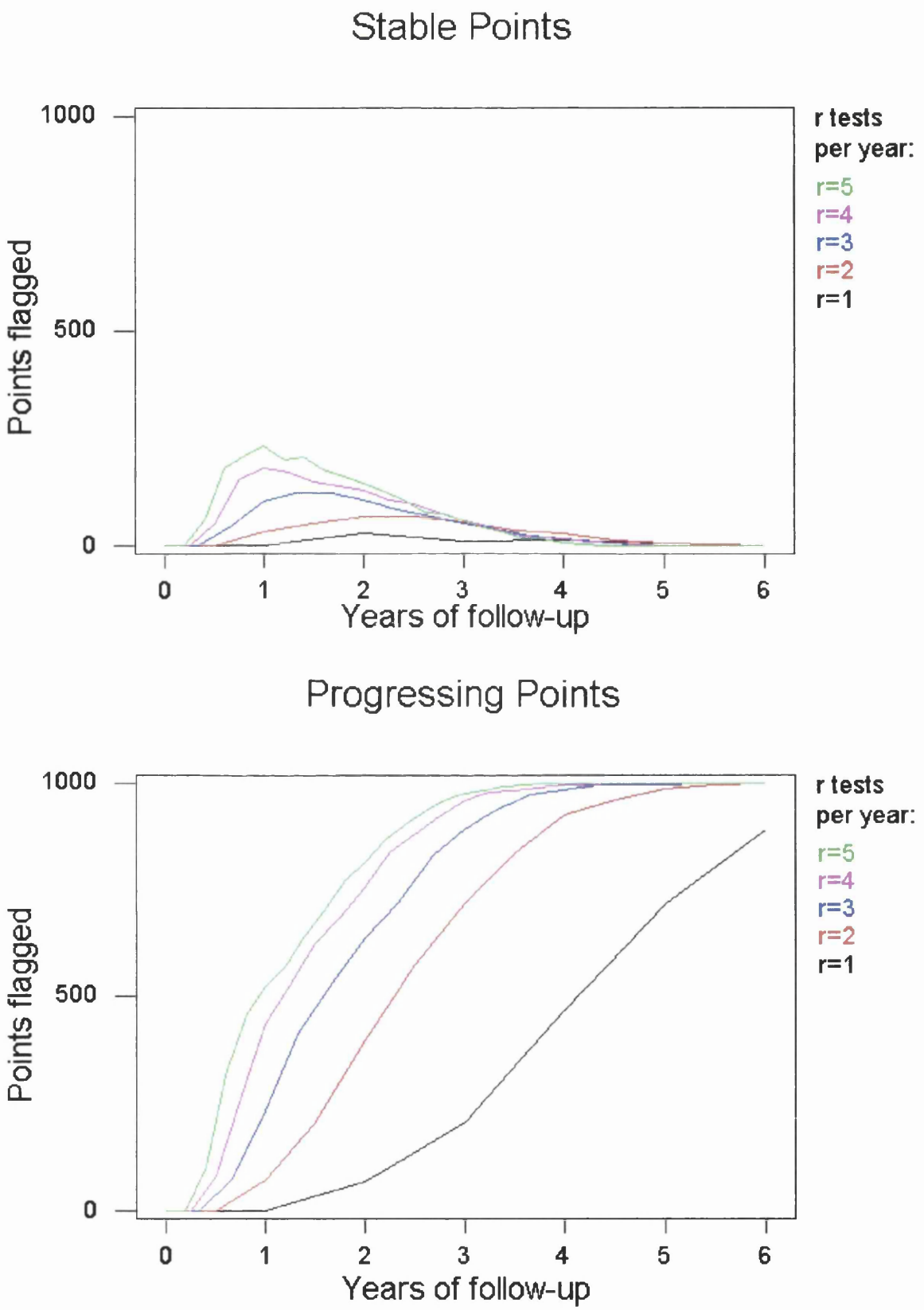


Figure 3.1: The effect of changing the frequency of testing on the proportion of points flagged as progressing, when the noise-free behaviour of the eye is Stable (top) and Progressing (bottom).

3.3 Conclusions

This experiment has shown that although increasing the number of tests per year speeds up the detection of progression (in the lower graph in Figure 3.1), it does so at the expense of, initially at least, falsely flagging far more stable points as progressing (in the upper graph of Figure 3.1); the extra tests are actually making the performance worse. This is because the noise becomes far more significant than the amount of change that may or may not have occurred over the shorter period of time. If the PLR slope based on one test per year for five years is k, then the same sensitivities compressed into just one year (i.e. five tests per year) would give a PLR slope of $5k$, even though the eye may not actually be deteriorating any faster.

The first priority is, in many cases, specificity; because this reduces the chances of incorrect changes in clinical management. If, for example, there is a 3% probability that any given stable point will be incorrectly flagged as progressing (which is the level of false positives seen after as much as three or four years of follow-up in Figure 3.1), then the chance of at least one point out of the 52 test locations in a 24-2 Humphrey visual field (excluding the physiological blind spot) being incorrectly flagged is:

$$\begin{aligned} Prob(\geq 1\ Incorrect\ Point) &= 1 - Prob(A\ Given\ Point\ is\ Incorrect)^{52} \\ &= 1 - 0.97^{52} \\ &= 79.5\% \end{aligned}$$

This alarmingly poor performance is improved dramatically by a seemingly small improvement in the specificity at one point, as seen in Table 3.1:

Prob(Given Point is FP)	*Prob(≥1 Point is FP)*
3%	79.5%
2%	65.0%
1%	40.7%

Table 3.1: The effect of changes in the specificity of a single point on the specificity of the entire visual field.

Hence we have shown that having an increased number of tests per year reduces specificity when the follow up is relatively short. Although the graphs in Figure 3.1 do converge as the series lengthens, a clinical management decision based on a significant slope would be unwise with frequent testing in say the first two or three years. One strategy is to use one or more confirmation fields which may or may not be part of the actual follow up; this is addressed in the experiment described in Chapter 4.

The second clinical priority is to improve the sensitivity of the test; in other words detecting true visual field progression. From the lower graph in Figure 3.1 it is clear that the detection of progressing points with just one test carried out per year is extremely poor given the level of noise used in the simulation. Even with two tests per year, points deteriorating at 2dB/year would only be picked up on in 75% of cases after three years; by this time a loss of 6dB has occurred, which is severe enough that any decent method should have identified it. Clinically, this means that a substantial fraction of visual field locations have to

be truly deteriorating before progression can be reliably detected. Therefore, approximately three tests per year, resources permitting, seems to achieve a better success rate at determining which points are progressing and which are not. This supports and extends the findings on patient data in glaucoma from Viswanathan et al (1991c).

The Virtual Eye model provides a much-simplified version of reality. The assumption of noise being normally distributed, though commonly used, is unproven, and the amount of noise present in readings may well be larger than the estimates used here (Flammer et al 1984; Heijl et al 1989; Chauhan & Johnson 1999b; Hutchings et al 2000; Spry et al 2001). The question of estimating the distribution and magnitude of the noise will be examined in Chapter 7. As the amount of noise increases, the performance of PLR becomes even worse than in Figure 3.1. Several years' worth of follow-up is needed before satisfactory results can be obtained. Clinically, it is common practice to use confirmation fields to look for points that are persistently progressing (Hitchings et al 1994; Leske et al 1999; Membrey et al 2000; Membrey et al 2001; Westcott et al 2001; Heijl et al 2002; Kim et al 2003); this issue will be addressed further in Chapter 4. The uses of confirmation fields, and different levels of noise, were tried using the Virtual Eye and although the values for specificity and sensitivity altered, the qualitative results (most importantly the recommendations on ideal frequency of testing) were unchanged. For clarity, the simplest case, without confirmation fields, has been described in this chapter.

It is also common practice to look for clusters of points that are all progressing, rather than individual points. While this is beyond the scope of this model (because the spatial pattern of deterioration is not clearly known, and varies according to the location in the eye), it is still desirable to have an accurate determination of progression for each of the points in the cluster. The findings, in the present form, may have limited clinical generalisability; but it is believed that the conclusions drawn from them are more widely applicable. Non-glaucomatous change (e.g. effect of concomitant cataract) is currently indistinguishable from glaucomatous change using PLR and so cannot be included in the Virtual Eye simulation. Nevertheless use of simulation enables results to be found which would be extremely hard to achieve using patient data, because for a real-world patient, the underlying noise-free state of the eye is, as yet, unknown.

It is clear that the current PLR methodologies still need improvement. Methods are needed which are more sensitive without compromising specificity; this way, accurate results and faster diagnoses could be obtained with just one or two tests per year over shorter periods of time than at present. This relies primarily on reducing the amount of noise present in the readings, something which will be addressed later in this thesis in Chapters 5 and 6.

This experiment gives rise to the conclusion that three tests per year is a good compromise between sensitivity and specificity. There are certainly large benefits to be had in sensitivity when compared with carrying out just one test per year.

4. Confirmation Fields Methods

There is currently no agreement over which of the several different PLR criteria are the best. Indeed, the method that is most suitable for use will depend on the purpose to which the results are to be put. When the outcome of an eye being labelled as progressing would be the patient undergoing a risky or costly change in clinical management, a method should be used which is known to falsely label very few stable eyes as progressing (i.e. a high degree of specificity). Conversely, when failing to treat a progressing eye would be much more damaging than treating a stable eye (perhaps because the patient is already blind in the other eye, for example), it is better to use a test which will correctly flag progressing eyes quicker (i.e. a more sensitive test). Clearly, the key is which way the method flags eyes in the middle-ground, which could be judged to be either progressing or stable; normally tests which are more sensitive are less specific, and vice-versa. There is no consensus on what value of regression slope and p-value constitutes progression; and whether it should be maintained in subsequent fields. Indeed the latter idea of 'confirmation fields' or 'confirmation criteria' has been shown to improve the specificity of other methods for detecting visual field progression (Schulzer et al 1994; Chauhan et al 1999a) but has yet to be formally examined for PLR; yet such criteria are still used 'ad hoc' to demonstrate the 'benefits' of treatment changes in glaucoma (Membrey et al 2001). In this chapter, a selection of seven different methods for using confirmation fields to decide on the diagnosis in PLR are evaluated both theoretically and using the Virtual Eye. Two of these methods are new and have not been examined before. This work formed a paper published in Investigative Ophthalmology & Visual Science in May 2002 (Gardiner & Crabb 2002b).

4.1 The PLR Methods Examined

Seven different PLR methods of determining whether a point in an eye is progressing or not, given the sensitivities at n time points (equally-spaced visual field tests in time), will be compared; the aim being to rank them according to their specificity and sensitivity. We define these as:

1. *Standard Criteria*: a point is flagged as progressing if it shows a significantly negative slope at the 1% level, together with an observed slope of at least -1dB/year in the sensitivity; this will be written as Z_n=1. The significance level is calculated by comparing the slope to the t-distribution, with (n-1) degrees of freedom. This is the simplest commonly used PLR method used in published studies (Birch et al 1995a; McNaught et al 1996; Smith et al 1996; Katz et al 1997; Nouri-Mahdavi et al 1997; Viswanathan et al 1997a; Viswanathan et al 1997c; Wild et al 1997; Viswanathan et al 1999; Spry et al 2000; Hofmann et al 2001) and is used clinically as an indicator for change in some centres (Kamal & Hitchings 1998).

2. *Two out of Two*: a point is flagged as progressing if it satisfies the standard criteria, and continues to satisfy them on addition of a further observed (confirmation) point; i.e. Z_n=1 and Z_{n+1}=1. This was used by Hitchings et al (1994).

3. *Three out of Three*: a point is flagged if it satisfies the standard criteria, and continues to satisfy them after the addition of each of a further two observed (confirmation) points in turn; i.e. Z_n=1, Z_{n+1}=1 and Z_{n+2}=1. This approach, applied to a form of Glaucoma Change Probability analysis, has being used in a clinical trial to evaluate the role of immediate intra-ocular pressure reduction in glaucoma (Leske et al 1999; Heijl et al 2002), and has been used in a study to determine the prevalence of glaucomatous progressive functional damage (Kim et al 2003).

4. *Two out of Three*: a point is flagged if it satisfies the standard criteria, and continues to do so on the addition of either the following one point or two points; i.e. Z_n=1 and either Z_{n+1}=1 or Z_{n+2}=1 (or all three). This has recently been used in a study comparing PLR and the AGIS visual field score (Westcott et al 2001).

5. *Three out of Four*: a point is flagged if it satisfies the standard criteria, and continues to do so with the addition of two of three successive points to the series; i.e. Z_n=1 and at least two of Z_{n+1}=1, Z_{n+2}=1 and Z_{n+3}=1. This method has been used in studies examining the treatment of normal tension glaucoma (Membrey et al 2000; Membrey et al 2001).

And our two new proposed methods:

6. *Two Omitting*: a point is flagged if it satisfies the standard criteria, and the slope obtained by adding one confirmation point but excluding point n also satisfies the criteria; this will be written as Z_n=1 and Z'_{n+1}=1.

7. *Three Omitting*: a point is flagged if it satisfies the standard criteria, and the two slopes obtained by using points 1 to (n-1) and either point (n+1) or point (n+2) both satisfy the criteria; i.e. Z_n=1, Z'_{n+1}=1 (when point n is omitted) and Z'_{n+2}=1 (when points n and (n+1) are both omitted).

The logic behind these new methods is that if the latest sensitivity reading in the series is significantly worse than preceding readings purely by chance because of the noise, then a confirmatory regression, carried out once the following point has been added to the series, will be biased by this low sensitivity reading; and so omitting the point will give a more conservative estimate of the true rate of progression. As seen in Figure 4.1, when the point is actually stable, there will be less chance of it incorrectly being flagged as progressing, i.e. fewer false positives.

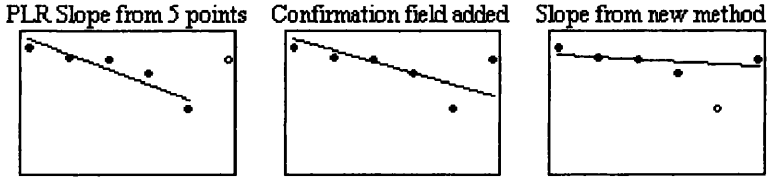


Figure 4.1: The effect of omitting a point from the regression on the PLR slope.

4.2 Theoretical Comparison

4.2.1 Assumptions

Several simplifying assumptions need to be made to make the theoretical problems tractable. Firstly, the noise is assumed to be normal (as with the Virtual Eye simulation); this is a reasonably accurate assumption but may be less true when the sensitivity is at the extremes of the range of measurement values because of the censoring nature of the instrument. In other words a point with a sensitivity of say 5dB will have positively skewed noise, since no readings below 0dB are possible; whilst a stimulus seen at dim levels such as 36dB will have negatively skewed noise. The mean for this normal distribution is estimated based on the readings taken so far. Secondly, it is assumed that the visual fields for each patient are taken at regularly spaced intervals. Both of these assumptions are in keeping with published work (Spry et al 2000).

Since it is known that a more specific method of determining progression will always be less sensitive if the level of noise is unchanged (methods of reducing the amount of noise are a separate issue entirely), it is only necessary to look at the specificity of each method; that is, how many stable eyes will be incorrectly flagged as progressing.

4.2.2 Derivation

Suppose we have a series of visual fields, taken at times t_1, t_2... t_n. At a given point in the field, the deviations from the age-corrected normal sensitivities are X_1, X_2... X_n. Rather than the raw values, it is chosen to work with the deviations X_i below the age-corrected normal sensitivities; if the expected sensitivity at that point for a patient of this age is S_E, then $X_i = S_E - S_i$. This means that at a stable point, the true (noise-free) deviation will remain constant; and so the X_i will all have the same distribution. If the point is progressing, the X_i will increase over time.

Now, according to our first assumption described above, we can say that for a stable eye, the X_i will be normally distributed about a constant c; $X_i \sim N(c, \sigma^2)$. Note that the assumption is that the actual deviation of the point is constant, but we do not assume that the first or last readings are exactly equal (unlike in Spry et al 2000), because those two readings will also have noise, and so the readings will not be the same as the actual deviation in the eye. The constant c is estimated by $\hat{c}$, the mean of the readings taken so far; so that the sum of the first $(n\text{-}1)$ readings will be $(n\text{-}1)\hat{c}$. But the PLR slope will be unaffected by the addition of a constant to each reading X_i; and so if we define $Y_i = X_i - \hat{c}$, we can carry out PLR on the Y_i to produce the same results; then $Y_i \sim N(0, \sigma^2)$ and the sum $Y_1 + Y_2 + \ldots + Y_{n-1} = 0$. So, the Y_i are pure noise; as if there were no noise, Y_i would equal 0 for each $i = 1 \ldots n\text{-}1$. We are also assuming (as mentioned above) that $t_i = i$; i.e. the readings have been taken at equally spaced time intervals.

Now, let β_n be the PLR slope based on the first n readings in the series, and using our new *Two Omitting* method the new PLR slope is θ_{n+1}. Now for a pointwise linear least-squares regression, the slope is calculated by the equation

$$\left(\Sigma t_i^2 - \frac{(\Sigma t_i)^2}{n}\right)\beta_n = \Sigma t_i X_i - \frac{\Sigma t_i}{n}\Sigma X_i$$

Equation 4.1

where the sums are taken over 1 to n.

Because $t_i=i$ in this case, and using the substitution $Y_i=X_i-c$, we can simplify this to

$$\frac{1}{12}n(n+1)(n-1)\beta_n = \Sigma iY_i - \frac{n+1}{2}\Sigma Y_i$$

Equation 4.2

Next, using our new *Two Omitting* method, we can calculate the new PLR slope θ_{n+1} (based on the values at points 1, 2,… (n-1) and (n+1)) from Equation 4.1, giving:

$$\frac{1}{12n}(n^2-1)(n^2+12)\theta_{n+1} = \sum_{i=1}^{n-1} iY_i + (n+1)Y_{n+1} - \left(\frac{n+1}{2}+\frac{1}{n}\right)\left(\sum_{i=1}^{n-1} Y_i + Y_{n+1}\right)$$

Equation 4.3

We are interested in the case where X_n has been added to the series, and has made the PLR slope significant; and the clinician wishes to carry out a confirmation test. So we know that $\beta_n > \beta_{n-1}$; or equivalently, from Equation 4.2,

$$\frac{(n-2)(n-1)}{2}Y_n > 3\sum_{i=1}^{n-1} iY_i$$

Equation 4.4

Now from our definition of Y_i, given the first (n-1) readings, we know that the expected values $E(Y_n)=$ $E(nY_n)= E(Y_{n+1})= E((n+1)Y_{n+1})=0$. Remembering that $\sum_{i=1}^{n-1} Y_i = 0$, we see from Equation 4.2 and 4.3 that when $\sum_{i=1}^{n-1} iY_i \geq 0$;

$$
\begin{aligned}
\mathrm{E}(\beta_{n+1} \text{ given } Y_1,\ldots Y_n) &= \frac{12}{n(n+1)(n+2)}\left(\sum_{i=1}^{n} iY_i - \frac{n+2}{2}\sum_{i=1}^{n} Y_i\right) \\
&= \frac{12}{n(n+1)(n+2)}\left(\sum_{i=1}^{n-1} iY_i - \frac{n+2}{2}\sum_{i=1}^{n-1} Y_i + \left(n - \frac{n+2}{2}\right)Y_n\right) \\
&= \frac{12}{n(n^2-1)}\left(\frac{n-1}{n+2}\sum_{i=1}^{n-1} iY_i + \frac{(n-2)(n-1)}{2(n+2)}Y_n\right) \\
&> \frac{12}{n(n^2-1)}\left(\frac{n-1}{n+2}\sum_{i=1}^{n-1} iY_i + \frac{3}{n+2}\sum_{i=1}^{n-1} iY_i\right) \\
&= \frac{12n}{n^2(n^2-1)}\sum_{i=1}^{n-1} iY_i \\
&\geq \frac{12n}{(n^2+12)(n^2-1)}\sum_{i=1}^{n-1} iY_i \\
&= \mathrm{E}(\theta_{n+1} \mid Y_1,\ldots Y_n)
\end{aligned}
$$

Equation 4.5

Also, since $\beta_n>0$ (since the PLR slope satisfies the progression criteria), if $\sum_{i=1}^{n-1} iY_i < 0$ then from Equation 4.2, $\frac{n-1}{2}Y_n > -\sum_{i=1}^{n-1} iY_i$; and so in this case,

$$
\begin{aligned}
\mathrm{E}(\beta_{n+1} \text{ given } Y_1,\ldots Y_n) &= \frac{12}{n(n+1)(n+2)}\left(\sum_{i=1}^{n} iY_i - \frac{n+2}{2}\sum_{i=1}^{n} Y_i\right) \\
&= \frac{12}{n(n+1)(n+2)}\left(\sum_{i=1}^{n-1} iY_i - \frac{n+2}{2}\sum_{i=1}^{n-1} Y_i + \left(n - \frac{n+2}{2}\right)Y_n\right) \\
&= \frac{12}{n(n^2-1)(n+2)}\left((n-1)\sum_{i=1}^{n-1} iY_i + \frac{(n-2)(n-1)}{2}Y_n\right) \\
&> \frac{12}{n(n^2-1)(n+2)}\left((n-1)\sum_{i=1}^{n-1} iY_i - (n-2)\sum_{i=1}^{n-1} iY_i\right) \\
&= \frac{12n}{n^2(n^2-1)(n+2)}\sum_{i=1}^{n-1} iY_i \\
&\geq \frac{12n}{(n^2+12)(n^2-1)}\sum_{i=1}^{n-1} iY_i \\
&= \mathrm{E}(\theta_{n+1} \mid Y_1,\ldots Y_n)
\end{aligned}
$$

Equation 4.6

where the final inequality is true whenever $n\geq 2$.

Therefore in either case, the *Two Omitting* method for PLR would be expected to give a slope closer to the actual slope of zero for a stable eye than the more standard *Two out of Two* method.

Now, the advantage to assuming the Y_i to be Normally distributed is that any linear combination of them will also be Normally distributed; in particular, β_{n+1} and θ_{n+1} both are. The variance of β_{n+1} is:

$$\begin{aligned}
Var(\beta_{n+1}) &= Var\left(\frac{12}{n(n+1)(n+2)}\left(\sum_{i=1}^{n+1} iY_i - \frac{n+2}{2}\sum_{i=1}^{n+1} Y_i\right)\right) \\
&= \left(\frac{12}{n(n+1)(n+2)}\right)^2 \left(\sum_{i=1}^{n+1} i^2\sigma^2 - \left(\frac{n+2}{2}\right)^2 \sum_{i=1}^{n+1}\sigma^2\right) \\
&= \left(\frac{12}{n(n+1)(n+2)}\right)^2 \left(\frac{(n+1)(n+2)(2n+3)}{6} - \frac{(n+2)^2(n+1)}{4}\right)\sigma^2 \\
&= \left(\frac{12}{n(n+1)(n+2)}\right)^2 (n+1)(n+2)\left(\frac{2n+3}{6} - \frac{n+2}{4}\right)\sigma^2 \\
&= \frac{12\sigma^2}{n(n+1)(n+2)}
\end{aligned}$$

Equation 4.7

where σ^2 is the variance of each Y_i (assumed earlier to be constant). Similarly,

$Var(\theta_{n+1}) = \dfrac{12n\sigma^2}{(n^2+12)(n^2-1)}$. And so, writing $se(X) = \sqrt{Var(X)}$ for the standard error of X,

Equations 4.2, 4.3 and 4.7 show that for testing the significance of the PLR slopes, when $\sum_{i=1}^{n-1} iY_i \geq 0$;

$$\begin{aligned}
\mathrm{E}\left(\frac{\theta_{n+1}}{se(\theta_{n+1})}\middle| Y_1,\ldots Y_n\right) &= \sqrt{\frac{12n}{(n^2+12)(n^2-1)\sigma^2}}\left(\sum_{i=1}^{n-1} iY_i\right) \\
&= \sqrt{\frac{12}{n(n+1)(n+2)\sigma^2}}\sqrt{\frac{n^2(n+2)}{(n^2+12)(n-1)}}\left(\sum_{i=1}^{n-1} iY_i\right) \\
&= \sqrt{\frac{12}{n(n+1)(n+2)\sigma^2}}\sqrt{\frac{n^2(n-1)}{(n^2+12)(n+2)}}\left(\frac{n+2}{n-1}\sum_{i=1}^{n-1} iY_i\right) \\
&\leq \sqrt{\frac{12}{n(n+1)(n+2)\sigma^2}}\left(\frac{n+2}{n-1}\sum_{i=1}^{n-1} iY_i\right) \\
&= \sqrt{\frac{12}{n(n+1)(n+2)\sigma^2}}\left(\sum_{i=1}^{n-1} iY_i + \frac{3}{n-1}\sum_{i=1}^{n-1} iY_i\right) \\
&< \sqrt{\frac{12}{n(n+1)(n+2)\sigma^2}}\left(\sum_{i=1}^{n-1} iY_i + \frac{n-2}{2}Y_n\right) \\
&= \sqrt{\frac{12}{n(n+1)(n+2)\sigma^2}}\left(\sum_{i=1}^{n} iY_i - \frac{n+2}{2}\sum_{i=1}^{n} Y_i\right) \\
&= \mathrm{E}\left(\frac{\beta_{n+1}}{se(\beta_{n+1})}\middle| Y_1,\ldots Y_n\right)
\end{aligned}$$

Equation 4.8

where the last inequality follows from Equation 4.4. Similarly, when $\sum_{i=1}^{n-1} iY_i < 0$;

$$
\begin{aligned}
\mathrm{E}\left(\frac{\theta_{n+1}}{se(\theta_{n+1})}\middle|Y_1,\ldots Y_n\right) &= \sqrt{\frac{12n}{(n^2+12)(n^2-1)\sigma^2}}\left(\sum_{i=1}^{n-1} iY_i\right) \\
&= \sqrt{\frac{12}{n(n+1)(n+2)\sigma^2}}\sqrt{\frac{n^2(n+2)(n-1)}{(n^2+12)}}\left(\frac{1}{n-1}\sum_{i=1}^{n-1} iY_i\right) \\
&< \sqrt{\frac{12}{n(n+1)(n+2)\sigma^2}}\left(\frac{1}{n-1}\sum_{i=1}^{n-1} iY_i\right) \\
&= \sqrt{\frac{12}{n(n+1)(n+2)\sigma^2}}\left(\sum_{i=1}^{n-1} iY_i - \frac{n-2}{n-1}\sum_{i=1}^{n-1} iY_i\right) \\
&< \sqrt{\frac{12}{n(n+1)(n+2)\sigma^2}}\left(\sum_{i=1}^{n-1} iY_i + \frac{n-2}{2}Y_n\right) \\
&= \mathrm{E}\left(\frac{\beta_{n+1}}{se(\beta_{n+1})}\middle|Y_1,\ldots Y_n\right)
\end{aligned}
$$

Equation 4.9

with the last inequality following from Equation 4.2 as before.

So, the slope θ_{n+1} would be expected to be less steep (from Equations 4.5 and 4.6) than β_{n+1}; and also less significant (from Equations 4.8 and 4.9) (if $\theta_{n+1}>0$; if $\theta_{n+1}<0$ then the slope will not be flagged as progressing anyway); showing that our new *Two Omitting* method is more specific than the current *Two out of Two* method.

Naturally, it is straightforward to extend these two results to prove that the *Three Omitting* PLR method is more specific than the *Three out of Three* method.

It is obvious (and easily provable) that *Three out of Four* will flag more points as progressing than *Three out of Three*; since any series which satisfies the criteria for the latter method necessarily satisfies the criteria for the former method. So we can sum up all the information in Figure 4.2; where when a line connects two methods, it means that the upper method has been shown from theoretical results to be more specific than the lower method.

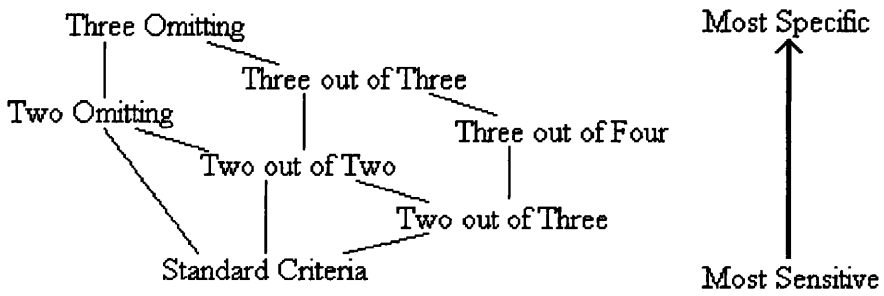


Figure 4.2: Theoretical relative specificity and sensitivity of the different PLR methods.

4.3 Virtual Eye Simulation Experiment

4.3.1 Method

We simulate sets of readings for a whole visual field in two circumstances:

Stable Eye:

A Humphrey 30-2 visual field for a 'Virtual Eye' was constructed and is shown in greyscale form at the top of Figure 4.3. This left eye has an early to moderate inferior arcuate scotoma and some general depression of sensitivity in the superior field. This field has an abnormal Humphrey mean deviation of approximately –4dB. The individual pointwise sensitivities that make up this field are assumed to be the 'true' physiological values for this eye. These will not generally be identically the same as the measured or estimated values derived at a visual field test. Three typical examples of the latter, where noise has been added to the true visual field, are shown as greyscales at the bottom of Figure 4.3.

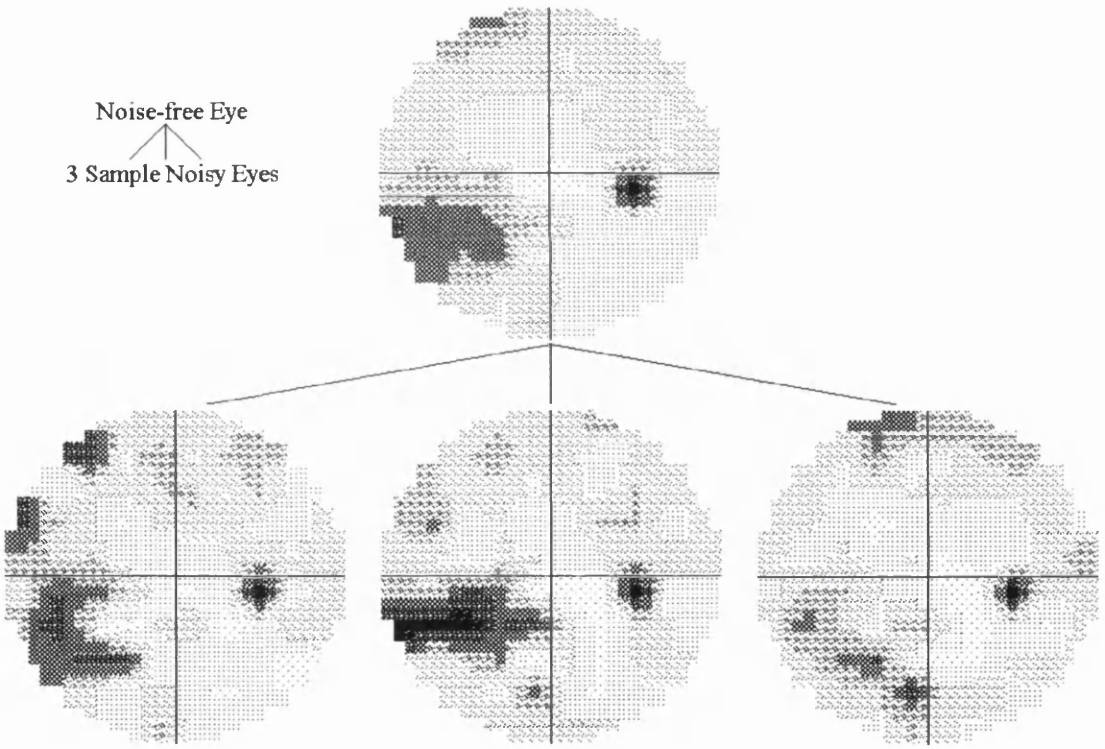


Figure 4.3: The effect of adding noise to the actual input sensitivity values for each point in a Virtual Eye. Visual fields are represented as 30-2 Humphrey greyscales with darker areas indicating lower sensitivity readings at that point. The three noisy eyes generated are noticeable different both from each other, and also from the true physiological values represented by the top greyscale.

One extra feature of the stable Virtual Eye used for this experiment is that a normal age related decline of 0.1dB per year was subtracted from the true value in all cases. This means that if the initial true sensitivity is 28dB, then the measured sensitivity at that point recorded five years after the first test is given by the simulated value drawn randomly from a Normal distribution with mean 27.5dB and standard deviation of

2.84dB. The standard deviation is calculated from Equation 2.1 in Chapter 2; imitating the established fact that variability is dynamic, increasing as the sensitivity declines.

Deteriorating Eye:

The stable Virtual Eye described above forms the basis and starting point for the deteriorating eye. A progressive visual field defect is then added to sites in and around the initial inferior arcuate defect. More precisely this means that six points (with starting sensitivities of 32, 28, 24, 20, 16 and 12 dB) in the visual field are each given a rate of loss of 2dB per year. The consequence of this magnitude of progression on the noise-free eye over a six-year period is illustrated in Figure 4.4 (naturally, the simulation subsequently adds noise to each point to generate visual field series). This is the same eye shown earlier in Figure 2.4.

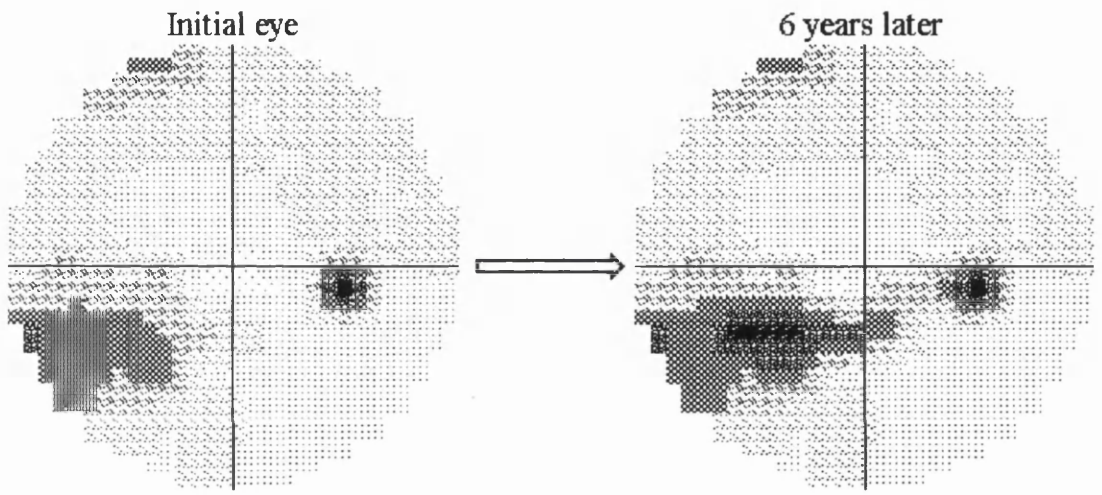


Figure 4.4: The progression in our noise-free eye, as used in the Virtual Eye simulation.

The visual field at six years would have a Humphrey MD of approximately –5 dB, as compared to -4dB initially. This level of deterioration could be considered clinically moderate and with noise added in, it is doubtful that progression would be diagnosed by visual inspection or monitoring changes in global indices. Nevertheless, the cluster of points are decaying at more than 10 to 20 times age related values (Heijl et al 1991b; Spry & Johnson 2001), and the threat of the progressing defect to the functionally important area of fixation means that it is precisely the type of progression that we suggest any decent pointwise method should detect with reasonable sensitivity and specificity.

As with the stable eye, the important feature of this type of simulation is the noise and this is added to points in the visual field as described in Chapter 2. For example, the progressing point at (–3°, –3°) on the Humphrey visual field starts with a true value of 32 dB. The measurement at this point is imitated by the simulated value drawn randomly from a Normal distribution with mean 32 dB and standard deviation of 1.97dB. After a period of follow-up of three years, the true value at this point is 26 dB and the simulated measurement is given by the value drawn randomly from a Normal distribution with mean 26 dB and standard deviation of 3.20dB.Visual field series for this deteriorating eye can be generated in a similar fashion to those for the stable eye.

One thousand series from the stable eye and one thousand series from the progressing eye, all with two tests per year over a six year follow up period (thirteen fields in total for each series), were generated. Then PLR was applied to each of the 74 non-blind spot locations within each 30-2 visual field series, sequentially for each test starting from the fourth test (i.e. using the first one and a half years of readings from the series) and finishing at the thirteenth test (using all six years of readings). Then, each criterion for determining progression was used in turn, and the eye labelled as progressing if one point in the field satisfied that particular criteria. The test at which the eye was first detected as progressing was recorded. For those methods using at least two fields (one confirmation field), it means that the fifth field was the earliest that progression can be flagged; and for those using two confirmation fields it means the sixth test was the earliest possible. Also note that if, for example, the *Two out of Two* criterion is satisfied by using the ninth and tenth fields, then the visual field series was only labelled as progressing at the tenth field.

The SPlus program, incorporating the Virtual Eye, which was written to carry out this experiment is given in Appendix 1.

4.3.2 Results

The graphs in Figures 4.5 and 4.6 show the cumulative percentage of visual field series of the Virtual Eye (out of one thousand) that were flagged as progressing at, or before, each point in time using the specified methods. In each graph, the uppermost line is flagging most eyes as progressing, whether this is incorrect (for a stable eye; this method is then the least specific) or correct (for a deteriorating eye; this method is the most sensitive). For example, Figures 4.5 and 4.6 show that after three years of follow up the *Standard Criteria* had already falsely labelled 74.6% of the stable visual field series as progressing. After six years virtually all of the stable eyes had been falsely labelled as progressing. In contrast the *Three Omitting* method had falsely labelled just 11.6% of the stable eyes after six years. The sensitivity (proportion of correctly labelled progressing eyes) of the *Standard Criteria* and the *Three Omitting* methods after six years were 97.1% and 65.7% respectively (a significant but much smaller difference).

Figure 4.5 shows that out of the three methods which use one confirmation fields (*Two out of Two, Two out of Three* and *Two Omitting*), the *Two Omitting* method gives a significant improvement in specificity, while the accompanying loss in sensitivity is much less marked. Figure 4.6 shows that, similarly, out of the three methods which use two confirmation fields (*Three out of Three, Three out of Four* and *Three Omitting*), the *Three Omitting* method gives a significant improvement in specificity, while the accompanying loss in sensitivity is much less marked. So it is clear that, as predicted theoretically, our new *Two Omitting* and *Three Omitting* methods are the most specific out of those using one and two confirmation fields respectively; the currently-used methods can incorrectly flag twice as many stable eyes as being progressing. These two new methods improve the specificity of PLR, without having a severe negative effect on the sensitivity.

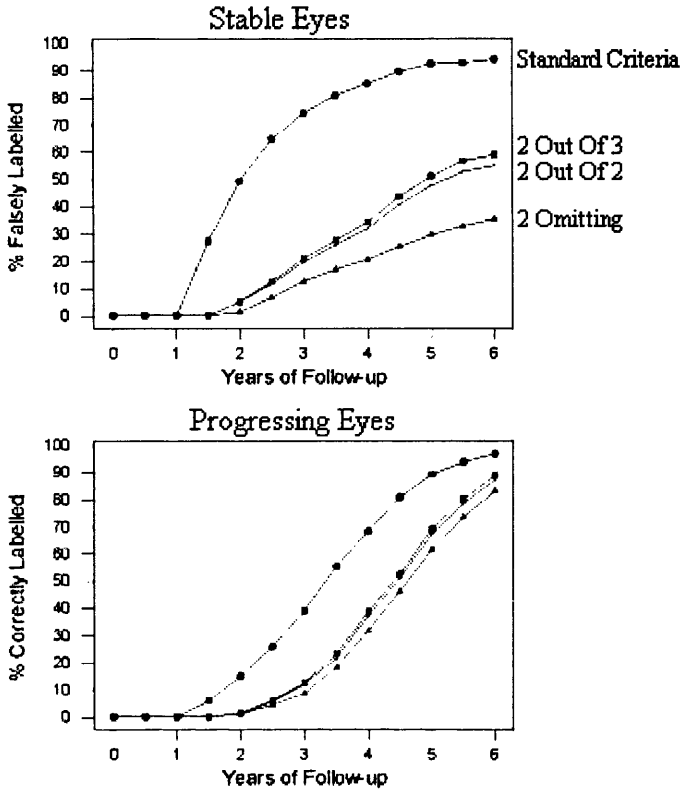


Figure 4.5: The relative specificity and sensitivity (respectively) of four different PLR methods, as determined by the Virtual Eye simulation.

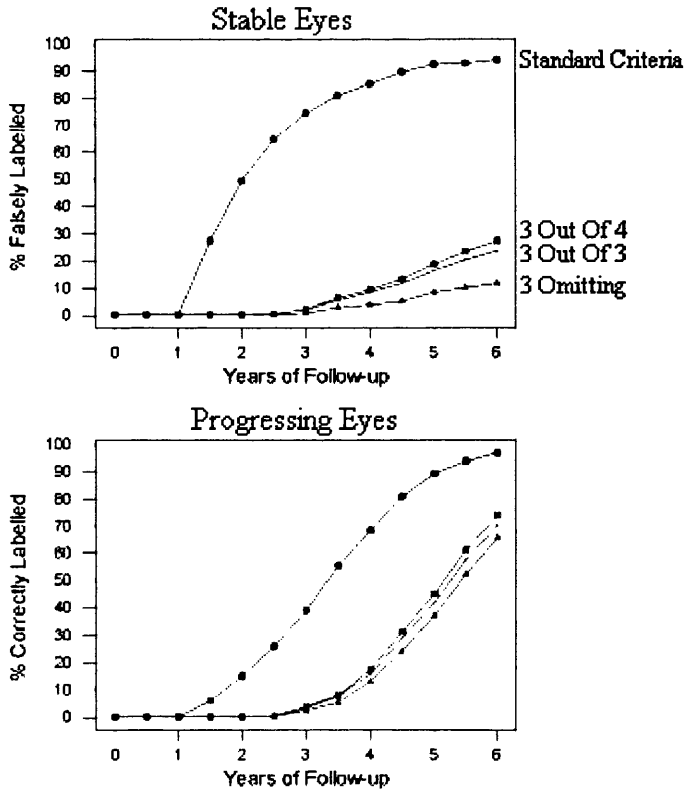


Figure 4.6: The relative specificity and sensitivity (respectively) of the remaining three different PLR methods, plus the Standard Criteria method again for reference, as determined by the Virtual Eye simulation.

4.4 Conclusions

When choosing a method for determining whether a visual field is progressing or stable, the specificity and sensitivity both have to be taken into account. PLR has previously been shown to be a sensitive technique for detecting visual field progression in patients with glaucoma (Birch et al 1995; Smith et al 1996; Katz et al 1997; Viswanathan et al 1997a; Nouri-Mahdavi et al 1997). Confirmation tests (or fields) require more patient examinations but afford better specificity. Out of the seven PLR methods we have examined in this chapter, the *Standard Criteria* method was the most sensitive (it successfully detected progressing eyes quicker); whereas the *Three Omitting* method was the most specific (it wrongly identified fewest stable eyes as being progressing). This work has also highlighted that fulfilling the *Standard Criteria* for PLR at just one point in the visual field without any confirmation criteria is clinically unreliable as a means of diagnosing progression, because of its absurd levels of specificity; and as discussed in Section 3.3, specificity is crucial to any good method.

Statistically speaking, there is very little justification for classifying a point within a visual field as either stable or progressing; any cut-off point will be in essence arbitrary. The rationale behind such a distinction is that if one or more points are progressing, then treatment is needed. However, a clinical decision of whether to treat will be based on much more complex factors than whether the slope satisfies one strict criteria; for example the history of the patient and their family, and the condition of the other eye. Although we have chosen, in accordance with published studies, to use a significant slope of -1dB/year to indicate progression, this figure is still fairly arbitrary; as is describing a typical progressing point as being one with a deterioration of -2dB/year. Any software used clinically for the analysis of visual field data should allow the user to alter the level of slope required for a point to be flagged, according to their own clinical judgment based on other factors; this can be done with some software, for example Progressor (Fitzke et al 1996). Also, when comparing two treatments, for example, it may be better to compare the distribution of slopes of points, rather than a perfunctory comparison of the proportion of slopes that satisfy such criteria.

This study has been confined to a consideration of different PLR methods. No comparison is made against non-linear or non-pointwise methods. There is disagreement over whether or not PLR really is the best method of determining progression, but it is a commonly examined method (Hitchings et al 1994; McNaught et al 1995; Smith et al 1996; Katz et al 1996; Fitzke et al 1996; McNaught et al 1996; Viswanathan et al 1997a; Viswanathan et al 1997c; Nouri-Mahdavi et al 1997; Bhandari et al 1997; Wild et al 1997; Kamal & Hitchings 1998; Viswanathan et al 1999; Membrey et al 2000; Spry et al 2000; Membrey et al 2001; Hoffman et al 2001), and as such any potential refinements to the method should be considered. Moreover, PLR may be suitable for examining any deepening of an existing defect; but its effectiveness at examining enlargement of defects is more open to question. Potential disadvantages of PLR are not addressed by the current study. However, the conclusion that omitting techniques may be of benefit to regression methods has wider applications than the limited set of conditions used here. For example, different levels of noise, different defect sizes and different rates of progression would all affect the quantitative results in Figures 4.5 and 4.6; but they would not affect the qualitative comparisons

between the methods. The simulation should be viewed as an example, which supports the theoretical comparisons between the methods.

Another way of evaluating the new method would be to use real patient data. However, when the actual state of the eye is unknown (as in the case of real patient data), there is currently no gold standard against which to judge different techniques. A further advantage of the techniques used here is that they are not limited to glaucomatous eyes; and so the methods developed and the principles behind confirmation techniques could be applied in other situations in related fields. Nevertheless, before recommending any methods for widespread use, their usefulness and practicality would need to be tested in a clinical situation.

In such a clinical situation, the interpretation of any method will remain a subjective matter; but using the approach described in this chapter provides decent estimates of sensitivity and specificity for particular methods, arming the clinician to make more informed patient management decisions than with other PLR methods. It is very clear from these results how much specificity is gained by adopting a confirmation approach. We have demonstrated that (when the criteria of -1dB/year and a 1% significance level are fixed) any gain in specificity when the PLR method is changed will be accompanied by a loss of sensitivity. But because of the disproportionate effect of a small change in specificity at each point, as demonstrated in our results, and because progressing eyes may often form a small proportion of the study population, seeking out specific methods is generally more important than sensitivity. For this reason, the new *Three Omitting* method is suggested as an alternative to current confirmation methods applied to PLR, and may be a suitable option when PLR methods are used in clinical trials.

5. Derivation Of Filter

Reliable detection of glaucomatous visual field defects, and follow-up of current defects to determine whether or not they are spreading or deepening, is crucial for the correct management of patients. Yet this is extremely difficult given the inaccuracy of threshold perimetry (in terms of the high inter-test and intra-test variability) and the various components of variability (or noise) associated with the perimetric process (Flammer et al 1984; Chauhan & House 1991; Wild et al 1991; Spenceley & Henson 1996; Spry & Johnson 2002), as discussed in Section 1.2.1. Therefore, it is sought to find ways of reducing the noise present in the readings. Reducing this noise is particularly relevant when detecting visual field progression (Spry & Johnson 2002); this is because true change in a glaucomatous visual field has to be larger than the noise before it becomes statistically distinguishable (Heijl et al 1989; Chauhan & Johnson 1999b). Improved methods of data acquisition such as SITA (Bengtsson et al 1997) and SITA Fast (Bengtsson & Heijl 1998) have tended to focus on reducing the time of perimetric examination rather than making the measurement more accurate (Wild et al 1999b; Artes et al 2002). Carrying out more readings during the testing procedure is also clearly undesirable. Far better would be to process the data in such a way that the noise would be reduced, without any additional testing time. One way of doing this relies on exploiting the relationships between the actual sensitivities of different points; in essence, if one point has a reduced sensitivity, then its neighbours are more likely to also have reduced sensitivities. This principle points towards spatial filtering of the data as a possible solution.

The work in this chapter and in Chapter 6 formed a paper published in Vision Research (Gardiner et al 2003). It was presented by the author at the Research Students' Conference in Probability and Statistics at Warwick University, UK from 18-21 March 2002; at ARVO (the Association for Research in Vision and Ophthalmology) conference in Fort Lauderdale, Florida, USA from 5-10 May 2002; and at the International Perimetric Society conference in Stratford-upon-Avon, UK, from 26-29 June 2002.

5.1 Previous Work

Spatial filtering is an image processing technique widely used to improve the quality of digital information. Spatial filtering applied to perimetric threshold sensitivity values using a Gaussian filter (fully described elsewhere (Fitzke et al 1995; Crabb et al 1995), and illustrated in Figure 5.1) has been shown to reduce test-retest variability and measurement noise (Crabb et al 1997). In the latter paper, the authors showed that when the first five visual fields in a series were used to predict the field one year later, 72% of the predicted sensitivity values were within ±5dB of the corresponding measured threshold; this improved to 83% after application of a Gaussian filter to the first five fields in the series. Similarly, when a field two years after the fifth test was predicted, the performance improved from 56% to 73% after Gaussian filtering (based on the same criterion). Moreover, benefits from post-test filtering of the data are accrued without any extra patient testing or alteration to the perimetric process.

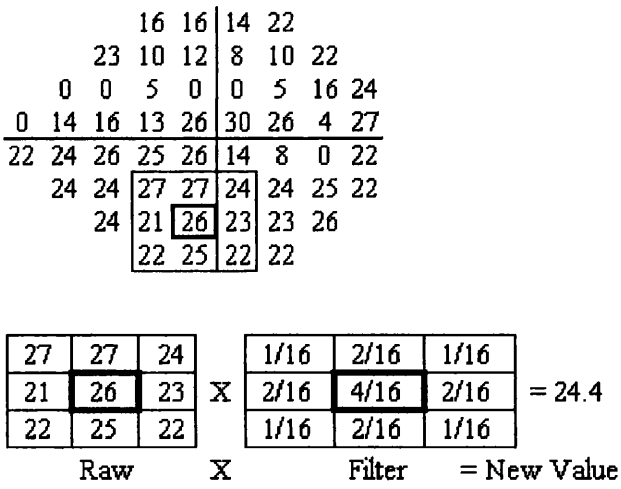


Figure 5.1: How the filtered sensitivity is calculated by the Gaussian filter. In this method, the raw sensitivity value is replaced by one derived from a linear combination of the sensitivities at the nine points in a square centred on the point of interest. This is repeated for each point in the field in turn, each time looking at the points in a square surrounding the point of interest.

However, there are problems with the basic Gaussian filter when applied to perimetric data. The visual field test locations, unlike say pixel values in digital images, are not physiologically linked in a 'grid-like' fashion to their immediate neighbours as indicated by the matrix of values on the visual field chart, and the Gaussian filter takes no account of the actual anatomical structure of the optic disc. For example, glaucomatous defects are generally limited to either the superior hemifield or the inferior hemifield; they rarely cross the horizontal meridian that passes through the fovea (Lachenmayr & Vivell 1993). Yet the Gaussian filtered value at a point adjacent to this meridian is based in part on values from the opposite hemifield. There is general (albeit vague) agreement about the shape of the retinal nerve fibre layer (RNFL), and that the shapes of glaucomatous defects follow the nerves; the Gaussian filter ignores this information. It has been reported that applying the Gaussian filter to data may reduce the proportion of false positives (i.e. the number of healthy points flagged as being part of a defect), but it also reduces the detection of true positives, blurring out true defects. Spry et al (2002) found that when small defects of two or three horizontally adjacent locations progressing at -1dB/year were simulated, the Gaussian filter consistently reduced the sensitivity of PLR, for example after eight tests from around 20% to just 2-3%. Sensitivity was only improved by Gaussian filtering when large defects (nine or eighteen locations) progressing rapidly (-2.5dB/year) were considered. An example of a greyscale where a glaucomatous localised defect is being blurred out by Gaussian filtering is shown later in this chapter in Figure 5.3. This chapter aims to derive a filter taking into account the true structure of the retinal nerve fibre layer, which will reduce the noise without obscuring true defects.

5.2 Mathematical Derivation

The filter is based entirely on real patient data, as described in Section 5.3. It aims to find the relationships between sensitivities at different points in the visual field, and build the filter based on that information.

First, the covariances and correlations between sensitivities at each pair of points in the eye were found. The covariance between sensitivities at points A and B is given by

$$Cov(S_A, S_B) = E(S_A S_B) - E(S_A)E(S_B)$$

Equation 5.1

where $E(S_A)$ is the expected (mean) value of the sensitivity at A, etc. Now, suppose we could accurately predict the sensitivity at point A by a linear combination of the sensitivities at all the other points;

$$\hat{S}_A = k_1 S_1 + k_2 S_2 + k_3 S_3 + \ldots$$

Equation 5.2

for some constants k_1, k_2, ...; then from Equation 5.1,

$$Cov(S_A, S_B) = \sum_i k_i Cov(S_i, S_B)$$

Equation 5.3

We choose to perform regressions on the covariances in Equation 5.3 rather than on the basic sensitivities in Equation 5.2. This is because an apparent relationship between two points, X and Y say, may actually be the product of relationships between each of these points and some third point Z; looking at covariances takes account of this second order of complexity.

Before performing multiple regressions to find these coefficients k_i for each point A, a few constraints were placed on the regressions. Firstly, each k_i had to be non-negative. Secondly, the expected (mean) predicted value must equal the mean raw value for the filter to make sense; so

$$E(\hat{S}_A) = \sum_i k_i E(S_i) = E(S_A)$$

Finally, if the correlation between the sensitivities at point A and point i was low (less than 0.7), k_i was set to equal zero, as those points were considered to be unrelated (this was done purely for computational reasons. In tests on small samples with fewer points it had no effect on the outcome; the regression still identified the same points as being good predictors with associated high values of k_i. In fact a small minority of the k_i were constrained to be zero because of this rule; far more were found to be zero after the regressions were carried out).

So, for each point A, regressions were performed on these series of equations (one equation for each point B) to produce constants k_i, and hence an algorithm for predicting the sensitivity at point A based on those at points elsewhere in the visual field.

The next stage is to produce the filtered value S_A^f for the sensitivity at point A; which should be a combination of the predicted sensitivity and the raw sensitivity at that point:

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

Equation 5.4

The value of c in Equation 5.4 will vary according to the position in the eye; some points are more predictable than others, and these points should have a correspondingly high value of c, hence giving more weight to the predicted value. (Note that in the Gaussian filter, c=12/16 for all points not adjacent to either the edge of the field or the blind spot, as shown in Figure 5.1). Further, the more points that are used in the prediction (i.e. the greater the number of non-zero k_i), the more accurate this predicted value will be, in terms of reduced variability. Now, if the predictions were entirely accurate, the variance of the filtered value would be:

$$\begin{aligned} Var(S_A^f) &= Var(c\hat{S}_A + (1-c)S_A) \\ &= Var(c\sum_i k_i S_i + (1-c)S_A) \\ &= \left(c^2 \sum_i k_i^2 + (1-c)^2 \right)\sigma^2 \end{aligned}$$

Equation 5.5

where σ^2 is the variance of the noise at each point, assumed for this to be constant throughout the eye. Now, we want to choose c to minimise this variability in the filtered value from Equation 5.5, and so reduce the noise as much as possible. Mathematically, the minimum of a function occurs when its differential equals zero. So, differentiating:

$$\begin{aligned} \frac{d}{dc}\left(Var(S_A^f)\right) = \left(2c\sum_i k_i^2 - 2(1-c) \right)\sigma^2 &= 0 \\ \left(\sum_i k_i^2 + 1 \right)\hat{c} &= 1 \\ \hat{c} &= \left(\sum_i k_i^2 + 1 \right)^{-1} \end{aligned}$$

Equation 5.6

So, we have a prescription from Equation 5.6 for choosing c if the predicted values were completely accurate. However, this is clearly not the case; indeed, the predictions at some points will be closer to the raw values than elsewhere in the visual field. So, the correlations $Corr_A$ between the predicted and raw values at each point A were found by calculating the predicted sensitivity at point A for each visual field in the database; so $Corr_A = Corr(S_A, \hat{S}_A)$. This then provides a measure of the predictability of point A; the higher this correlation is, the more accurate the predicted sensitivity and so the more weight should be given to it in the filter. Now, $\hat{c}$ was found in Equation 5.6 based on the assumption that the predicted values were 100% accurate; so in the final filter, the weighting given to the predicted sensitivity in the filter, is given by $c = \hat{c} \times Corr_A$. So, having found the k_i for point A by multiple regressions based on Equation 5.3 (as described above), the final filtering algorithm for point A is:

$$S_A^f = \frac{Corr_A}{\sum_i k_i^2 + 1}\left(\sum_i k_i S_i \right) + \left(1 - \frac{Corr_A}{\sum_i k_i^2 + 1} \right) S_A$$

Equation 5.7

Obviously, the k_i differ according to which point in the visual field is being considered (point A); and so the filter is, in effect, a matrix containing the coefficients for filtering each point in the field, coefficients which vary from point to point.

5.3 The Data Used

The work is based on a visual fields database consisting of patients seen by the Moorfields Eye Hospital Glaucoma Service in London, UK. This means that the filter is based on, and designed for use in, a tertiary glaucoma referral service. It would be feasible to utilise the methods in Section 5.1 to produce a family of filters for different situations, such as population screening events or general ophthalmic clinical settings; however the process would be hindered by the increased proportion of non-glaucomatous defects. It would be expected that the results would be extremely similar.

The database contains 98,821 visual fields, taken from 14,675 individual suspected glaucoma patients. Fields were measured using the Humphrey Visual Field Analyser (Humphrey Instruments Inc, Dublin, California, USA), as described in Section 1.2. The data goes back as far as 1985; it consists of both 30-2 fields and more recently 24-2 fields (all standard white-on-white, full threshold tests), although only complete 24-2 fields (i.e. the non-edge points in the 30-2 fields) were used. This database, consisting of around five million individual threshold values, provides an excellent and unmatched resource giving a comprehensive and representative cross-section of all tests carried out at glaucoma clinics. The database was cleaned to remove duplicated visual fields; and each patient given a unique ID number. This way, the data was completely anonymised. Visual fields from right eyes were transposed, and their mirror images used instead as 'left' eyes.

When deriving the new filter, no patients were excluded. It was sought to use as representative a sample as possible of all the patients entering the clinic, on an "intention-to-treat" basis. This means that the data used is not restricted to patients who actually have glaucoma, merely patients who are sent to the clinic because it is believed at the time that they may have glaucoma. A filter derived purely on data from glaucomatous patients may not be accurate for healthy eyes, and vice-versa; so all available data was used in the belief that this gives added weight to our results. This means that the results are truly representative of the average test carried out in a glaucoma clinic.

It is important to note that unreliable fields, or fields from patients presenting at the glaucoma clinic who turn out to have other pathologies (e.g. pituitary tumours, myelinated nerve fibres etc), were not removed from the database prior to analysis. This is because we seek to represent the typical test carried out at a glaucoma clinic, and so these patients should not be ignored. The principle of "intention-to-treat" is well founded, and is very commonly used throughout the scientific and statistical literature.

A further important distinction should be drawn between using every available test from each patient (making the results representative of the average visual field test carried out in a glaucoma clinic) and

using one randomly selected field from each patient (which would make the results representative of the average patient seen by a glaucoma clinic). Since patients with true glaucoma would be expected to have more fields in the database than patients with other pathologies, and since it is more important to deal with patients with glaucoma who make repeated visits to the clinic, the choice was made to use all the fields from each patient.

There are arguments both for and against these positions, as set out below.

5.3.1 Use of Unreliable or Non-glaucomatous Fields

The purpose of the work is to develop a spatial filter that takes into account the structure of the RNFL, but the large database may include patients with defects that do not respect this anatomic structure, caused by diseases such as pituitary tumours and other neurological defects. These patients should be a small minority of those in the database, because the data comes from a glaucoma referral service. Moreover, the non-exclusion of unreliable visual fields may also produce defects which do not resemble the true structure of the RNFL. The effect of these fields on the derivation of the filter would be to reduce the covariances in Equation 5.1 between points which are actually anatomically close in the RNFL, while slightly increasing the covariances between anatomically distant points. In effect, the clarity would be marginally reduced; it makes it less certain which points are related and which are not. The process of carrying out regressions on the covariances will reduce the effect of this loss of clarity further. It will have a minimal effect on the final filter, but potentially not a negligible effect.

The database will inevitably contain a proportion of normal fields. These would not be expected to provide as much information, if any, about the functional correlate of the RNFL anatomy. The effect of including these fields will be to raise all the pair-wise covariances between points. The subsequent regressions carried out on the covariances should largely eliminate this effect, since points anatomically close will still have a higher covariance than points anatomically distant, assuming that including normal fields has raised all covariances by the same amount. This assumption, though reasonable, has not been tested.

Against these arguments, including all patients will increase the available amount of data. The size of the database is considerable, so removing these patients would still leave a database of a more than acceptable size (an order of magnitude larger than used in most published studies); but it is always better to use as much data as possible.

Excluding unreliable fields brings the question of what should constitute an unreliable field; what should the cut-off point be for parameters such as False Negatives and False Positives? Similarly, there is no clear definition as to what constitutes a glaucomatous defect, and so choosing a cut-off point for whether a defect is glaucomatous or from another disease would be problematic. Again, the question arises as to how to distinguish between normal and glaucomatous fields; even if it were feasible to examine each field individually, agreement between clinicians is poor, and any rigid statistical cut-off point would be, in essence, arbitrary.

"Intention to treat" is a well-established principle in medical statistics; indeed drug regulatory bodies often insist on it. The aim of this study is to develop a filter for use in a glaucoma clinic, and if unreliable and non-glaucomatous patients are present (as is the assumption behind the other side of the discussion), the filter must also work for those patients. Similarly, excluding healthy eyes may possibly result in a filter applicable only to damaged eyes, but unreliable on healthy eyes, which make up the majority of the population. Whether this would be the case or not would be interesting to test, by deriving a filter from the same methods but based solely on glaucomatous eyes, and comparing the results; however this again would cause the problem of defining a healthy or glaucomatous eye, as mentioned above.

If healthy eyes were to be excluded because they bring negligible information about the structure of the RNFL, then logically would eyes which are severely diseased and with zero or near-zero sensitivity throughout the eye (or even throughout one hemifield) also have to be excluded? Further, if for example an arcuate defect is present in the upper hemifield, but the lower hemifield is perfectly normal, would the lower hemifield be excluded?

Excluding unreliable fields may give a more accurate picture of the relationships between points. However, it would do so at the expense of losing information about the noise. By definition, unreliable fields are those with more significant – and therefore more interesting – noise.

In conclusion, it is the author's belief that any detrimental effects which these fields have on the results are minimal, as explained above, because of the regressions carried out. Contrastingly, the benefits they bring (principally a more representative database), and the problems caused by omitting them (such as choosing the cut-off points), are both significant. So, in balance, it was concluded that the arguments against excluding patients outweigh the arguments for.

5.3.2 Use of All Fields from Each Patient

Data from fields obtained from a single subject over time are correlated with each other, thus they are inputting the same or related defects multiple times in the derivation of the new filter, introducing a bias that may not reflect the anatomy of the RNFL. It is certainly true that the same defect from the same patient eye will appear more than once in the database, and this may bias the filter structure towards the spatial relationships between test locations for individuals with more visual field tests in the database.

However, selecting one field at random from each patient would bias the database towards healthy eyes or eyes in the early stages of disease, because eyes tested only once or twice (and so likely to be healthy eyes which did not require further testing) would make up a far larger proportion of the database. This would greatly worsen any problems associated with including healthy or non-glaucomatous eyes in the database, as described in Section 5.3.1. Also, as before, the size of the database would also be greatly reduced, from 98,821 fields down to 14,675 fields. While still a very large database, this reduction is significant.

It is more crucial that the filter works on patients who attend regularly over a period of many years, since they are generally the patients who have progressive glaucoma, or are most at risk of progression. Improving their test results is more important than improving the test results of a patient with no defect who only comes to the clinic once. Therefore it is logical to give more weighting to patients with longer series of visual fields; this is best achieved by including all fields for each patient. A key potential use for the filter is to improve methods for detecting progression. In these circumstances, a series of several fields is present. Again, this makes it more important to accurately reflect the structure of such eyes.

As stated earlier, using every field for each patient produces a filter representative of the average test performed in the clinic, rather than representative of the average patient. It is our belief that this makes the filter more useful more of the time. It will also lessen the effect of eyes with non-glaucomatous defects, as discussed in Section 5.3.1. Any negative effects caused by the decision above to include all patients, irrespective of normality, other diseases and unreliability, will be reduced still further by using all the visual fields.

Using every field for each patient would seem liable to bias the results in favour of reproducing the structure of patients who attend repeatedly; principally those patients who have progressing glaucomatous defects. This is not necessarily a bad thing. The arguments on either side are not definitively conclusive. However, especially in view of the first decision above, it was decided that the arguments for including every field outweighed the arguments for using just one field per eye per patient.

5.4 The Filter

The filter conforms to the accepted physiological shape of the retinal nerve fibre layer. Figure 5.2 shows the points that are used to filter a few 'Central Points' (i.e. those points with a non-zero k_i when the relevant Central Point is 'point A' in the derivation in Section 5.2). If a point is connected to the Central Point by a line, it indicates that that point is a predictor for the Central Point; the thicker the line, the larger the effect it has on the prediction (i.e. the higher k_i is for that point), as shown in the key on the right. The remaining contribution to the filtered value comes from the Central Point itself. It is seen that predictors are not necessarily neighbours of the Central Point (as they would be if the Gaussian filter was being considered), but they follow the expected arcs. As seen in Figure 5.1, this diagram for the Gaussian filter would simply connect each Central Point with all of its immediate neighbours.

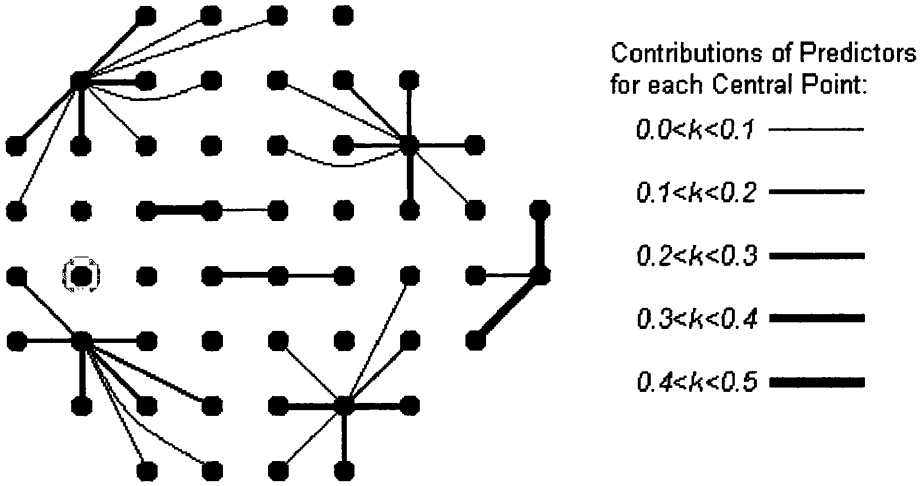


Figure 5.2: An illustration of the shape of the filter. The grey area represents the blind spot.

For example, for point 49 (the Central Point nearest the bottom right in Figure 5.2), the filter is given by:

$$S_{49}^{f} = 0.016S_{34} + 0.031S_{41} + 0.164S_{43} + 0.219S_{48} + 0.180S_{50} + 0.055S_{53} + 0.078S_{54} + 0.218S_{49}$$

Furthermore, points on opposite sides of the horizontal meridian only turn out to be significant predictors on one occasion (not including the points beyond the blind spot, to the far left in Figure 5.2). This is very promising, since the derivation of the filter at no point took account of the relative positions of point in the visual field. The only location where a point does turn out to be a significant predictor of a point on the opposite side of the horizontal meridian is at the nasal step, as shown on the right hand side of Figure 5.2. This observation is interesting in itself, since nasal step defects are seen to cross the horizontal meridian in clinical practice. One possible explanation is the anatomical notion that some nerve fibres from upper and lower hemiretinas interdigitate at the temporal end of the horizontal raphe (Vrabec 1966; Sakai et al 1987); alternatively, the phenomenon could be caused by local effects from the glaucomatous process or a testing artefact. Also, the filter as derived above always assigns a sensitivity of zero to the point labelled in grey in Figure 5.2, which corresponds to the blind spot (10 to 15 degrees temporal of the fovea, just below the horizontal meridian); again, this was not pre-determined by the method of deriving the filtering algorithm.

Clearly, the noisier the initial (raw) visual field is, the more obvious to the naked eye will be the difference caused by filtering. The effects of the filter are subtle; although improvements undetectable when simply viewing greyscales of the data may have much larger effects over time when a computer analysis, such as PLR (Fitzke et al 1996) or the Glaucoma Change Probability (Heijl et al 1991a), is used. However, there are cases when the benefits are clear. Figure 5.3 shows a 24-2 greyscale from an eye, taken from a textbook example (Budenz 1997), noted as having moderate to early glaucoma. There is a scotoma superior to the nasal horizontal meridian partially obscured by the noise present throughout the field. Filtering emphasises this defect, at the same time as removing most of the random noise elsewhere in the field. In contrast, the Gaussian filter almost entirely blurs out this defect. Note also that, unlike the Gaussian filter, the new filter leaves the blind spot intact.

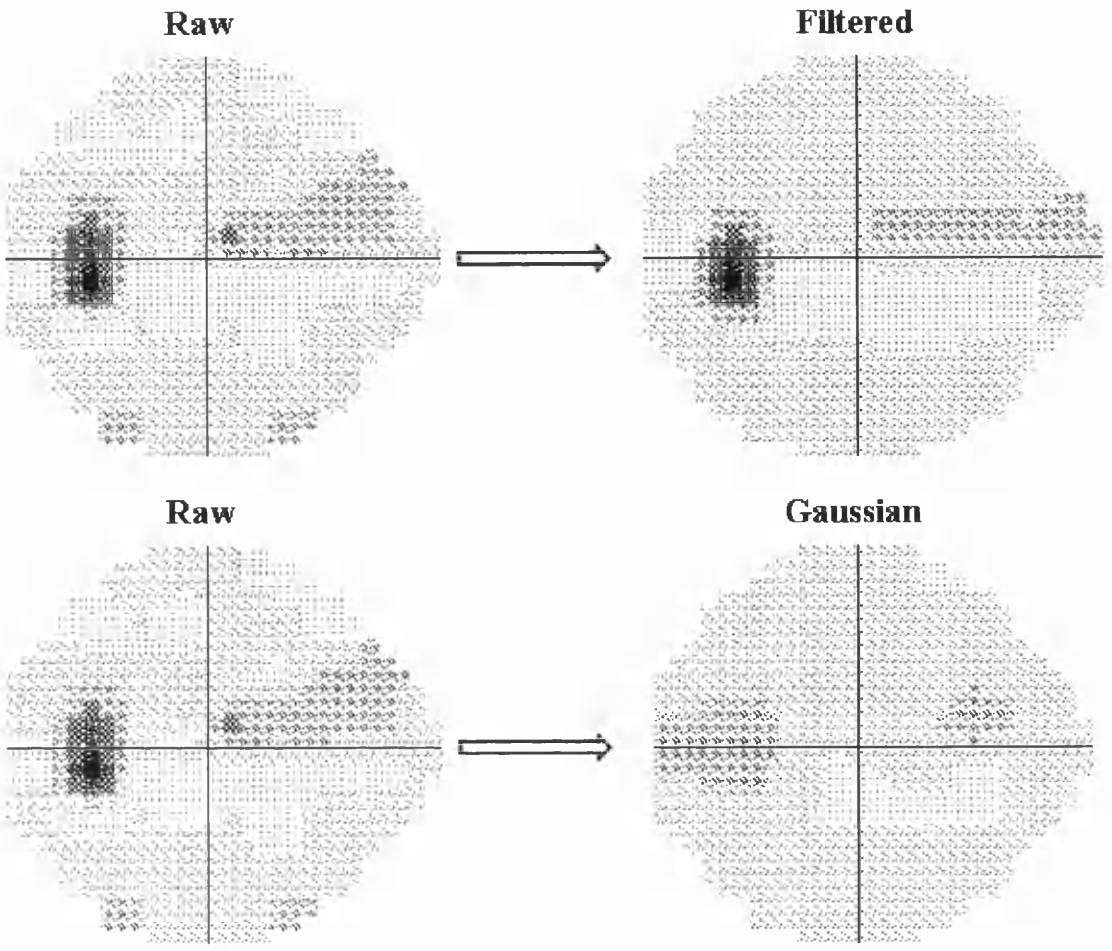


Figure 5.3: The effect of the new filter on a visual field exhibiting a scotoma in the superior hemifield, compared with the effect of applying the Gaussian filter to the same data.

So, the new filter not only appears to conform to the presumed shape of glaucomatous defects, but also appears to reduce the noise in the case of this example. However, more formal testing is naturally required. This will be addressed in Chapter 6.

In order for this filter to be used clinically, it must be implemented into currently-available software. Appendix 2 gives details of the coding necessary to implement the filter into the Progressor program, replacing the Gaussian filter. It is hoped that this will be included in any versions of Progressor which are released to clinicians in the future.

An example of a visual field from the demonstration patient provided by Progressor is shown on the left in Figure 5.4; the effect of applying the spatial filter to this visual field is shown in the greyscale on the right. The result is a clearer, less noisy picture, without the glaucomatous defect being blurred out and disappearing.

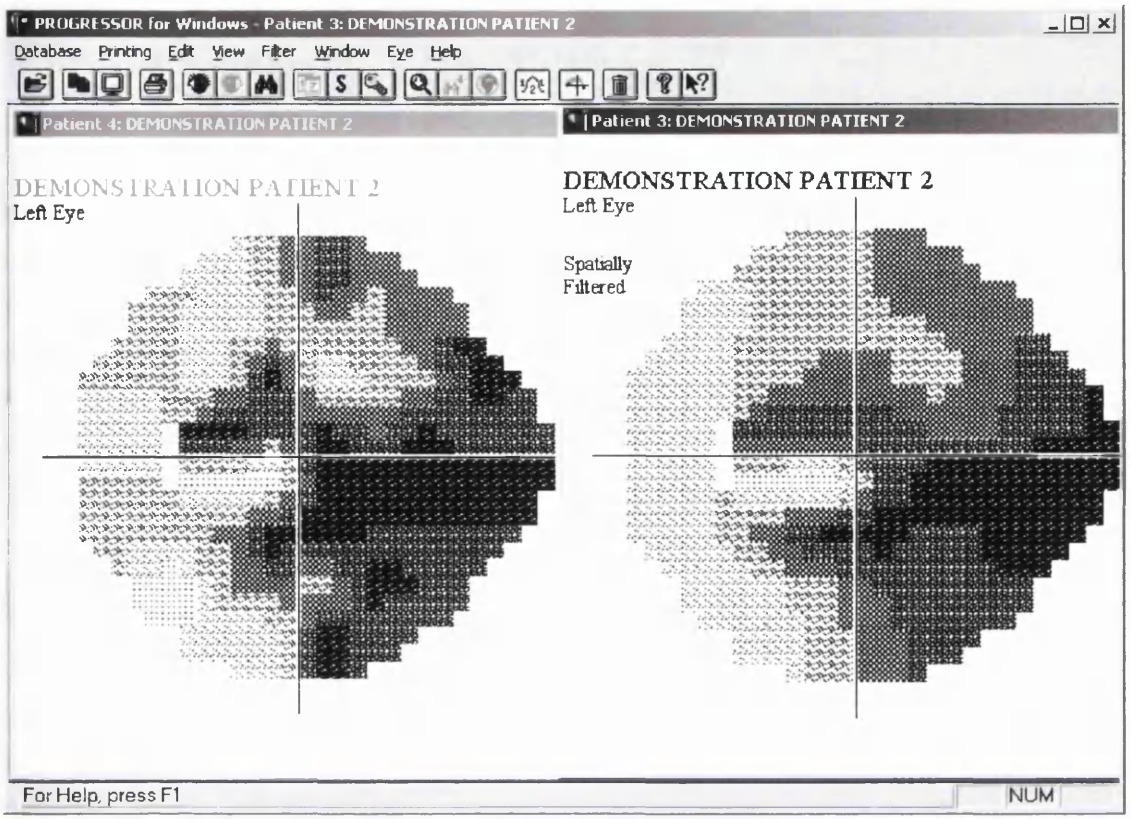


Figure 5.4: An example of the use of Progressor to apply the filter to a visual field from one of the program's demonstration patients. The same field is shown twice, with the raw field on the left and the filtered field on the right.

6. Testing The Filter

There are two key questions that need to be addressed when considering the effectiveness or otherwise of the new filter derived in Chapter 5. Firstly, does the filter reduce noise; this after all is the purpose of the filter. Secondly, does the filter blur out actual glaucomatous defects, a criticism which has previously been levelled at the Gaussian filter (Spry et al 2002). Two methods are described below for testing the filter. The first uses the longitudinal properties of the database to assess its effect on pointwise linear regression (PLR); the second and main examination of the technique uses the Virtual Eye simulation developed in Chapter 2.

6.1 Longitudinal Data

PLR, as described in Section 1.3, examines how the sensitivity at a given point in the visual field changes over time by fitting a linear regression model to the series of sensitivity values for each individual point in the field. It would be expected that if the noise were reduced, the sensitivity values would conform more closely to the trend over time; that is, the fit of a linear regression model would be improved. If the between-test variability is taken as being the sum of short-term variability and long-term variability, as described in Section 1.2.1, then the filter would be hoped to greatly reduce the short-term variability component. Indeed, the Gaussian filter has previously been shown to improve the consistency over time of visual fields (Fitzke et al 1995; Crabb et al 1995).

The database described in Section 5.3 contains a number of eyes which have been tested repeatedly over a long period, in some cases as many as forty times over fifteen years. This longitudinal feature of the data was not used in the derivation of the filter (which treated each visual field as being equally important). So, it is proposed that using this feature could provide one indication of the effectiveness of the filter, without the results being biased by being taken from the same database.

For this purpose, all series of length at least twenty (i.e. twenty visual fields taken from the same eye of the same patient on different dates) were extracted from the database. There were 303 such series. For those series of longer lengths, only the first twenty fields were taken. Now, the idea of filtering is to reduce the noise present in the readings; so if this has happened, there should be less variability over time in the sensitivity of a given point; the correlation between the sensitivity and the date of testing will increase. This is the same as saying that the R-Squared value obtained when carrying out linear regression to find the rate of change at that point will improve upon filtering. So, for each series of twenty fields, this analysis was carried out at each of the 54 points in the central 24-2 field in turn; giving a total of 303x54=16362 R-Squared values. The same analysis was then repeated after filtering the data (by each of the two filtering methods – using the Gaussian filter, and using the new filter developed in the previous chapter – in turn).

A high R-Squared value, or equivalently a high correlation between the sensitivity at a point and time, indicates that there is little between-test variability at that point, as the sensitivities closely follow the overall trend over time. As seen from the density functions shown in Figure 6.1, the R-Squared values generally increased after filtering, whether it is using the Gaussian filter or the new filter.

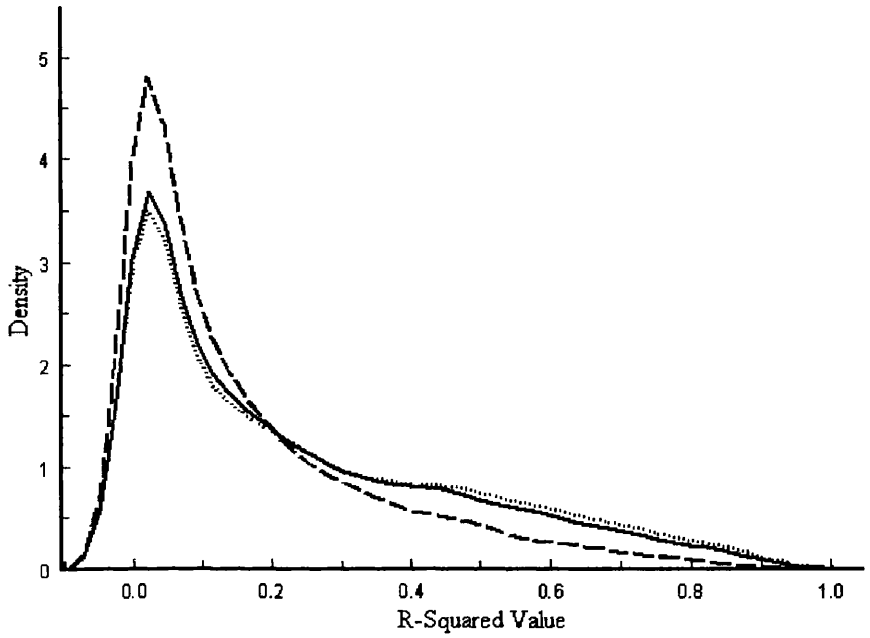


Figure 6.1: Frequency Distribution of R-Squared values for series of Raw sensitivities (dashed line); Gaussian-filtered sensitivities (dotted line) and Filtered sensitivities (solid line).

Out of the 16362 regressions performed (54 points from each of 303 series of fields), 1164 series of raw sensitivities (7.1%) had an R-Squared value of greater than 0.5; whilst 2446 series of filtered sensitivities (14.9%) had a similarly high R-Squared value. The 95% confidence interval for the mean of the changes in R-Squared values brought about by filtering was (0.066, 0.070), highly statistically significant ($p<0.001$) according to a paired t-test. This suggests the process of filtering (by either method) is significantly reducing the noise, as the sensitivities are more consistent over time.

6.2 Virtual Eye Simulation

It has previously been reported that although the Gaussian filter may reduce noise, it also reduces the signal, as described by Spry et al (2002) and discussed in Section 5.1; i.e. methods which use filtering will be less successful at detecting localised defects than methods based on the raw data. It was decided to investigate this by looking at the effect of filtering on simulated visual field series, using pointwise linear regression (PLR) as the testing tool.

6.2.1 The Simulated Visual Fields

The simulated data was generated using the "Virtual Eye" described in Chapter 2. Series of six fields (from annual testing over a period of five years) were simulated based on input noise-free visual fields;

the simulation adds random noise to each point of each of the six fields in the series. It was sought to make these noise-free series as simple as possible, to simplify and clarify the results obtained. So, each point in the field was assigned as one of:

1. *Stable*: the sensitivity remains constant at 30dB throughout the series, before noise is added.
2. *Defective*: the noise-free sensitivity is reduced by 2dB per year, from 30dB to 20dB over the five years.
3. *Border*: the noise-free sensitivity is reduced by 1dB per year, from 30dB to 25dB.

This process was first carried out on a stable visual field; i.e. one where all 54 points in the visual field (including the two points coinciding with the blind spot) were *Stable*.

Next, localised defects were chosen with reference to the map of the physiological optic nerve head (ONH) locations for each visual field test point produced by Garway-Heath et al (2000b), shown in Figure 6.2. The rationale for using this map is that localised defects are more likely to occur in clusters determined by the anatomical entry position of nerve fibre bundles into the optic disc. This rationale partly satisfies the need for determining localised defects separate from one based on perimetric data alone, avoiding the use of purely arbitrary and subjective defects such as those used by Spry et al (2002). Using the map, a 'Central Point' was chosen, and its location in the ONH noted. This point and all other points located within ±5 degrees of the Central Point in the ONH were assigned as being *Defective* points with a glaucomatous loss of 2dB per year. Further points between six and ten degrees away from the Central Point (in terms of ONH location) were assigned as being *Border* points, with a glaucomatous loss of 1dB per year. Points elsewhere in the eye were assumed to be *Stable*.

			268	262	252	245		
		264	274	281	275	260	246	
	271	285	291	296	298	283	253	229
278	287	291	298	312	329	318	.	218
83	76	68	55	34	11	13	.	167
	85	78	66	56	48	60	95	136
		88	81	77	80	93	112	
			93	95	100	108		

Figure 6.2: The optic nerve head location, in degrees, for each non-blind spot visual field test point in a right eye. The temporal meridian (9-o'clock position) was designated as 0°, and degrees counted clockwise from there. Reproduced from Garway-Heath et al (2000).

Each of the 54 points in the standard visual field was considered in turn as the Central Point, with the exceptions of the two points forming the blind spot. Some examples of defects generated in the manner described are shown in Figure 6.3.

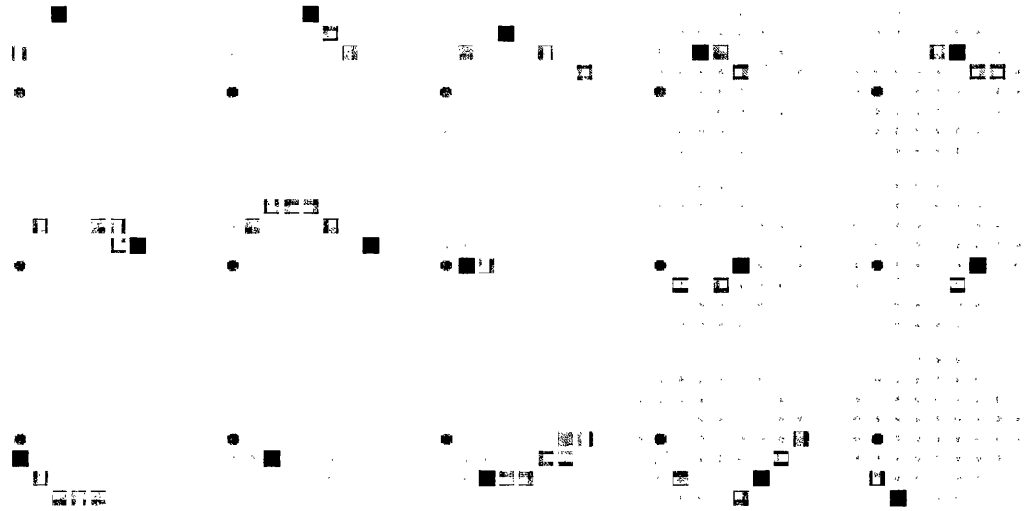


Figure 6.3: The shapes of some of the 52 generated defects tested, superimposed onto a Humphrey 24-2 visual field. The black square in each case is the Central Point. Dark grey squares represent the *Defective* points (within five degrees of the Central Point in the ONH); light grey squares represent the *Border* points (between six and ten degrees away from the Central Point in the ONH). The large grey circle towards the left of each field represents the location of the physiological blind spot.

Anderson's criteria (Anderson & Patella 1999) define a clinically significant progressing localised defect as being a cluster of three or more points that have sensitivities occurring in fewer that 5% of the normal population, at least one of which must occur in less that 1% of the population, according to the pattern deviation probability plot. Hence, localised defects generated in this way which were found to consist of fewer than three adjacent points, and therefore not constituting an identifiable defect according to Anderson's criteria, were separated out and will be commented further upon later (this comprised six 'defects' consisting of a solitary point, and six 'defects' of just two points, one example of which is shown in the centre of Figure 6.3).

Finally, the process was carried out on a field with a non-glaucomatous defect (as shown in Figure 6.4), consisting of three *Defective* points with one *Border* point at each end. This was done in the hope that a filter designed to identify glaucomatous defects will not pick up defects which would not occur in glaucoma. Such a defect may be genuine, caused by for example a neurological disease; but since the purpose of this work is to more readily identify the existence and progression of glaucomatous defects, and it is not designed for use when other suspected conditions are being looked for, it is desirable that such non-glaucomatous defects be blurred out.



Figure 6.4: The shape of the non-glaucomatous defect used, superimposed onto a Humphrey 24-2 visual field. As before, the grey circle represents the physiological blind spot.

Each input series was simulated five thousand times. After each run, points were tested to see whether they would be flagged as progressing using PLR. The criteria for a progressing point used in these simulation experiments were a deterioration of at least 1dB/year, statistically significant at the 1% level.

This criteria has been used in published studies (Viswanathan et al 1999; Nouri-Mahdavi et al 1997; Membrey et al 2001). This analysis was repeated after filtering each field in the noisy simulated series using each of the Gaussian and our new filter in turn. So for each of the five thousand runs for each input series, the points were tested three times. In this way, the sample probabilities of the points being flagged as progressing based on each of the Raw, Gaussian and Filtered data were calculated. In the case of the input localised defects, the probabilities of the Central Point being (correctly) flagged as progressing in each case will hereafter be referred to as the three Detection Rates for that defect.

6.2.2 Results from the Virtual Eye Experiments

When the stable eye was tested (i.e. one where, before the addition of noise, all the points remain at a sensitivity of 30dB throughout), the proportion of points still being flagged as progressing (all of which are therefore false positives) fell dramatically from 0.59% to 0.04% upon applying our new filter. This means that the probability of at least one of the 52 non-blind spot points being wrongly flagged as progressing has fallen from 26.5% to just 2.1%. This is further evidence that the filter is reducing the noise and increasing specificity, as found in Section 6.1.

For each of the artificial localised defects based on the ONH structure (as in the examples in Figure 6.3), the percentage of the five thousand runs of the simulation which resulted in the Central Point satisfying the standard criteria for progression under PLR (i.e. a slope of at least –1dB/yr, statistically significant at the 1% level) was calculated. This percentage was found after the same five thousand noisy series had been filtered using the Gaussian filter (marked as 'Gaussian'), and using the new filter (marked as 'Filtered'), as well as before filtering (marked as 'Raw'). The results are shown in Figure 6.5.

The detection rates for the Central Points vary considerably after filtering, since they depend in part on the status (*Stable*, *Defective* or *Border*) of neighbouring points. Without any filtering, the detection rates are consistently around 20% for the Central Point and other *Defective* points (which are deteriorating at 2dB/yr) and around 8% for the *Border* points (which are deteriorating at 1dB/yr); because PLR without filtering regards each point as being independent, this value is approximately constant. It is seen that in almost all cases, the percentage of Central Points correctly identified as progressing increased significantly after filtering. Further, the new filter is performing much better than the Gaussian filter in this regard; it was common for defects to be blurred out by the Gaussian filter. The mean detection rate for the localised defects was 19.9% for the unfiltered, raw data; rising to 29.9% after application of the Gaussian filter, and 38.0% after using our new filter. This indicates a dramatic improvement in sensitivity when the field is filtered before it is analysed.

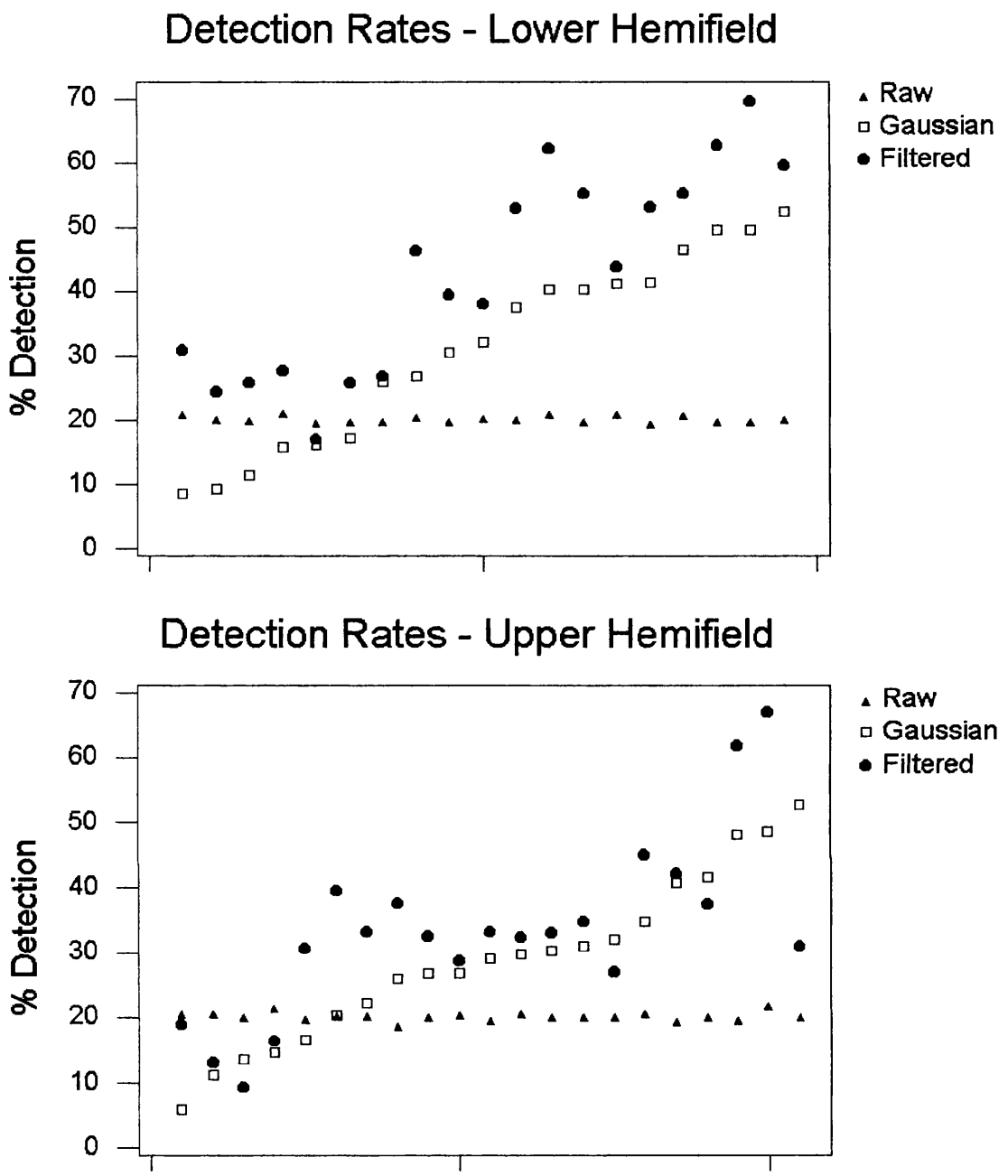


Figure 6.5: Percentage of Central Points correctly flagged as progressing for each of the tested defects. Each column contains three symbols, representing the detection rates for the raw data, the data after application of the Gaussian filter, and the rate after application of our new filter; there is one column per point tested. The points have been split into the two hemifields. The points are in each case shown in order of increasing detection rate after application of the Gaussian filter (rather than in order of location), to make the figure clearer to understand. No relationship was found between the area of the field which a point was in and the effect of filtering on the detection rate.

As mentioned in Section 6.2.1, of the 54 possible Central Points in the central visual field, fourteen resulted in defects consisting of fewer than three adjacent points and so did not satisfy the criteria of Anderson & Patella (1999) (comprising the two points coinciding with the physiological blind spot, six solitary point 'defects', and six 'defects' consisting of two points). Because they did not produce realistic glaucomatous defects (according to this definition), results from these fourteen Central Points are not included in Figure 6.5. However, this does not necessarily mean that these small defects were not picked up. Indeed, the two-point localised defect in the centre of the field next to the blind spot, illustrated by the

central field in Figure 6.3, resulted in the detection rate increasing from 19.9% to 37.1% after filtering. Certainly, as with any type of spatial filtering, if a defect consists of only one solitary point its detection rate will decrease. However, one deteriorating location may not constitute clinical glaucomatous progression.

Of more concern are the five Central Points for which the detection rate in Figure 6.5 appears to be reduced by filtering. These five defects are those shown in Figure 6.6. In each case, the apparent reduction in the detection rate caused by filtering is explained by the differences in shape between the defect being tested (generated from the ONH map) and the shape of the filter at that point which is shown in the bottom row of Figure 6.6 (generated from visual field data). A large proportion of the filtered value for the Central Point is based on points not appearing in the defect; over 50% for defects *A*, *D* and *E*, and between 40% and 50% for defects *B* and *C*. These five isolated cases may appear to behave in this way because of the method used for generating simulated defects, and are not necessarily reason to doubt the usefulness of the filter; they may reasonably be assumed to be the result of the difference between functional and structural relationships, as discussed in Section 1.2, and the subject of continuing work in the field (Shareef et al 2002).

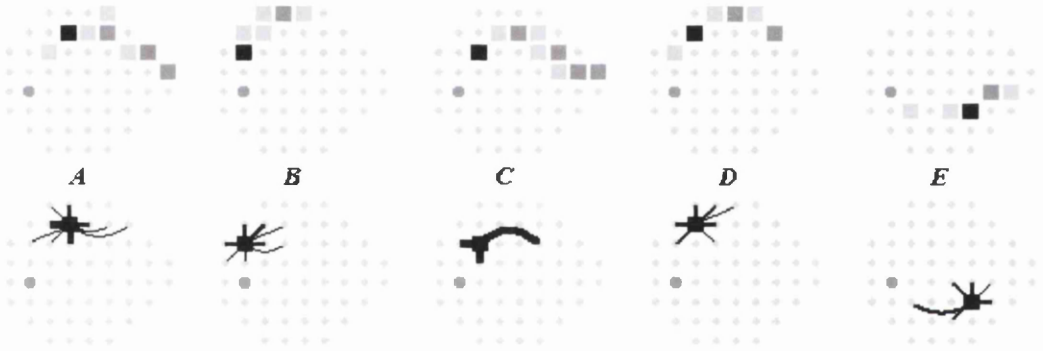


Figure 6.6: The five generated localised defects whose detection rate appears to be reduced by filtering in Figure 6.5, using the same symbols as in Figure 6.3. The lower line shows the points involved in the filtering algorithm for that particular Central Point; the thicker the line, the more effect the point has on the filtered value for the Central Point, as in Figure 5.2.

Furthermore, in the first three cases, the overall detection rate taken over all the points which are deteriorating at 2dB/yr (the Central Point, represented by the black square, and the *Defective* points represented by dark grey squares), rather than just the Central Point, has actually improved. With defect *A*, the average detection rate has in fact increased from 19.9% to 29.1% after filtering (Gaussian: 22.3%); although the Central Point itself is being flagged as progressing less frequently, the rest of the defect is more likely to be picked up, and so the overall detection rate has still improved. Similarly for defect *B*, where the average has increased from 19.6% to 21.5% (Gaussian: 23.6%); and for defect *C*, where the average has increased from 20.1% to 31.5% (Gaussian: 26.9%).

The defect shown in Figure 6.4 was tested in the same way. Averaged over the three *Defective* points, the filter reduced the percentage of runs of the simulation which resulted in points being flagged as progressing from 19.7% to 10.0%; blurring out non-glaucomatous defects, as hoped. The Gaussian filter did not have this beneficial effect; in fact it increased the proportion flagged from 19.7% to 29.6%.

6.3 Conclusions

The principle of filtering real-world data is long established as a tool for reducing noise. Indeed, even in a standard measurement of the visual field with a Humphrey perimeter, double determinations of the sensitivity are carried out at certain points in the field; and while this is done principally to provide an estimate of the variability present in the readings, it also provides a basic form of filtering, as clinicians may base their judgment on the mean of these double determinations. The semi-Bayesian approach of using the prior expected sensitivity at a point to predict it before it is measured is also utilised in the SITA algorithm (Bengtsson et al 1997). Spatial filtering, in the form of a Gaussian filter, has also been applied to visual field data with the benefit of reducing between test variability and improving the predictability of future visual field changes (Fitzke et al 1995; Crabb et al 1995; Crabb et al 1997). However, a recent study using a computer simulation of progressive glaucomatous visual field loss (Spry et al 2002) concluded that Gaussian filtering does not offer a consistent benefit over the analysis of raw visual field data, and in some instances significantly inhibits the ability to detect small, gradual progressive field changes. In support of these findings, our testing of the Gaussian filter on our selection of localised defects also showed that, in some (but not all) cases, it blurred out smaller progressive defects. With our new improved filtering algorithm, the detection of progressing points not only did not worsen, but it actually improved significantly, even in cases of very small clusters of progressing points. The new filter also offers dramatic improvement in specificity when compared with unfiltered fields (reducing the false positive rate from 0.59% to 0.04%), vital for pointwise methods for detecting progression. Indeed, a seemingly small improvement in specificity may be more important than a larger improvement in sensitivity, as shown in Section 3.3 (although here we are in the happy position of improving both!).

For any new method to become widespread in clinical use, it must be widely tested and demonstrated to provide a notable improvement on the current methodology. Opinions will naturally differ on what criteria are necessary for such a new method to be adopted. In this paper, a simulation method has been used to test the filter. Simulation, which is becoming an increasingly utilised tool in visual field problems (Johnson et al 1992; Glass et al 1995; Spenceley & Henson 1996; Spry et al 2000; Spry et al 2002; Gardiner & Crabb 2002 a,b; Vesti et al 2002; Artes et al 2003), was used to test not only whether the noise was being reduced by filtering, but also whether true localised defects were being blurred out. The details of the simulation used, and its benefits and disadvantages, are discussed in Chapter 2 and in Section 3.3. With any simulation, the results should be viewed qualitatively rather than quantitatively, due to uncertainty over such issues as the amount of noise used etc. Nevertheless, these experiments have clearly shown that defects are emphasised by the new filter, whilst there are less false positives from stable points and non-glaucomatous defects.

The use of a map of the ONH for generating localised defects (Garway-Heath et al 2000) has the notable advantage that it is not wholly based on perimetric results, thus making it suitable for independently selecting progressive defects. The rare instances of progressive defects where the new filter did not emphasise progression may tell us more about that lack of relationship between functional and structural

change; or equally, may be a result of correlations forged as part of the testing strategy of the perimeter itself in establishing thresholds. Some correlation between points, whether contiguous or not, may well be the result of the testing strategy and not of the physiological connections. The effect of this is that the range of defects generated may not be entirely realistic or comprehensive. Further, it means that the detection rate being reduced by filtering in the cases of two of the tested defects (defects *D* and *E* in Figure 6.6) need not be considered a problem with the filter; more, it may point to a difference between the structural and functional maps of the eye at those points. It is perfectly believable, and in fact to be expected, that there will be a few points in the eye where the results from this testing technique would not be perfect even if the filter were. This is a limitation of using this method for choosing realistic localised defects; yet this is preferable to choosing defects in a completely subjective manner based on common beliefs about visual fields, which would be biased towards the expected shapes.

Naturally, any further testing is to be welcomed. It would be desirable to also test whether or not localised defects are blurred out by filtering by looking at real patient data. This is hampered by the lack of a definition in the literature of what constitutes an expected defect; progressing or otherwise (hence why our simulation uses shapes of the localised defects generated by looking at a physiological map of the ONH). One possibility is to look at the predictive power of trend analysis methods, in the hope that this power would increase once the data had been filtered; however to do this would require a clean, independent database, and so it has not been carried out to date. Despite these inherent problems, and although the results of simulation should not be underestimated, they still need to be confirmed by clinical observations.

It should be emphasised that the filter described here is designed solely for suspected glaucoma patients. It is based on data from a glaucoma clinic, and so resembles the shapes of glaucomatous defects. This means that it is unsuitable for use when other conditions are suspected; neurological defects, for example, will be blurred out by the use of this filter, as demonstrated by the effect of the filter on the defect shown in Figure 6.4. Of course, when the issue is whether the patient has glaucoma or not, or whether their glaucoma is progressing, it is entirely desirable that non-glaucomatous defects be blurred out.

Even without further testing, there is sufficient evidence presented here to say that the new filter shows clinical potential, especially since it requires no changing of visual field testing or extra patient test time. It is to be hoped that the filter technique described here, and implemented into the Progressor program by the code given in Appendix 2, could in the future become a widely accepted tool in glaucoma clinics.

7. Modelling Noise

When using simulation techniques as a substitute for real patient data, as in the Virtual Eye, it is naturally essential that the noise is modelled in the best possible manner. The Virtual Eye, in common with other simulations (Crabb et al 1999; Spry et al 2000; Vesti et al 2002), assumes the noise to be normally distributed about zero. However, this assumption is not strictly accurate. From studies looking at the test-retest variability of visual fields (Heijl et al 1989; Artes et al 2002), it is known that the noise becomes increasingly negatively skewed as the sensitivity decreases. The assumption of normality appears reasonable for an eye which is healthy or shows only a small defect; but for more advanced defects the assumption breaks down. Artes et al (2002) report that for an initial baseline sensitivity of between 8dB and 10dB, measured using the SITA Standard testing strategy, the distribution of the retest sensitivity has a lower quartile of 4dB and an upper quartile of 20dB, with a median of 14dB; not only is this clearly non-symmetrical, but it is also not centred on the initial baseline sensitivity. The change in the distribution of the noise with the SITA testing strategy may be partly because the algorithm tries to fit the data to one of two models, either normal or defective, rather than a continuum of states.

This non-normality may be due in part to the convention of using the decibel scale to measure sensitivity. As described in Chapter 1.2, the sensitivity S in decibels (dB) is given by $S = 10 * \log (1/L)$, where L is the stimulus intensity measured in Lamberts (1Lambert = 10,000 apostilbs). This means that a small change in decibels may correspond to a large reduction in the numbers of light-sensitive cells when sensitivity is near normal, and a small reduction in the numbers of light-sensitive cells when the sensitivity is low (Garway-Heath et al 2000a). Intuitively, this would seem to indicate that the noise should be of a log-normal distribution, if the errors in measuring (albeit indirectly) the cell numbers were normally distributed. Yet this is also not the case; the testing strategies employed by automated perimeters (including, but not limited to, the Full Threshold strategy described in Chapter 1.2) are based on the scale being linear, and the strategy inevitably effects the distribution of the noise. Therefore, a new improved model for decibel noise which would be applicable at all sensitivity levels is desirable.

7.1 Using The Filter To Estimate Noise

Previously, noise has been estimated by looking at test-retest variability. An eye is tested twice within a short period of time (typically a week), and the visual fields compared. This is too short a time period for any significant change to have occurred, and so the difference between the two fields must be due to noise. There are two major problems with this approach:

- Firstly, it is not known what the true physiological sensitivity of each point is; it may be that, for example, the first and second readings are both too high (because of the inherent noise). This means that getting an accurate profile of the noise when the actual sensitivity is, say, 20dB is

impossible; instead, it is only possible to obtain a profile of the noise when the first noisy reading is 20dB.

- Secondly, the test-retest variability is taken as being the difference between the first and second sensitivity readings. However, since both readings are noisy, this measure in fact gives the difference between the noise components of the two readings; and so it gives, in terms of the variance, double the actual noise of each individual reading.

An ideal measurement of the noise present would look at the difference between the actual physiological sensitivity and the threshold estimate produced from the visual field test. Of course, there is currently no way of knowing the actual sensitivity; but the filter developed in Chapter 5 is designed to come closer than any current measure. It would be over-optimistic to expect any spatial filter to remove all the noise from a visual field reading, but it is feasible (and backed up by the testing in the previous chapter) that it removes the majority of the noise. Therefore, it is proposed that the difference between the raw and filtered sensitivities at each point be used as a measure of the noise at that point. This would be expected to slightly under-estimate the noise, not least because the filtered value is based in part on the raw value. However, it alleviates the two problems mentioned above, and should give a far more accurate picture of the distribution of the noise at different sensitivity levels. Crabb et al (1995) considered this idea using a simple median filter, where the differences between raw and filtered values were summarized in a measure called the local spatial variability; this measure was found to be closely correlated with short-term fluctuation measures made by the HFA, in groups of normal and glaucomatous eyes.

Moreover, because only one visual field is required to obtain an estimate of the noise (rather than two fields separated by a sufficiently short time interval for test-retest noise estimation), a much larger sample size can be used when deriving the estimated noise distributions; in fact the entire database of visual fields could be used. However in this case, because some patients have more visual field readings present in the database (described in Section 5.2) than others, and because the noise profile of individual patients could vary, it was chosen to use only the last eight visual fields from each eye from each patient (eyes which did not have at least eight complete visual fields were excluded); this way, no individual patient will have an undue effect on the results. This is particularly important because the patients who have very long series of visual fields in the database are typically those who have deteriorating glaucoma, and so their eyes may exhibit higher levels of noise than stable eyes; including all fields from each patient would therefore bias the results. The final eight fields for each patient were used instead of just one field per patient, because otherwise there would be too few fields left with moderate to advanced damage, therefore making it very difficult to obtain an accurate picture of the noise at lower sensitivities. This still left 15608 visual fields from the database, taken from 1951 different eyes.

7.2 Estimates Of The Noise

Each visual field in the database was considered in turn. The field was first filtered using the spatial filter described in the previous two chapters. Then, the estimated noise at each point was calculated as the

difference between the raw and filtered sensitivities at that point. These estimated values for the noise were separated according to the filtered value, rounded to the nearest half decibel; so, for example, one column of the output contained the noise estimates for all points whose filtered value was between 21.25 (inclusive) and 21.75dB (exclusive).

Following this procedure, the distribution of the noise estimates (raw value minus filtered value) can be shown for different rounded filtered sensitivity values; three examples are given in Figure 7.1.

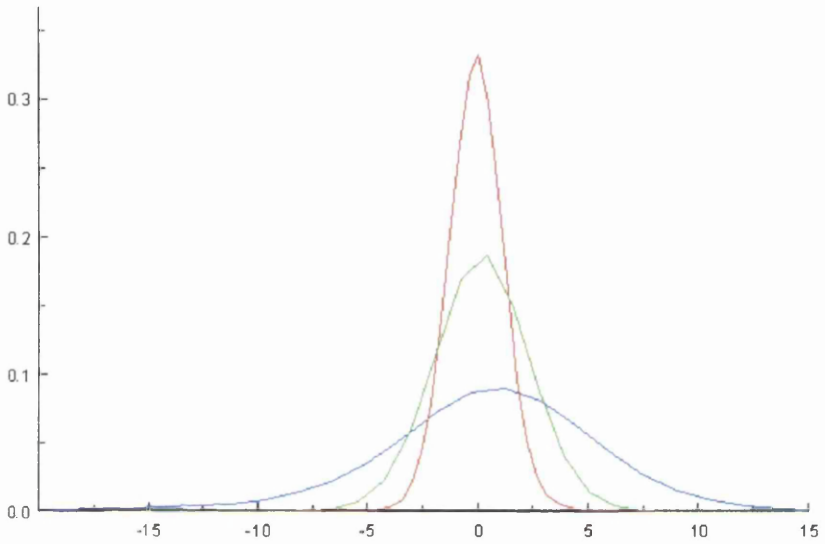


Figure 7.1: The empirical distribution of the noise estimates for different filtered sensitivity values. The red line shows the pdf for the noise estimates when the filtered sensitivity is between 32.25 and 32.75dB; the green line is for filtered sensitivities between 26.75 and 27.25dB; and the blue line is for filtered sensitivities between 16.75 and 17.25dB.

It can be seen that as the filtered sensitivity decreases, the distribution not only becomes more spread out (indicating a higher level of noise, as would be expected from Henson et al (2000) and others, and discussed in Section 2.1); but also the distribution becomes negatively skewed (the left tail is longer than the right tail, whilst the mode of the distribution is to the right of the centre). A similar graph of the pdf when the filtered sensitivity is even lower, say 5dB, would exhibit a markedly negatively-skewed distribution, but with a large peak at -5dB corresponding to the censoring of the raw values (which cannot be below 0dB). These qualitative findings are in keeping with published work based on test-retest variability (Heijl et al 1989; Heijl et al 1991; Chauhan & House 1991; Chauhan & Johnson 1999b; Henson et al 2000; Artes et al 2002), and so support the idea of using the filter to derive noise estimates. As predicted above, the method described here produces lower estimates of the amount of noise than the estimates based on test-retest variability, because it is estimating the amount of noise for just one reading rather than the difference between the amounts of noise at two consecutive readings.

The non-normality of the noise distribution could have major consequences. Firstly, if a better model can be found, it would clearly improve computer simulation techniques such as the Virtual Eye, making them truer to life. This would strengthen the arguments discussed in Chapter 2.1 for using simulated data rather than actual patient data when evaluating testing strategies. Secondly, if the noise distribution deviates significantly from a normal distribution, this has far-reaching implications for the diagnostic tools

analysing visual field data; the validity of pointwise linear regression relies on the implicit assumption of normality, since the least-squares regression assumes normal errors. If a mathematical model for the noise can be derived, then the extent of the deviation from normality can potentially be assessed.

7.3 Pearson Densities

Karl Pearson showed that given the first four sample moments, a continuous density function can be constructed that is consistent with those four moments (Rose & Smith 2002). Pearson density functions are the family of solutions $p(x)$ to the differential equation:

$$\frac{d}{dx}p(x) = -\frac{a+x}{c_0 + c_1 x + c_2 x^2}p(x)$$

Equation 7.3.1

that yield well-defined density functions (Johnson & Kotz 1969). The shape of the resulting distribution will depend on the Pearson parameters (a, c_0, c_1, c_2). These parameters can be expressed in terms of the first four moments of the distribution.

Given a random sample (X_1, X_2, ..., X_n) of size n, the rth sample moment is given by:

$$\hat{m}_r = \frac{1}{n}\sum_{i=1}^{n} X_i^r$$

and the rth sample central moment is given by:

$$m_r = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^r$$

where $\overline{X} = \hat{m}_1$ is the sample mean. Then, m_2 is the sample variance; $\beta_1 = \frac{m_3^2}{m_2^2}$ is the square of the sample skewness (a standard measure of the symmetry of the distribution); and $\beta_2 = \frac{m_4}{m_2^2}$ is often used as a measure of the sample kurtosis (a measure of the peakedness and tail weight of the distribution).

The Pearson parameters (a, c_0, c_1, c_2) in Equation 7.3.1 can be expressed in terms of these moments. The calculations are greatly simplified by choosing to work about the mean (i.e. subtracting the mean from each value in the sample); this yields a distribution with the same second, third and fourth moments as the empirical sample distribution, but a different mean value. This is easily corrected by adding a constant at the end to correct the mean value, so that all four moments are correct. Because this method is simpler, and so has less chance of rounding or calculation errors, it is this method which will be used. The formulae are then:

$$a = -\frac{m_3\left(3m_2^2 + m_4\right)}{2\left(9m_2^3 + 6m_3^2 - 5m_2m_4\right)}$$

$$c_0 = \frac{m_2\left(3m_3^2 - 4m_2m_4\right)}{2\left(9m_2^3 + 6m_3^2 - 5m_2m_4\right)}$$

$$c_1 = -\frac{m_3\left(3m_2^2 + m_4\right)}{2\left(9m_2^3 + 6m_3^2 - 5m_2m_4\right)}$$

$$c_2 = \frac{6m_2^3 + 3m_3^2 - 2m_2m_4}{2\left(9m_2^3 + 6m_3^2 - 5m_2m_4\right)}$$

Equations 7.3.2

There are in fact seven different types of Pearson density, depending on the values of β_1 and β_2 (or equivalently on the values of m_2, m_3 and m_4). In this case, as will be seen in the next section, the kurtosis is large (average 5.3), consistent with a peaked distribution with long tails; and the skewness is quite small (in the range [-1,+1]) and so the data is fitted by a Type IV Pearson distribution. This type occurs when the quadratic expression $(c_0 + c_1x + c_2x^2)$ does not have real roots, and it has the solution:

$$p(x) = e^{\frac{(c_1 - 2ac_2)\,ArcTan\left(\frac{c_1 + 2c_2x}{\sqrt{-c_1^2 + 4c_0c_2}}\right)}{c_2\sqrt{-c_1^2 + 4c_0c_2}}} \left(c_0 + x(c_1 + c_2x)\right)^{\frac{1}{2c_2}} C[S]$$

Equation 7.3.3

over the range (-∞, ∞); where numerical integration is required to find the constant of integration $C[S]$, for example by use of the trapezium rule.

Pearson densities, and a computer program called mathStatica (Wolfram Research Inc., Illinois, USA) capable of calculating them, are described fully by Rose & Smith (2002).

7.4 Modelling The Noise

The sample variance, skewness, etc of the noise distribution vary as the sensitivity varies. If these can be modelled, then an estimated Pearson distribution for the noise can be derived from them.

As described in Section 7.2, sample noise distributions were found for each interval of filtered sensitivities from [-0.25, 0.25) to [32.25, 32.75). For each of these intervals, the sample mean, variance, skewness and kurtosis were calculated. These were then plotted against sensitivity (where the sensitivity in the middle of each interval was used for the plot, i.e. 17, 17.5, etc). This produced the graphs in Figure 7.2.

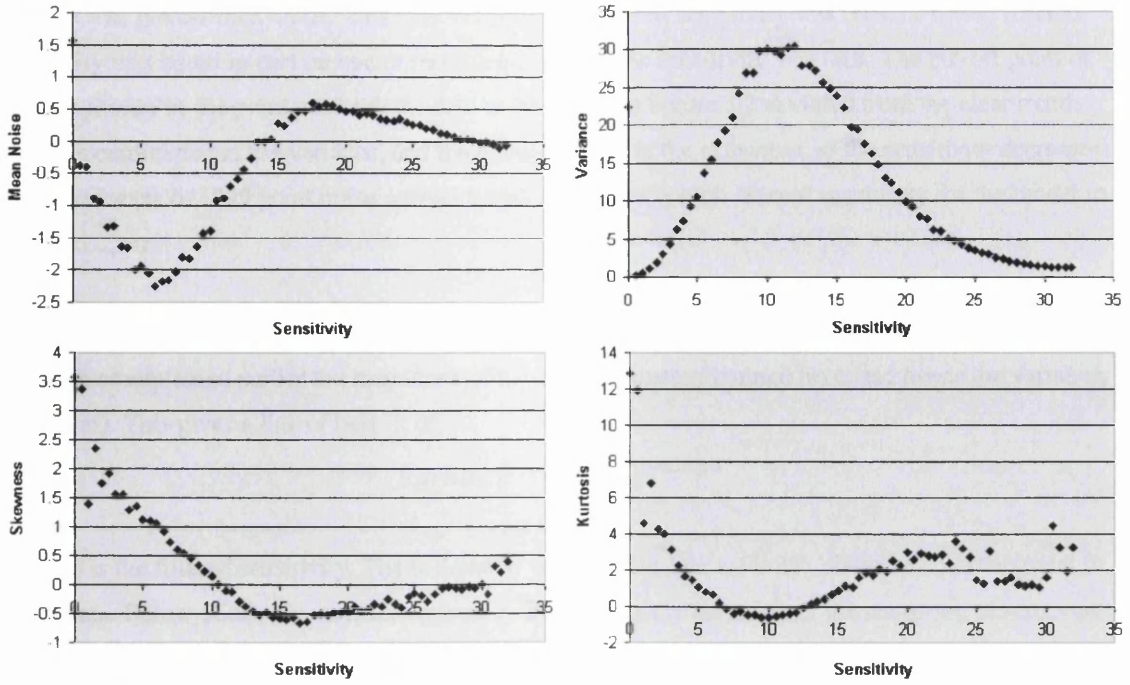


Figure 7.2: Graphs of the four distribution measures.

A major problem is the censoring nature of the perimeter. A sensitivity reading of 0dB actually means 'less than or equal to 0dB'. This is why in graphs of test-retest variability, there is always a peak corresponding with the retest sensitivity reading being zero. When the initial sensitivity is low (say 10dB), the probability of the retest sensitivity being below zero becomes highly significant. For example, Figure 7.3 shows the distribution of the noise when the rounded filtered sensitivity is 9.5dB (with 9.5dB added to each reading to give an indication of the scale of the distribution of the raw sensitivities).

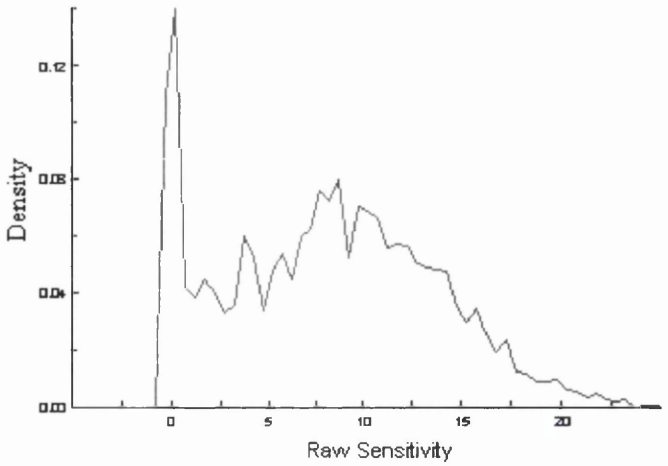


Figure 7.3: Empirical pdf of the raw sensitivity when the filtered sensitivity is between 9.25 and 9.75dB, showing the peak corresponding with the censoring effect at zero.

This makes modelling the noise distribution at low sensitivities unfeasible. The way to get around this problem is to model the noise distribution at higher sensitivities, and then extrapolate these modelled distributions to obtain a prediction of the distribution at lower sensitivities; before censoring these predictions and checking the fit against the empirical distributions. Therefore, the model will be based only on those noise estimates derived from points whose filtered sensitivity (rounded to the nearest half

decibel) was greater than 16dB. This removes points whose raw sensitivity was 0dB, or whose filtered sensitivity was based in part on one or more points whose raw sensitivity was 0dB. The cut-off point of 16dB is chosen as the point at which the data in the graphs in Figure 7.2 deviated from the clear trends (the exponential rise in the variance, and the linear decrease in the skewness, as the sensitivity decreases). This still leaves 644099 point noise estimates with a sufficiently high filtered sensitivity for the model to be based upon.

For the variance, there is a clear exponential curve above 16dB (agreeing with Henson et al (2000), although as explained earlier the magnitude of the noise estimates obtained here, and hence the variance, is smaller). This gives a line of best fit of:

$$Variance = 448.87\, e^{-0.1920\, S}$$

Equation 7.4.1

where S is the filtered sensitivity. The R-Squared value for this line is 99.3%, showing an exceptional fit to the data. Below 16dB, the variance appears to decrease; this is the result of the censoring at zero, since the variability of points below zero has been removed, hence reducing the sample variance.

Similarly, the positive skewness at lower sensitivities is caused by removing the left-hand tail of the noise distribution and replacing it with an artificially high density at zero; the censoring effect will clearly make the skewness of the data in Figure 7.3 higher than it would otherwise have been. For the higher sensitivities, there is again a clear pattern in the skewnesses above 16dB; in this case, a linear fit is appropriate, with a very good fit indicated by an R-Squared value of 87.8%:

$$Skewness = 0.0519S - 1.5041$$

Equation 7.4.2

The kurtosis does not show any significant trend as the sensitivity changes (amongst sufficiently high sensitivities as before); instead it hovers between four and eight. Due to this lack of a trend, rather than fitting a line of best fit (which would have a very poor R-Squared value) it was chosen to use the mean value for the kurtosis; so it will henceforth be assumed that the kurtosis is constant at 5.305 (to 3 decimal places). Thankfully the kurtosis has far less effect on the shape of the resulting Pearson distribution than the variance or skewness, so this assumption is not too critical.

These modelled estimates for the skewness and kurtosis ensure that a Pearson density will be of Type IV (as described in the Chapter 7.3) whenever the true sensitivity is greater than 8dB, which is sufficient for our purposes. The model for the noise developed here will not be well-defined at lower sensitivities; but it will still be an improvement on the normally-distributed noise currently used, which is only a valid model down to at best 20dB (and, more conservatively, down to just 25dB). Clinically – and hence in any simulation experiments – points whose sensitivity is below 20dB are still of interest; yet points whose sensitivity has deteriorated below 8dB are so badly damaged that their future status becomes less important.

Modelling the mean is harder. The mean is less affected by the censoring at zero than the variance or skewness. Yet a linear model for the mean only fits the data down to a sensitivity of around 18dB, a

higher sensitivity than would be achieved if the non-linearity of the data was purely caused by the censoring at zero. So a non-linear model is required which will fit the data down to 8dB (the lower limit of the model). Now, the values for the mean noise shown in the top left graph in Figure 7.2 seem to oscillate about zero as the sensitivity increases from 8dB towards 35dB. Further, if it were possible to continue the graph further right towards infinity, the mean noise would be expected to asymptotically tend towards zero (rather than ∞ or -∞), because the variance of the noise would reduce to zero. Therefore a damped oscillation model suggests itself; the best fit is given by:

$$Mean = -6.5414\, e^{-0.1154(S+1.4866)} \sin\left(0.2034(S+1.4866)\right)$$

Equation 7.4.3

The goodness of fit of this model can be illustrated by plotting modelled values for the mean noise and comparing with the empirical estimates, as in Figure 7.4.

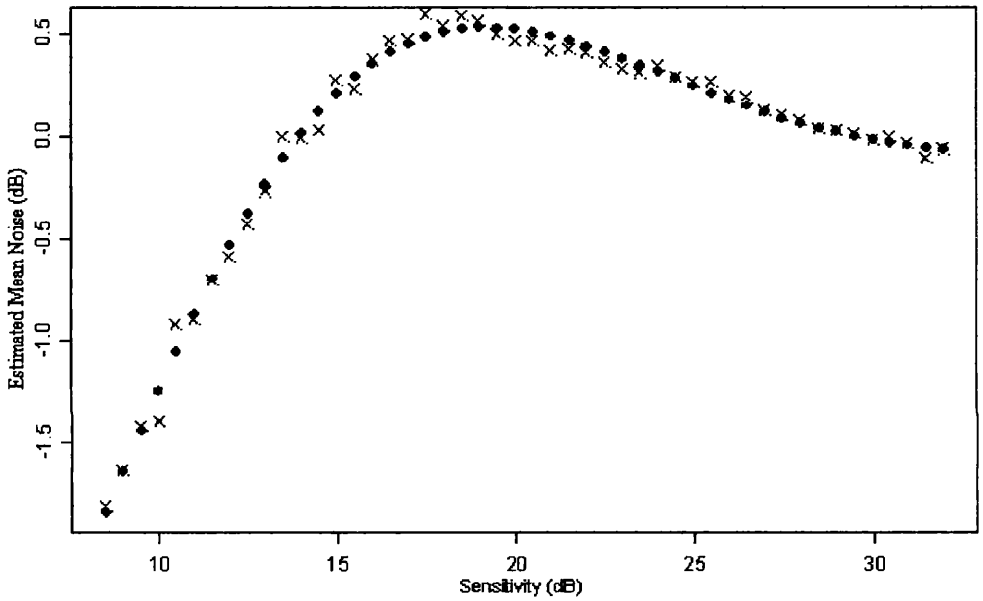


Figure 7.4: Comparing the empirical estimates of the mean noise (shown by crosses) with estimates calculated from the model (shown by dots) at different sensitivities.

Intuitively it would seem that the noise distribution should have a mean of zero for all sensitivities. Yet this is clearly not the case. The reasons for this are explained below in Chapter 7.4.1.

Now, having predicted these four parameters of the noise distribution for a given input sensitivity (from Equations 7.4.1-3, and the given constant value for the kurtosis), the moments of the noise distribution can be calculated as follows:

$$m_1(S) = Mean$$

$$m_2(S) = Variance$$

$$m_3(S) = \sqrt{\left(Skewness^2 \times Variance^3\right)}$$

$$m_4(S) = Kurtosis \times Variance^2$$

Equations 7.4.4

Next, going back to Equations 7.3.2, the Pearson parameters can be easily calculated. These parameters can then be put into Equation 7.3.3 to find distribution $p_S(x)$ at sensitivity S, using numerical integration to find the constant of integration and ensure a well-defined distribution function.

The mean of this distribution can be calculated from $\mu_S = \int x p_S(x) dx$ (estimated numerically by use of, for example, the trapezium rule). The distribution must be shifted along the axis so that the mean value for the distribution becomes equal to the desired mean value for the noisy sensitivity readings, namely $(S + m_1(S))$. The distribution for the noisy sensitivity readings is therefore given by $q_S(x) = p_S(x - S - m_1(S) + \mu_S)$, where x is the input. Now, if the output sensitivity is less than zero, then because the perimeter censors the data (it cannot measure sensitivities below 0dB), these sensitivities must be set equal to zero. So the final distribution for the simulated output sensitivity readings is given by:

$$\begin{cases} x > 0: & f_S(x) = p_S(x - S - m_1(S) + \mu_S) \\ x = 0: & f_S(x) = 1 - \int_{x>0} p_S(x - S - m_1(S) + \mu_S). \\ x < 0: & f_S(x) = 0 \end{cases}$$

Equation 7.4.5

In summary, the steps for simulating the distribution of noisy sensitivity readings, given that the input noise-free sensitivity is SdB, are:

Step 1. Calculate the predicted variance and skewness of the distribution using Equations 7.4.1 and 7.4.2.

Step 2. Convert these, and the permanent estimate of 5.305 for the kurtosis, into the second, third and fourth moments of the distribution using Equations 7.4.4.

Step 3. Use these estimated moments to derive estimates for the Pearson coefficients, using Equations 7.3.2.

Step 4. Insert these coefficients into Equation 7.3.3 to produce the distribution $p_S(x)$.

Step 5. Use the trapezium rule to estimate the constant of integration $C[S]$ (converting $p_S(x)$ into a well-defined distribution) and the mean μ_S of the distribution.

Step 6. Calculate the desired mean of the distribution from Equation 7.4.3.

Step 7. The modelled noisy sensitivity distribution, including censoring, is then given by $f_S(x)$ as defined in Equation 7.4.5.

7.4.1 The sample mean of noise estimates

This sub-section attempts to explain why the sample mean noise counter-intuitively changes with sensitivity. It is clear from Figure 7.2 that as the filtered sensitivity decreases, the mean sample noise initially increases; that is, the raw sensitivity becomes increasingly likely to be higher than the filtered sensitivity. On further reduction of the filtered sensitivity, the mean noise then becomes negative; the raw sensitivity is now likely to be lower than the filtered sensitivity. This is not a characteristic of the filter; rather, it is a product of the choice to categorise points according to their filtered value.

Indeed, when the same analysis is carried out using the Gaussian filter (which is known to blur out low sensitivity points, as shown by Spry et al (2002)), the same findings occur; a medium sensitivity Gaussian-filtered value (around 20dB) is likely to have come from a point whose raw sensitivity was slightly higher. However when the noise is categorised according to the raw sensitivity of the point, rather than the filtered value, the opposite occurs; as the raw sensitivity decreases from 30dB towards 20dB, the mean sample noise decreases, indicating that the Gaussian filtered sensitivity becomes increasingly likely to be higher than the raw sensitivity. How can this be?

What the results are saying is that when the filtered sensitivity (using either the Gaussian filter or the new filter developed in this work) is 20dB, it is more likely that the raw sensitivity at that point was 22dB than that it was 18dB. When the raw sensitivity is 20dB, it is more likely that the filtered sensitivity will be 22dB than that it will be 18dB.

To explain this phenomenon, Bayes' Law must be taken into consideration. Over the entire database, a sensitivity of 22dB is more common than one of 18dB. Thus, the probability that the raw sensitivity is 22dB (written below as $P(R=22)$) is larger than the probability that the raw sensitivity is 18dB (written as $P(R=18)$). Therefore, writing R for the raw sensitivity and F for the filtered sensitivity:

$$P(R=22) > P(R=18)$$

$$P(R=22)P(F=20 \textit{ given } R=22) > P(R=18)P(F=20 \textit{ given } R=18)$$

$$\frac{P(R=22)P(F=20 \textit{ given } R=22)}{\int P(R=R')P(F=20 \textit{ given } R=R')dR'} > \frac{P(R=18)P(F=20 \textit{ given } R=18)}{\int P(R=R')P(F=20 \textit{ given } R=R')dR'}$$

$$P(R=22 \textit{ given } F=20) > P(R=18 \textit{ given } F=20)$$

(note that $P(F=20 \textit{ given } R=22)$ is not identical to $P(F=20 \textit{ given } R=18)$ because of the skewness in the noise distribution, but they are similar enough for this effect not to outweigh the inequality)

Similarly, over the entire database, a filtered sensitivity of 22dB is more common than one of 18dB. So by the same logic:

$$P(F=22) > P(F=18)$$

$$P(F=22)P(R=20 \textit{ given } F=22) > P(F=18)P(R=20 \textit{ given } F=18)$$

$$P(F=22 \textit{ given } R=20) > P(F=18 \textit{ given } R=20)$$

This is why, whatever the filtered sensitivity is, the raw sensitivity at that point is (on average) likely to be closer to the centre of the distribution of raw sensitivities. Therefore, despite initially appearing curious, the change in the sample mean of the noise estimates as the sensitivity changes is not a cause for alarm. It is only correct that this effect should be taken into account when modelling noisy sensitivities, hence the adjustment to the desired mean, caused by Equation 7.4.3, which appears in Equation 7.4.5.

7.5 Examples

Three examples of using this new model for the noise will be given; when the actual sensitivity is 30dB (healthy), 20dB (damaged) and 10dB (severely damaged). The equivalent values for a standard normal distribution (N(0,1)) are also given for reference. Fitting these sensitivity values into Equations 7.4.1 and 7.4.2 (Step 1) and Equation 7.4.3 (for later use in Step 6) gives the following estimates for the characteristics of the noise distributions (to four decimal places):

Sensitivity	30dB	20dB	10dB	Normal
Mean	-0.0209	0.5163	-1.2529	0
Variance	1.4144	9.6478	65.8075	1
Skewness	0.0529	-0.4661	-0.9851	0
Kurtosis	5.305	5.305	5.305	3

Table 7.1: Estimated distribution characteristics for the three example sensitivities of 30dB, 20dB and 10dB.

Now, using Equations 7.4.4 (Step 2), values can be estimated for the second, third and fourth moments of the noise distributions (again, to four decimal places):

Sensitivity	30dB	20dB	10dB	Normal
m_2	1.4144	9.6478	65.8075	1
m_3	0.0890	-13.9677	-525.8878	0
m_4	10.6134	493.7930	22973.9536	3

Table 7.2: Estimated distribution moments for the three example sensitivities.

From these values, the Pearson coefficients are calculated (Step 3) according to Equations 7.3.2:

Sensitivity	30dB	20dB	10dB	Normal
a	0.0149	-0.3706	-2.8356	0
c_0	0.8568	6.1165	51.4785	1
c_1	0.0149	-0.3706	-2.8356	0
c_2	0.1314	0.1220	0.0726	0

Table 7.3: Estimated Pearson coefficients for the three example sensitivities.

Clearly the main difference between these noise distributions and a standard normal distribution is the increase in variance which occurs when the sensitivity decreases; however there is also a noticeable change in the skewness and related moments and coefficients. Next, using Equation 7.3.3, we can derive the estimated density functions (Step 4). So the densities are:

$$p_{30}(x) = e^{0.1248\,ArcTan(0.0222+0.3917x)}\left(0.8568+0.0149x+0.1314x^2\right)^{-3.8048} C[30]$$
$$p_{20}(x) = e^{-1.3608\,ArcTan(-0.2196+0.1446x)}\left(6.1165-0.3706x+0.1220x^2\right)^{-4.0981} C[20]$$
$$p_{10}(x) = e^{-12.7102\,ArcTan(-1.0792+0.0552x)}\left(51.4785-2.8356x+0.0726x^2\right)^{-6.8889} C[10]$$

Equations 7.5.1

(note that the constants of integration, $C[10]$, $C[20]$ and $C[30]$, will not be the same). Next, the constants of integration and the means of these distributions must be found; this is done by evaluating $p_S(x)$ at x=40, 39.5, 39, ... and then using the trapezium rule (in this case with step size of 0.5, although obviously other step sizes could be used) to calculate $C[S]^{-1}=\Sigma p_S(x)/2$ (so that the constant of integration will normalise the distribution), and $\mu_S=\Sigma x p_S(x)/2$. This gives:

Sensitivity	30dB	20dB	10dB
$C[S]^{-1}$	4.6673	0.00551	1.088×10^{-6}
μ_S	-3.913×10^{-7}	0.00107	0.2861

Table 7.4: Estimated distributional constants obtained from numerical integration for the three example sensitivities.

Finally, Step 7 uses Equation 7.4.5 to give the final distribution of the simulated noisy sensitivities for each of the three input sensitivities.

These final distributions can be compared with the respective empirical distributions, as shown graphically in Figure 7.5. There are two points which must be taken into consideration when carrying this out:

1. As described in Section 7.4, the empirical distributions contain all estimates of the noise when the filtered sensitivity is between (for example) 19.75 and 20.25dB. To show accurately the distribution of noise, whilst still giving an appropriate indication of the scale, in the graphs below it is assumed that the filtered sensitivity is exactly 20dB in each case. The data for the empirical distribution of raw sensitivities has been adjusted accordingly.
2. Since the filtered sensitivity is only spread over a 0.5dB range, and the raw sensitivities are almost always integer decibel values (the only possible exception being when a point was measured twice during testing and the two readings averaged), the noise (raw-filtered) cannot take values in the whole range. For example, if the filtered sensitivity is from the range [29.75, 30.25) then the noise will always be in the range [z-0.25, z+0.25) where z is any integer. Therefore although the bars contain the ranges [0, 0.5); [0.5, 1); [1, 1.5) etc, in reality the noise estimates in those ranges will always be within [0, 0.25); [0.75, 1); [1, 1.25) etc.

For clarity and easier interpretation, a line graph has been used in Figure 7.5 for the empirical distributions rather than a histogram. In each case, the red line represents the empirical distribution of noisy sensitivities whose filtered sensitivity is within the given range (subject to the correction mentioned in point 1 above); the blue line represents the modelled pdf for noisy sensitivity values at the given input sensitivity; and the yellow line, included for reference, represents the pdf of a normal distribution of the same variance as the model, with a mean equal to the input sensitivity (and thus equal to the pdf of noisy sensitivity values which would be obtained from using the currently-used model based on noise being normally distributed).

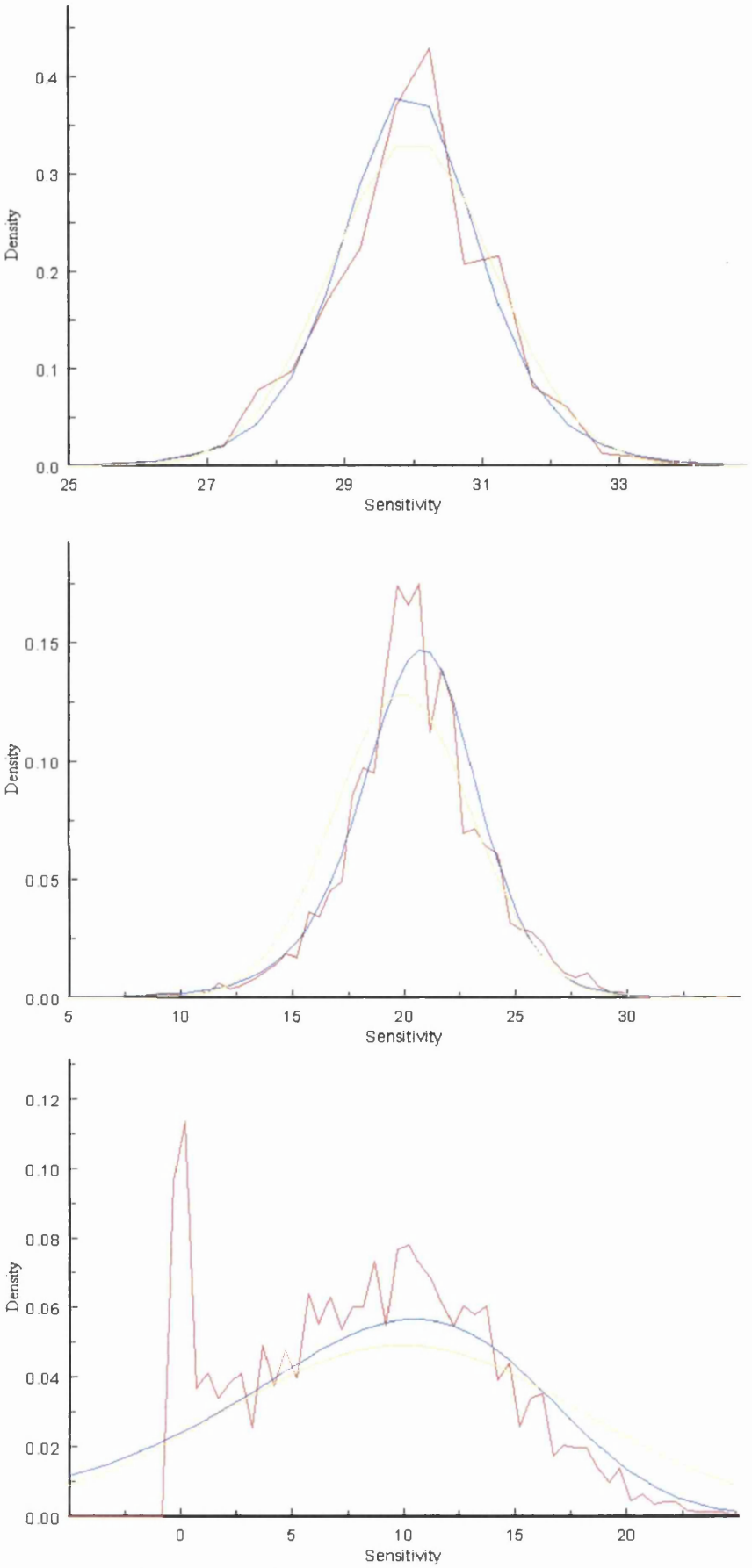


Figure 7.5: Comparing the pdf of the modelled distribution of noisy sensitivities (blue line) to the empirical pdf (red line), with the normal distribution (yellow line) for reference. The top graph shows this for an input sensitivity of 30dB; the middle graph for 20dB; and the bottom graph for 10dB. Note that the scales on the axes differ for each graph.

In each of the three examples illustrated in Figure 7.5, the model appears to fit the empirical distribution well. Importantly, this is still the case at 10dB, where the model is based on extrapolated data (recall that the models for the variance and skewness are based on data at sensitivities of 16dB and above). This supports the decision to extrapolate the trends downwards to remove the effect of the censoring.

7.6 Testing The Model

From a visual inspection of the three graphs in Figure 7.5, the fit in each case appears to be better with the model developed in this chapter (shown by the blue line in each case) than with the alternative normally distributed model (shown by the yellow lines). However a more formal test of goodness-of-fit is required. One method is to examine the Pearson chi-squared statistics.

Given an input sensitivity S, write O_i for the empirical number of points whose filtered sensitivity is S (when rounded to the nearest 0.5dB as before) and whose (rounded) raw sensitivity is i. Write M_i for the number of points expected to have a (rounded) noisy sensitivity of i according to the model developed here, when the input sensitivity is S; and similarly N_i for the number expected according to the alternative model of normally-distributed noise. Then the two test statistics can be calculated:

$$X_M^2 = \sum_i \frac{(O_i - M_i)^2}{M_i}$$

$$X_N^2 = \sum_i \frac{(O_i - N_i)^2}{N_i}$$

Equations 7.6.1

where the sums are carried out over all sensitivities i for which M_i and N_i respectively are non-zero (or for computational purposes, where the pdf using each model gives a density greater than 0.001). A larger value for the test statistic indicates a worse fit to the empirical data. This can be done for every input sensitivity S, and so the two models compared over the full range of sensitivities, as in Figure 7.6. Note that to interpret the test statistics fairly, it is necessary in each case to divide the statistics by the number of points falling into that category (i.e. the number of points whose filtered sensitivity is in the range $[S\text{-}0.25, S\text{+}0.25)$); this corrects the statistics so that they are all the same magnitude, without affecting the comparison between the two models. In effect it is altering the statistics so that the empirical and modelled pdfs are being compared, rather than the actual number of points.

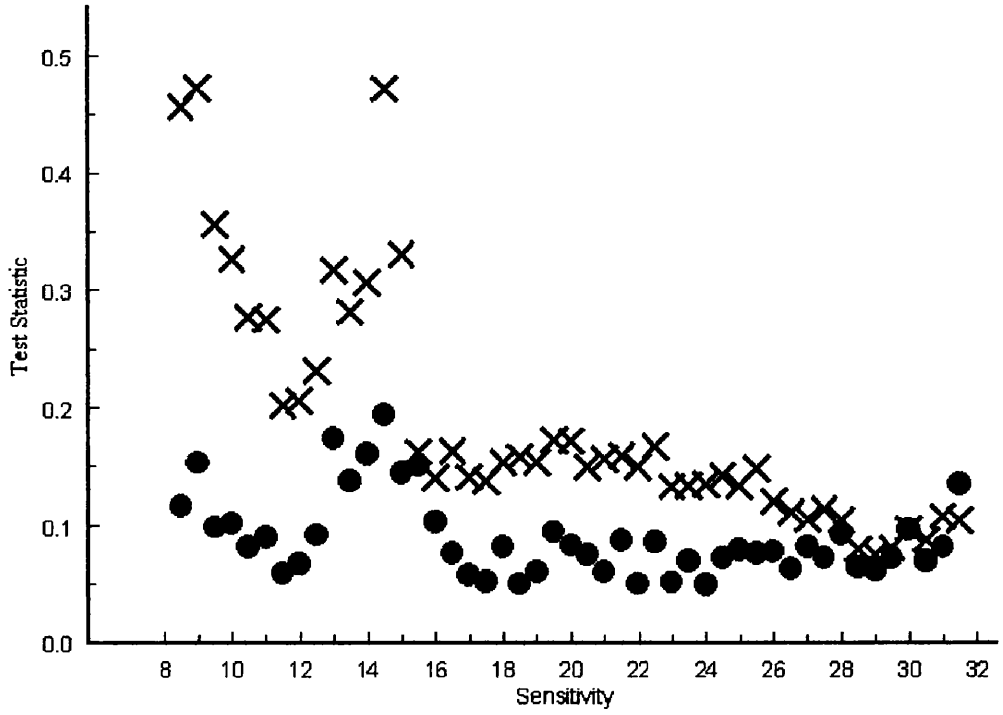


Figure 7.6: Comparing the test statistics (as defined in Equation 7.6.1) for the normally-distributed noise model (crosses) and the model derived in this chapter (circles).

It is clear to see that the normally-distributed noise model does not fit the empirical data as well as the model derived in this chapter. The difference is even more marked at lower sensitivities, where the fact that the model derived in this chapter is taking into account the skewness becomes more significant. A simple paired t-test on the chi-squared statistics confirms that the test statistics are much lower using the newly-derived model (with a p-value of less than 0.001).

The normal distribution model performs reasonably well at higher sensitivities; however it can be concluded that the new model fits the empirical distribution of noisy sensitivities better than the normal distribution and over a wider range of sensitivities, and so can improve upon the currently-used simulation models. Whether the improvement justifies changing the simulation models is less clear, largely because of the difficulties with sampling from the new distributions; this can be done with some software packages such as mathStatica (Wolfram Research Inc) (Rose & Smith 2002) which can generate Monte-Carlo random samples from any arbitrary distribution. The process will become more feasible over the coming years as the speed of computers increases and the software available improves.

8. Conclusions

Glaucoma is a leading cause of blindness and visual disability. The advent of automated perimetry has offered the prospect of a more objective analysis of visual field loss in glaucoma, and currently provides the most reliable strategy for the clinical follow-up of the disease. However, because of the large amount of variability present in perimetric readings, it is often difficult to determine whether a localized glaucomatous defect is present in an eye, and whether such defects are progressing. Answering these two questions reliably is crucial for making patient management decisions.

One way of determining whether a diseased eye is progressing is by Pointwise Linear Progression (PLR); whence each point in the perimetric visual field is considered individually and its change over time examined. While this method is not universally accepted as the best way to follow eyes over time, largely because it does not take into account the spreading of defects to neighbouring points, it is non-controversially useful because of its objectivity and because (unlike some other methods) it considers all the available data from a series of tests.

PLR is a relatively new method, and as such its development and refinement is ongoing. In Chapter 3, a 'Virtual Eye' simulation model was used to examine the optimum frequency of testing which will produce reliable results, in terms of the sensitivity and specificity of distinguishing progressing eyes from stable eyes. The conclusion was suggested that testing eyes three times per year provided the best compromise between sensitivity and specificity for points progressing at 2dB/year; for other rates of deterioration more or less frequent test intervals may be optimum.

In Chapter 4, the same Virtual Eye simulation model and theoretical modeling were both used to examine the use of confirmation fields; that is, whether a point which appears to be progressing should be re-tested at a later date before progression can reliably be diagnosed. The work showed that confirming progression was indeed desirable, as otherwise the specificity was unacceptable. It was also proposed that this process could be further improved by omitting from the confirmation testing the test result which had first suggested progression.

These two experiments combine to give a prescription for carrying out perimetric visual field testing which, if followed, would improve the ability to distinguish between progressing and stable points in an eye. Given that a diagnosis of progression may result in treatment being applied, possibly including surgery, such an improvement is crucial.

In Chapter 5, a data-based spatial filter was developed with the aim of reducing the amount of noise present in visual field tests. Successfully reducing the variability will aid both identifying localized glaucomatous defects, and also determining whether they are progressing or not. Testing of the filter in Chapters 5 and 6 showed very promising results in this regard; indeed it is realistic to hope that the filter presented here may be utilised clinically in the medium term.

Chapter 7 used the filter to produce a statistical model for the noise present when estimating the sensitivity of a point in the visual field by automated perimetry. This model could replace the existing assumption of normality present when simulating visual field data, as in the Virtual Eye, although there are practical problems with carrying this out and the benefits are not overwhelming.

The key contributions to work in this field are:

- Producing a simulation model which can be used as a test-bed for examining visual field testing without requiring large amounts of patient data, and whilst knowing the actual noise-free status of the eye;
- Examining the methodology for using PLR to analyse visual field data, and suggesting improvements in terms of the frequency of testing and the use of confirmation fields;
- Producing a spatial filter which, from testing carried out to date, appear to dramatically reduce the noise present in visual field readings, hence making glaucomatous defects easier both to identify and to follow over time.

Potential future work includes:

- Incorporating the model for the noise developed in Chapter 7 into the Virtual Eye simulation model. This would currently involve re-writing the Virtual Eye into another software package capable of sampling from the new distribution; in the future it is to be hoped that the process will become easier in other packages.
- Examining non-linear, non-polynomial models for the deterioration of pointwise sensitivity over time. For example, survival analysis techniques of proportional hazards models could be considered, or the use of logistic regression techniques to transform sensitivity values into a more amenable scale.
- Producing a model for the spread of glaucomatous defects to neighbouring points. One possibility could be to use the filter to spread a defect to related points in the RNFL, repeatedly at set time intervals.
- Testing the filter on a different database to the one used to derive the filter. This could be done by for example comparing the predictive power of PLR on series of filtered and unfiltered fields.

Appendix 1 – Virtual Eye Program

Below is the code for an SPlus program which uses the Virtual Eye (as described in Chapter 2) to simulate series of visual fields, which are then tested for progression using each of the seven criteria detailed in Section 4.1; this is the experiment described in Section 4.3.1. This gives results as in Section 4.3.2; each column in the results matrix 'Graphs' contains the number of simulated points flagged as progressing using one of the seven criteria.

```
#Detection of linear progression
#model


#Variable settings
##################

#Number of fields=================
#7 fields (7=3yrs; 9=4yrs; 11=5yrs)
flds<-15
tim<-c(seq(0,(flds[1]/2)-0.5,0.5))

#Settings
f1<-c(NA,NA,NA,NA,NA)
#set initial deviation
f1[1]<-0
#set Rate of loss per year (dB)
f1[2]<-0
#set change ch1 values to dummy
ch1<-c(0,0)
#set number of simulations
simx<-1000
#p-value
pvalue<-0.01
#vector for sensitivity values
hdat1<-rep(NA,flds[1])

#Results matrix.
#Column 1 for Standard Criteria
#Column 2 for Two out of Two
#Column 3 for Three out of Three
#Column 4 for Two out of Three
#Column 5 for Three out of Four
#Column 6 for Two omitting
```

```
#Column 7 for Three omitting
Graphs<-matrix(0,flds[1],7)
#####VIRTUAL EYE SIMULATION=======================

for (m in 1:simx[1])
{
    ###matrix for results (prog/stable)
    resx<-matrix(0,flds[1],5)

    #####Generate data###############
    ###set change variable
    ch1[1]<-0

    for (k in 1:flds[1])
    {
    ###simulate from normal dist mean/SD
    hdat1[k]<-round(rnorm(1,f1[1]+ch1[1],exp(3.27-(f1[1]+ch1[1]*0.081))))
    ###linear reduction ********
    ch1[1]<-ch1[1]+(f1[2]/2)
    }


    #####Carry out regressions necessary for each method###
    for(upto in 3:flds[1])
    {
        ###subset of values used for the fit
        subset<-c(rep(T,upto),rep(F,flds[1]-upto))

        ###linear regression analysis###########
        tester <- lm(hdat1 ~ tim, subset=subset)
        #calculate standard error (5 and 6 df for n=7)
        se1<-sqrt((sum(tester$resi * tester$resi)/(upto-2))/(var(tim) * (upto-
    1)))
        #calculate P-value (slope / serror)
        #NB. pt=function for calculating p-values from t distn df=n-2.
        p1<-2*(1-pt(abs(tester$coef[2]/se1[1]),(upto-2)))

        ###results matrix#######################
        resx[upto,1]<-round(tester$coef[2],2)
        resx[upto,2]<-round(p1,2)

        ###Code to detect significant changes (p=pvalue)###
        if (resx[upto,2] <= pvalue & resx[upto,1] >= 0.9) resx[upto,3]<-1

        #####Results using first 5 determination methods###
        if (resx[upto,3]==1) Graphs[upto,1]<-Graphs[upto,1]+1
```

```
   if (resx[upto,3]==1 && resx[upto-1,3]==1) Graphs[upto,2]<-
Graphs[upto,2]+1
   if (resx[upto,3]==1 && resx[upto-1,3]==1 && resx[upto-2,3]==1)
      Graphs[upto,3]<-Graphs[upto,3]+1
   if (resx[upto,3]==1 && (resx[upto-1,3]==1 || resx[upto-2,3]==1))
      Graphs[upto,4]<-Graphs[upto,4]+1
   if (resx[upto,3]==1 && ((resx[upto-1,3]==1 && (resx[upto-2,3]==1 ||
      resx[upto-3,3]==1)) || (resx[upto-2,3]==1 && resx[upto-3,3]==1)))
      Graphs[upto,5]<-Graphs[upto,5]+1



      #####Repeat, omitting penultimate point
   if (upto>3)
   {
      ###subset of values used for the fit
      subset[upto-1]<-F

      ###linear regression analysis, omitting point###
      tester <- lm(hdat1 ~ tim, subset=subset)
      #calculate standard error (5 and 6 df for n=7)
      se1<-sqrt((sum(tester$resi * tester$resi)/(upto-3))/(var(tim) *
(upto-2)))
      #calculate P-value (slope / serror)
      #NB. pt=function for calculating p-values from t distn df=n-2.
      #n=number of points used = upto-1.
      p1<-2*(1-pt(abs(tester$coef[2]/se1[1]),(upto-3)))

      ###results matrix#######################
      resx[upto,1]<-round(tester$coef[2],2)
      resx[upto,2]<-round(p1,2)

      ###Code to detect significant changes (p=pvalue)###
      if (resx[upto,2] <= pvalue & resx[upto,1] >= 0.9) resx[upto,4]<-1

      #####Results using 6th determination method###
      if (resx[upto,4]==1 & resx[upto-1,3]==1) Graphs[upto,6]<-
Graphs[upto,6]+1
   }



   #####Repeat again, omitting penultimate two points
   if (upto>4)
   {
      ###subset of values used for the fit
      subset[upto-2]<-F
```

```
        ###linear regression analysis, omitting points###
        tester <- lm(hdat1 ~ tim, subset=subset)
        #calculate standard error (5 and 6 df for n=7)
        se1<-sqrt((sum(tester$resi * tester$resi)/(upto-2))/(var(tim) *
    (upto-1)))
        #calculate P-value (slope / serror)
        #NB. pt=function for calculating p-values from t distn df=n-2.
        p1<-2*(1-pt(abs(tester$coef[2]/se1[1]),(upto-2)))

        ###results matrix#######################
        resx[upto,1]<-round(tester$coef[2],2)
        resx[upto,2]<-round(p1,2)

        ###Code to detect significant changes (p=pvalue)###
        if (resx[upto,2] <= pvalue & resx[upto,1] >= 0.9) resx[upto,5]<-1

        #####Results using 7th determination method###
        if (resx[upto,5]==1 && resx[upto-1,4]==1 && resx[upto-2,3]==1)
           Graphs[upto,7]<-Graphs[upto,7]+1
      }
    }


print(m)
###close m loop
}
################################
```

Appendix 2 –Progressor Code

Below is the new function written into the C++ code for the Progressor program to implement the new filter derived in Chapter 5. Because of its size, the matrix of values containing the filter (a 76x76 matrix called 'filter') has been omitted where shown below; one row of this matrix (row 12) has been shown as an example. This function filters the sensitivity values for one eye; a nearly-identical function in the corresponding program file filters sensitivity values for the other eye. Additionally, the programming for the user interface was altered, so that the new filter is utilised instead of the Gaussian filter. The program was then compiled and successfully tested; indeed it was used to produce Figures 5.3 and 5.4.

```
void CLeftView::OnFilterNew()
{
    WORD filter[76][76] ={ ###### Rows of matrix omitted ######

    {0,0,0,0,0,11,10,2,0,0,0,14,20,21,0,1,0,0,0,0,16,0,3,0,0,0,1,0,0,0,0,0,0,0,
    0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
    ,0,0,0,0},

    ###### Rows of matrix omitted ######};

    WORD total[76] =
    {0,0,0,0,0,100,98,100,98,0,0,101,99,100,98,101,99,0,0,99,103,100,102,100,10
    2,104,100,0,0,100,88,98,105,100,102,101,101,93,0,100,0,96,103,102,100,102,1
    02,94,0,100,101,99,100,100,102,102,101,0,0,101,102,98,102,100,100,0,0,100,1
    00,97,99,0,0,0,0,0};

    int fx;
    WORD divisor;
    DWORD filtered;

    int current_field, no_of_fields, pointno;

    WORD matrix[76];


    CProwin01Doc* pDoc = GetDocument();

    CStatusBar* pStatus = (CStatusBar*) AfxGetApp()->m_pMainWnd-
    >GetDescendantWindow(AFX_IDW_STATUS_BAR);

    if ( m_bRightEye )
```

```
    {no_of_fields = pDoc -> m_rightpointArray.GetSize() / 76;
    if (!no_of_fields)return;
        BeginWaitCursor();
        if (!(pDoc -> m_rightGaussianFiltered))
            {pStatus->SetPaneText(0,"Applying filter...",TRUE);
            pStatus->UpdateWindow();

            for (int i = 0; i < pDoc -> m_rightpointArray.GetSize(); i++)
                pDoc -> m_rightUndoArray.SetAtGrow(i, pDoc ->
m_rightpointArray[i]);

        for ( current_field = 0; current_field < no_of_fields;
current_field++ )
            {for ( pointno = 0; pointno < 76; pointno++)
               matrix[pointno] = MISSING_VALUE;

            for ( pointno = 0; pointno < 76; pointno++)
               { matrix[pointno] = pDoc -> m_rightpointArray.GetAt( pointno +
current_field * 76);}

            for ( pointno = 0; pointno < 76; pointno++)
                  {if ( matrix[pointno] == MISSING_VALUE )continue;
                  divisor = 0;
                  filtered = 0;
                  for ( fx = 0; fx < 76; fx++)
                        {if (matrix[fx] != MISSING_VALUE)
                           {filtered += matrix[fx] * filter[pointno][fx];
                           divisor += filter[pointno][fx];
                        }
                     }
                  if (pointno == 40) {filtered = 0;}
                     else if (divisor!=0) {filtered /= 100 * divisor /
total[pointno];}
                     else {filtered = MISSING_VALUE;};
                  pDoc -> m_rightpointArray.SetAt( pointno + current_field *
76, filtered);
            }
        }
        pDoc -> Regress();
        //pDoc -> SetModifiedFlag();
        pDoc -> m_rightGaussianFiltered = 1;
        pDoc -> UpdateAllViews(NULL,1L,NULL);

    }
    else
        {pStatus->SetPaneText(0,"Removing filter...",TRUE);
        pStatus->UpdateWindow();
```

```
        for (int i = 0; i < pDoc -> m_rightpointArray.GetSize(); i++)
            pDoc -> m_rightpointArray[i] =  pDoc -> m_rightUndoArray[i];

        pDoc -> m_rightUndoArray.RemoveAll();
        pDoc -> Regress();
        //pDoc -> SetModifiedFlag();
        pDoc -> m_rightGaussianFiltered = 0;
        pDoc -> UpdateAllViews(NULL,1L,NULL);
    }
}
else
    {no_of_fields = pDoc -> m_leftpointArray.GetSize() / 76;
    if (!no_of_fields)return;
        BeginWaitCursor();
        if (!(pDoc -> m_leftGaussianFiltered))
            {pStatus->SetPaneText(0,"Applying filter...",TRUE);
            pStatus->UpdateWindow();

            for (int i = 0; i < pDoc -> m_leftpointArray.GetSize(); i++)
                pDoc -> m_leftUndoArray.SetAtGrow(i, pDoc ->
m_leftpointArray[i]);

        for ( current_field = 0; current_field < no_of_fields;
current_field++ )
            {for ( pointno = 0; pointno < 76; pointno++)
               matrix[pointno] = MISSING_VALUE;

            for ( pointno = 0; pointno < 76; pointno++)
               { matrix[pointno] = pDoc -> m_leftpointArray.GetAt( pointno +
current_field * 76);
            }

            for ( pointno = 0; pointno < 76; pointno++)
                  {if ( matrix[pointno] == MISSING_VALUE )continue;
                  divisor = 0;
                  filtered = 0;
                  for ( fx = 0; fx < 76; fx++)
                        {if (matrix[fx] != MISSING_VALUE)
                           {filtered += matrix[fx] * filter[pointno][fx];
                           divisor += filter[pointno][fx];
                        }
                     }
                  if (pointno == 40) {filtered = 0;}
                     else if (divisor!=0) {filtered /= 100 * divisor /
total[pointno];}
                     else {filtered = MISSING_VALUE;};
```

```
                    pDoc -> m_leftpointArray.SetAt( pointno + current_field *
    76, filtered);
                }
            }
            pDoc -> Regress();
            //pDoc -> SetModifiedFlag();
            pDoc -> m_leftGaussianFiltered = 1;
            pDoc -> UpdateAllViews(NULL,2L,NULL);

        }
        else
            {pStatus->SetPaneText(0,"Removing filter...",TRUE);
            pStatus->UpdateWindow();

            for (int i = 0; i < pDoc -> m_leftpointArray.GetSize(); i++)
                pDoc -> m_leftpointArray[i] =  pDoc -> m_leftUndoArray[i];

            pDoc -> m_leftUndoArray.RemoveAll();
            pDoc -> Regress();
            //pDoc -> SetModifiedFlag();
            pDoc -> m_leftGaussianFiltered = 0;
            pDoc -> UpdateAllViews(NULL,2L,NULL);
        }
    }
    if (pDoc -> m_bProgIndDlgDisplayed) pDoc -> m_pProgIndDlg ->
    OnInitDialog();
    if ((m_bRightEye && pDoc -> m_rightGaussianFiltered) || (!m_bRightEye &&
    pDoc -> m_leftGaussianFiltered))
        pStatus->SetPaneText(0,"Filter applied. For Help, press F1",TRUE);
    else pStatus->SetPaneText(0,"Filter removed. For Help, press F1",TRUE);
    pStatus->UpdateWindow();
    EndWaitCursor();

}
```

References

AGIS Investigators. The Advanced Glaucoma Intervention Study (AGIS): 2.Visual field test scoring. *Ophthalmology* 1994;101:1445-1455.

AGIS Investigators. The Advanced Glaucoma Intervention Study (AGIS): 7.The relationship between control of intraocular pressure and visual field deterioration. *Am J Ophthalmol.* 2000;130:429-40.

Anderson DR, Patella VM. *Automated Static Perimetry.* 2nd Ed. St. Louis, MO: Mosby; 1999:147-159.

Artes PH, Iwase A, Ohno Y, Kitazawa Y, Chauhan BC. Properties of perimetric threshold estimates from Full Threshold, SITA Standard, and SITA Fast strategies. *Invest Ophthalmol Vis Sci.* 2002;43:2654-2659.

Artes PH, Henson DB, Harper R, McLeod D. Multisampling suprathreshold perimetry: a comparison with conventional suprathreshold and full-threshold strategies by computer simulation. *Invest Ophthalmol Vis Sci.* 2003;44:2582-7.

Aung T, Ocaka L, Ebenezer ND, Morris AG, Krawczak M, Thiselton DL, Alexander C, Votruba M, Brice G, Child AH, Francis PJ, Hitchings RA, Lehmann OJ, Bhattacharya SS. A major marker for normal tension glaucoma: association with polymorphisms in the OPA1 gene. *Hum Genet* 2002;110:52-6. (a)

Aung T, Ocaka L, Ebenezer ND, Morris AG, Brice G, Child AH, Hitchings RA, Lehmann OJ, Bhattacharya SS. Investigating the association between OPA1 polymorphisms and glaucoma: comparison between normal tension and high tension primary open angle glaucoma. *Hum Genet* 2002;110:513-4. (b)

Aung T, Okada K, Poinoosawmy D, Membrey L, Brice G, Child AH, Bhattacharya SS, Lehmann OJ, Garway-Heath DF, Hitchings RA. The phenotype of normal tension glaucoma patients with and without OPA1 polymorphisms. *Br J Ophthalmol* 2003;87:149-52.

Bengtsson B, Olsson J, Heijl A, Rootzen H. A new generation of algorithms for computerized threshold perimetry, SITA. *Acta Ophthalmol* 1997;75:368-375.

Bengtsson B, Heijl A. SITA Fast, a new rapid perimetric threshold test: description of methods and evaluation in patients with manifest and suspect glaucoma *Acta Ophthalmol Scand* 1998;76:431-437.

Bengtsson B, Heijl A. Comparing significance and magnitude of glaucomatous visual field defects using the SITA and Full Threshold strategies. *Acta Ophthalmol Scand.* 1999;77:143-146.

Bhandari A, Crabb DP, Poinoosawmy D, Fitzke FW, Hitchings RA, Noureddin BN. Effect of surgery on visual field progression in normal tension glaucoma. *Ophthalmology* 1997;104:1131-1137.

Birch MK, Wishart PK, O'Donnell NP. Determining progressive field loss. In: Mills RP, Wall M, eds. *Perimetry Update 1994/1995.* Amsterdam: Kugler & Ghedini; 1995:31-36. (a)

Birch MK, Wishart PK, O'Donnell NP. Determining progressive visual field loss in serial Humphrey visual fields. *Ophthalmology* 1995;102:1227-1234. (b)

Blumenthal EZ, Sample PA, Berry CC et al. Evaluating sources of variability for standard and SWAP visual fields in glaucoma patients, suspects and normals [ARVO abstract]. *Invest Ophthalmol Vis Sci.* 2000;41:S85.

Budenz DL. *Atlas of Visual Fields.* Philadelphia, PA: Lippincott-Raven; 1997:146-155.

Buhrmann RR, Quigley HA, Barron Y, West SK, Oliva MS, Mmbaga BBO. Prevalence of glaucoma in a rural East African population. *Invest Ophthalmol Vis Sci.* 2000;41:40-48.

Chauhan BC, Drance SM, Lai C. A cluster analysis for threshold perimetry. *Graefe's Arch Clin Exp Ophthalmol* 1989;227:216-220.

Chauhan BC, Drance SM, Douglas GR. The use of visual field indices in detecting changes in the visual field in glaucoma. *Invest Ophthalmol Vis Sci.* 1990;31:512-20.

Chauhan, BC, House, PH. Intratest variability in conventional and high-pass resolution perimetry *Ophthalmology* 1991;98,79-83.

Chauhan BC, Drance SM. The relationship between intraocular pressure and visual field progression in glaucoma. *Graefe's Arch Clin Exp Ophthalmol* 1992;230:521-526.

Chauhan BC, Tompkins JD, LeBlanc RP, McCormick TA. Characteristics of frequency-of-seeing curves in normal subjects, patients with suspected glaucoma, and patients with glaucoma. *Invest Ophthalmol Vis Sci.* 1993;34:3534-40.

Chauhan BC, Drance SM, LeBlanc RP, Lieberman MF, Mills RP, Werner EB. Technique for determining glaucomatous visual field progression by using animation graphics. *Am J Ophthalmol.* 1994;118(4):485-91.

Chauhan BC, House PH, McCormick TA, LeBlanc RP. Comparison of conventional and high-pass resolution perimetry in a prospective study of patients with glaucoma and healthy controls. *Arch Ophthalmol.* 1999;117:24-33. (a)

Chauhan BC, Johnson CA. Test-retest variability of frequency-doubling perimetry and conventional perimetry in glaucoma patients and normal subjects. *Invest Ophthalmol Vis Sci* 1999;40:648-656. (b)

Chen PP, Budenz DL. The effect of cataract on the visual field of eyes with chronic open-angle glaucoma. *Am J Ophthalmol* 1998;125:325-33.

Choplin N, Edwards R. *Visual Fields.* Thorofare, NJ: SLACK Inc; 1998:142-145.

Coffey M, Reidy A, Wormald RPL, Xian WX, Wright L, Courtney P. Prevalence of glaucoma in the West of Ireland. *Br J Ophthalmol* 1993;77:17-21.

Coleman AL. Glaucoma. *Lancet* 1999;354:1803-10.

Crabb DP, Edgar DF, Fitzke FW, McNaught AI, Wynn HP. New approach to estimating the variability in visual field data using an image processing technique. *Br J Ophthalmol.* 1995;79:213-217.

Crabb DP, Fitzke FW, McNaught AI, Edgar DF, Hitchings RA. Improving the prediction of visual field progression in glaucoma using spatial processing. *Ophthalmology* 1997;104:517-524.

Crabb DP, Fitzke FW, Hitchings RA. Detecting gradual and sudden sensitivity loss in series of visual fields. In *Perimetry Update 1998/1999.* (Eds M.Wall and J.M. Wild) Kugler, Amsterdam. 1999:131-138.

Fitzke FW, Crabb DP, McNaught AI, Edgar DF, Hitchings RA. Image processing of computerised visual field data. *Br J Ophthalmol* 1995;79:207-212.

Fitzke FW, Hitchings RA, Poinoosawmy DP, McNaught AI, Crabb DP. Analysis of visual field progression in glaucoma. *Br J Ophthalmol* 1996;80:40-48.

Flammer J, Drance SM, Zulauf M. Differential light threshold - short-term and long-term fluctuation in patients with glaucoma, normal controls, and patients with suspected glaucoma. *Arch Ophthalmol* 1984;102:704-706.

Flammer J, Zulauf M. The frequency distribution of the deviations in static perimetry. *Doc Ophth Proc Ser.* 1985;42:17-24.

Gardiner SK, Crabb DP. Frequency of testing for detecting visual field progression. *Br J Ophthalmol.* 2002;86:560-564. (a)

Gardiner SK, Crabb DP. Examination of different pointwise linear regression methods for determining visual field progression. *Invest. Ophthalmol. Vis. Sci.* 2002;43:1400-1407. (b)

Gardiner SK, Crabb DP, Fitzke FW, Hitchings RA. Reducing noise in suspected glaucomatous visual fields by using a new spatial filter. *Vision Res* 2003; in print.

Garway-Heath DF, Caprioli J, Fitzke FW, Hitchings RA. Scaling the hill of vision: the physiological relationship between light sensitivity and ganglion cell numbers. *Invest. Ophthalmol. Vis. Sci.* 2000;41:1774-1782. (a)

Garway-Heath DF, Poinoosawmy D, Fitzke FW, Hitchings RA. Mapping the visual field to the optic disc in normal tension glaucoma eyes. *Ophthalmology* 2000;107:1809-1815. (b)

Glass E, Schaumberger M, Lachenmayr BJ. Simulations for FASTPAC and the standard 4-2 dB full threshold strategy of the Humphrey field analyser. *Invest Ophthalmol Vis Sci.* 1995;36:1847-1854.

Goldbaum MH, Sample PA, White H, Colt B, Raphaelian P, Fechtner RD, Weinreb RN. Interpretation of automated perimetry for glaucoma by neural network. *Invest Ophthalmol Vis Sci.* 1994;35:3362-73.

Guzman G, Javitt JC, Glick H. Glaucoma in the United States population: the economic burden of illness. *Invest Ophthalmol Vis Sci* 1992;33:S759.

Haley MJ. *The Field Analyzer Primer,* 2nd ed. Allergan Humphrey, San Leandro, CA; 1987.

Hayreh SS. Progress in the understanding of the vascular etiology of glaucoma. *Curr Opinion Ophthalmol* 1994;5:11,26-35.

Heijl A, Lindgren G, Olsson J. A package for the statistical analysis of visual fields. *Doc Ophthal Proc Ser* 1987;49:153-168.

Heijl A, Lindgren G, Olsson J: Test retest variability in glaucomatous visual fields. *Am J Ophthalmol* 1989;108:130-135.

Heijl A, Lindgren G, Lindgren A, Olsson J, Asman P, Myers S, Patella VM. Extended empirical statistical package for evaluation of single and multiple fields in glaucoma: Statpac 2. In: Mills RP,

Heijl A, eds. *Perimetry Update 1990/1991.* Amsterdam: Kugler & Ghedini; 1991:303-315. (a)

Heijl A, Lindgren G, Lindgren A. Normal variability of static perimetric threshold values across the central visual field. *Arch Ophthalmol* 1991;105:1544-1549. (b)

Heijl A; Leske MC; Bengtsson B; Hyman L; Bengtsson B; Hussein M. Reduction of intraocular pressure and glaucoma progression: results from the early manifest glaucoma trial. *Arch Ophthalmol.* 2002;120:1268-1279.

Henson DB, Evans J, Lane C. Fixation accuracy of patients with glaucoma during full threshold perimetry. In: Mills RP, Wall M, eds. *Perimetry Update 1995.* Amsterdam: Kugler & Ghedini; 1995:249-55.

Henson DB, Evans J, Chauhan BC, Lane C. Influence of fixation accuracy on threshold variability in patients with open angle glaucoma. *Invest Ophthalmol Vis Sci.* 1996;37:444-50.

Henson DB. *Visual Fields*, 2nd ed. Oxford University Press; 2000.

Henson DB, Chaundry S, Artes PH, Faragher EB, Ansons A. Response variability in the visual field: comparison of optic neuritis, glaucoma, ocular hypertension, and normal eyes. *Invest Ophthalmol Vis Sci.* 2000;41:417-421.

Hitchings RA. Pyschophysical testing in glaucoma. *Br J Ophthalmol* 1993;77:471-472.

Hitchings RA, Migdal CS, Wormald R, Poinooswamy D, Fitzke FW. The primary treatment trial: changes in the visual field analysis by computer-assisted perimetry. *Eye* 1994;8:117-20.

Hitchings RA, Wu J, Poinooswamy D, McNaught A. Surgery for normal tension glaucoma. *Br J Ophthalmol* 1995;79:402-6.

Hoffman DC, Westcott MC, Gaasterland DE, Caprioli J. Comparison of methods to detect glaucomatous visual field progression [ARVO abstract]. *Invest Ophthalmol Vis Sci.* 2001;42:S815.

Hoffman DC, Li G, Jiang L, Gaasterland DE, Caprioli J. Identifying visual field progression with a linear mixed model. [ARVO abstract]. *Invest Ophthalmol Vis Sci.* 2003;44:E-Abstract 76.

Holmin C, Krakau CET. Regression analysis of the central field in chronic glaucoma cases. *Acta Ophthalmol* 1982;60:267-274.

Hudson C, Wild JM, O'Neill EC. Fatigue effects during a single session of automated static threshold perimetry. *Invest Ophthalmol Vis Sci* 1994;35:268-280.

Hutchings N, Wild JM, Hussey MK, Flanagan JG, Trope GE. The long-term fluctuation of the visual field in stable glaucoma. *Invest Ophthalmol Vis Sci* 2000;41:3429-3436.

Jay JL, Allan D. The benefit of early trabeculectomy versus conventional management in primary open-angle glaucoma relative to severity of disease. *Eye* 1989;3:528-535.

Johnson NL, Kotz S. *Distribution in Statistics: Continuous Univariate Distributions I.* New York: Wiley; 1969.

Johnson CA, Chauhan BC, Shapiro LR. Properties of staircase procedures for estimating thresholds in automated perimetry. *Invest Ophthalmol Vis Sci.* 1992;33:2966-2974.

Johnson CA. Modern developments in clinical perimetry. *Curr Opinion Ophthalmol* 1993;4:7-13.

Johnson CA, Samuels SJ. Screening for glaucomatous visual field loss with frequency-doubling perimetry *Invest Ophthalmol Vis Sci* 1997;38:413-425.

Johnson CA, Cioffi G, Van Buskirk. Frequency doubling technology using a 24-2 stimulus presentation pattern. *Optom Vis Sci.* 1999;76:571-581.

Kamal D, Hitchings R. Normal tension glaucoma - a practical approach. *Br J Ophthalmol.* 1998;82:835-840.

Kass MA, Heuer DK, Higginbotham EJ, Johnson CA, Keltner JL, Miller JP, Parrish RK, Wilson MR, Gordon MO. The ocular hypertension treatment study: a randomized trial determines that topical ocular hypotensive medication delays or prevents the onset of primary open-angle glaucoma. *Arch Ophthalmol* 2002;120:701-13.

Katz J, Gilbert D, Quigley HA, Sommer A. Estimating Progression of visual field loss in glaucoma. *Ophthalmology* 1997;104:1017-1025.

Katz J, Congdon N, Friedman DS. Methodological variations in estimating apparent progressive visual field loss in clinical trials of glaucoma treatment. *Arch Ophthalmol.* 1999;117:1137-1142.

Kim LS, Wild JM, Cunliffe LA, Hancock SA. Longitudinal changes in function and structure in primary open angle glaucoma and ocular hypertension. [ARVO abstract]. *Invest Ophthalmol Vis Sci.* 2003;44:E-Abstract 66.

Klien BEK, Klien R, Sponsel WE, Franke T, Cantor LB, Martone J, Menage MJ. Prevalence of glaucoma: the Beaver Dam eye study. *Ophthalmology* 1992;99:1499-1504.

Koseki N, Araie M, Shirato S, Yamamoto S. Effect of trabeculectomy on visual field performance in central 30 degrees in progressive normal-tension glaucoma. *Ophthalmology* 1997;104:197-201.

Lachenmayr BJ, Vivell PMO. *Perimetry and its Clinical Correlations.* Stuttgart: Georg Thieme Verlag; 1993.

Leske MC. The epidemiology of open angle glaucoma: A review. *Am J Epidemiol* 1983;118:166-191.

Leske MC, Connell AMS, Schachat AP, Hyman L. The Barbados eye study: Prevalence of open angle glaucoma. *Arch Ophthalmol* 1994;112:821-829.

Leske MC, Heijl A, Hyman L, Bengtsson B. Early manifest glaucoma trial - Design and baseline data. *Ophthalmology* 1999;106:2144-2153.

Leske MC, Wu SY, Nemesure B, Hennis A. Risk factors for incident nuclear opacities. *Ophthalmology* 2002;109:1303-1308.

Lietman T, Eng J, Katz J, Quigley HA. Neural networks for visual field analysis: how do they compare with other algorithms? *J Glaucoma.* 1999;8:77-80.

McNaught AI, Crabb DP, Fitzke FW, Hitchings RA. Modelling series of visual fields to detect progression in normal-tension glaucoma. *Graefes Arch Clin Exp Ophthalmol* 1995;233:750-755.

McNaught AI, Crabb DP, Fitzke FW, Hitchings RA. Visual Field Progression: comparison of Humphrey Statpac2 and pointwise linear regression analysis. *Graefes Arch Clin Exp Ophthalmol* 1996;234:411-418.

Membrey WL, Poinoosawmy DP, Bunce C, Fitzke FW, Hitchings RA. Comparison of visual field progression in patients with normal pressure glaucoma between eyes with and without visual field loss that threatens fixation. *Br J Ophthalmol* 2000;84:1154-1158.

Membrey WL, Bunce C, Poinoosawmy DP, Fitzke FW, Hitchings RA. Glaucoma surgery with or without adjunctive antiproliferatives in normal tension glaucoma: 2 Visual field progression. *Br J Ophthalmol.* 2001;85:696-701.

Mikelberg FS, Drance SM. The mode of progression of visual field defects in glaucoma. *Am J Ophthalmol* 1984;98:443-5.

Morgan RM, Feuer WJ, Anderson DR. Statpac 2 Glaucoma Change Probability. *Arch Ophthalmol* 1991;109:1690-1692.

Musch DC, Lichter PR, Guire KE, Standardi CL, The CIGTS Study Group. The Collaborative Initial Glaucoma Treatment Study: study designs, methods, and baseline characteristics of enrolled patients. *Ophthalmology* 1999;106:653-62.

Nouri-Mahdavi K, Brigatti L, Weitzman M, Caprioli J. Comparison of methods to detect visual field progression in glaucoma. *Ophthalmology* 1997;104:1228-1236.

O'Brien C, Schwartz B: The visual field in chronic open angle glaucoma. The rate of change in different regions of the field. *Eye* 1990;4:557-562.

O'Brien C, Schwartz B, Takamato T, Wu DC. Intraocular pressure and the rate of visual field loss in chronic open-angle glaucoma. *Am J Ophthalmol* 1991;111:491-500.

O'Brien C, Schwartz B. Point by point linear regression analysis of automated visual fields in primary open-angle glaucoma. In: Mills RP, ed. *Perimetry Update.* Amsterdam: Kugler and Ghedini, 1993:149-152.

Pang CP, Leung YF, Fan B, Baum L, Tong WC, Lee WS, Chua JKH, Fan DSP, Liu Y, Lam DSC. TIGR/MYOC Gene Sequence Alterations in Individuals with and without Primary Open-Angle Glaucoma. *Invest Ophthalmol Vis Sci.* 2002;43:3231-35.

Perneger TV. What's wrong with Bonferroni adjustments. *British Medical Journal* 1998;316:1236-1238.

Racette L, Medeiros FA, Pascual JP, Zangwill LM, Weinreb RN, Sample PA. Transforming the visual field scale from decibel to 1/Lambert: does is alter the structure-function relationship? [ARVO abstract]. *Invest Ophthalmol Vis Sci.* 2003;44:E-Abstract 77.

Rose C, Smith MD. *Mathematical statistics with Mathematica.* New York: Springer-Verlag; 2002:149-160.

Sakai T, Kuniyoshi S, Tsuzuki K, Makoto U, Kawamura Y. Temporal raphe of the retinal nerve fiber layer revealed by medullated fibers. *Jpn J Ophthal.* 1987;31:655-658.

Sarfarazi M, Child A, Stoilova D, Brice G, Desai T, Trifan OC, Poinoosawmy D, Crick RP. Localization of the fourth locus (GLC1E) for adult-onset primary open angle glaucoma to the 10p15-p14 region. *Am J Hum Genet* 1998;62:641-52.

Sample PA, Cook JN, Weinreb RN. Variability and sensitivity of short-wavelength color visual fields in normal and glaucoma eyes. *Ophthalmic & Visual Optics/Noninvasive Assessment of the Visual System Technical Digest,* Optical Society of America Washington, DC., USA; 1993:292-295.

Schulzer M, The Normal Tension Glaucoma Study Group. Errors in the diagnosis of visual field progression in normal tension glaucoma. *Ophthalmology* 1994;101:1589-1595.

Schwartz SH. *Visual perception. A clinical orientation.* 2nd Ed. Stuttgart: Georg Thieme Verlag; 1999:249-68.

Sekhar GC, Naduvilath TJ, Lakkai M, Jayakumar AJ, Pandi GT, Mandal AK, Honovar SG. Sensitivity of Swedish interactive threshold algorithm compared with standard full threshold algorithm in Humphrey visual field testing. *Ophthalmology* 2000;107:1303-1308.

Shapiro LR, Johnson CA, Kennedy RL. KRAKEN: A computer simulation procedure for static, kinetic, supra-threshold and heuristic perimetry. In: Heijl A, ed. *Perimetry Update.* Amsterdam: Kugler and Ghedini; 1989:431-438.

Shareef YS, Lesser GR, Imami NR, Darnley-Fisch D, Dahl D, Trick GL. Dissociation Between Functional And Morphological Changes In Glaucoma. *Invest Ophthalmol Vis Sci.* 2002;43:E-Abstract 1001.

Sharma AK, Goldberg I, Graham SL, Mohsin M. Comparison of the Humphrey Swedish interactive thresholding algorithm (SITA) and full threshold strategies. *J Glaucoma.* 2000;9:20-27.

Shirato S, Inoue R, Fukushima K, Suzuki Y. Clinical evaluation of SITA: a new family of perimeteric testing strategies. *Graefes Arch Clin Exp Ophthalmol.* 1999;237:29-34.

Smith RJH. Medical versus surgical therapy in glaucoma simplex. *Br J Ophthalmol* 1972;56:277-283.

Smith RJH. The enigma of primary open angle glaucoma. *Trans Ophthalmol Soc UK* 1986;105:618-633.

Smith SD, Katz J, Quigley HA. Analysis of progressive change in automated visual fields in glaucoma. *Invest Ophthalmol Vis Sci.* 1996;37:1419-1428.

Smith SD, Katz J, Quigley HA. Effect of cataract extraction on the results of automated perimetry in glaucoma. *Arch Ophthalmol* 1997;115:1515-19.

Spenceley SE, Henson DB, Bull DR. Visual field analysis using artificial neural networks. *Ophthalmic Physiol Opt* 1994;14:239-48.

Spenceley SE, Henson DB. Visual field test simulation and error in threshold estimation. *Br J Ophthalmol* 1996;80:304-308.

Sponsel WE, Dallas NL, Burbidge L. Visual field survival: the response to timolol therapy in open-angle glaucoma. *Br J Ophthalmol* 1983;67:220-227.

Spry PGD, Bates AB, Johnson CA, Chauhan BC. Simulation of longitudinal threshold visual field data. *Invest Ophthalmol Vis Sci.* 2000;41:2192-2200.

Spry PGD, Johnson CA, McKendrick AM, Turpin A. Variability components of standard automated perimetry and frequency-doubling technology perimetry. *Invest Ophthalmol Vis Sci.* 2001;42:1404-1410.

Spry PGD, Johnson CA. Senescent changes of the normal visual field: an age-old problem. *Optom Vis Sci.* 2001;78:436-441.

Spry PGD, Johnson CA, Bates AB, Turpin A, Chauhan BC. Spatial and temporal processing of threshold data for detection of progressive glaucomatous visual field loss. *Arch Ophthalmol.* 2002;120:173-180.

Spry PGD, Johnson CA. Identification of progressive glaucomatous visual field loss. *Surv Ophthalmol* 2002;47:158-173.

Stoilova D, Child A, Trivan OC, Crick RP, Coakes RL, Sarfarazi M. Localization of a locus (GLC1B) for adult-onset primary open angle glaucoma to the 2 cen-q13 region. *Genomics* 1996;36:142-50.

Stone EM, Fingert JH, Alward WLM, Nguyen TD, Polansky JR, Sunden SLF, Nishimura D, Clark AF, Nystuen A, Nichols BE, Mackey DA, Ritch R, Kalenak JW, Craven ER, Sheffield VC. Identification of a gene that causes primary open angle glaucoma. *Science* 1997;275:668-70.

Tate GW, Lynn JR. *Principles of Quantitative Perimetry.* Grune and Stratton, New York; 1977.

Thylefors B, Negrel AD. The global impact of glaucoma. *Bull World Health Organ.* 1994;72:323-6.

Tielsch JM, Sommer A, Katz J, Royall RM, Quigley HA, Javitt JC. The Baltimore eye survey - racial variations in the prevalence of primary open-angle glaucoma. *JAMA* 1991;266:369-374.

Tielsch JM, Katz J, Sommer A, Quigley HA, Javitt JC. Family history and risk of primary open-angle glaucoma - the Baltimore eye survey. *Arch Ophthalmol* 1994;112:69-73.

Traquair HM. Perimetry in the study of glaucoma. *Trans Ophthalmol Soc UK* 1931;51:585-599.

Trifan OC, Traboulsi EI, Stoilova D, Alozie I, Nguyen R, Raja S, Sarfarazi M. The third locus (GLC1D) for adult-onset primary open angle glaucoma maps to the 8q23 region. *Am J Ophthalmol* 1998;126:17-28.

Vesti E, Spry PGD, Chauhan BC, Johnson CA. Sensitivity differences between real-patient and computer-simulated visual fields. *J Glaucoma* 2002;11:35-45.

Viswanathan AC, Fitzke FW, Hitchings RA. Early detection of visual field progression in glaucoma: a comparison of Progressor and Statpac2. *Br J Ophthalmol* 1997;81:1037-1042. (a)

Viswanathan AC, Crabb DP, Westcott MC, McNaught AI, Fitzke FW, Hitchings RA. Inter-observer agreement on visual field progression in glaucoma: A comparison of methods [ARVO abstract]. *Invest Ophthalmol Vis Sci.* 1997;38:S2662. (b)

Viswanathan AC, Hitchings RA, Fitzke FW. How often do patients need visual field tests? *Graefes Arch Clin Exp Ophthalmol* 1997;235:563-568. (c)

Viswanathan AC, McNaught AI, Poinoosawmy D, Fontana L, Crabb DP, Fitzke FW, Hitchings RA. Severity and Stability of Glaucoma: Patient Perception compared with Objective Measurement. *Arch Ophthalmol.* 1999;117:450-454.

Viswanathan AC, Crabb DP, McNaught AI, Westcott MC, Kamal D, Garway-Heath DF, Fitzke FW, Hitchings RA. Inter-observer agreement on visual field progression in glaucoma: A comparison of methods. *Br J Ophthalmol.* 2003;87:726-30.

Vrabec F. The temporal raphe of the human retina. *Am J Ophthalmol.* 1966;62:926-938.

Wall M, Punke SG, Stickney TL, Brito CF, Withrow KR, Kardon RH. SITA Standard in Optic Neuropathies and Hemianopias: A Comparison with Full Threshold Testing. *Invest Ophthalmol Vis Sci* 2001;42:528-537.

Werner EB, Bishop KI, Koelle J, Douglas GR, LeBIanc RP, Mills RP, Schwartz B, Whalen WR,

Wilensky JT. A comparison of experienced clinical observers and statistical tests in detection of progressive visual field loss in glaucoma using automated perimetry. *Arch Ophthalmol.* 1988;106:619-23.

Werner EB. *Manual of visual fields.* Churchill Livingstone, New York; 1991.

Westcott MC, Hoffman D, Gaasterland D, Viswanathan AC, Caprioli J. Agreement between pointwise linear regression and AGIS score change in identifying visual field deterioration in glaucoma [ARVO abstract]. *Invest Ophthalmol Vis Sci.* 2001;42:S3005.

Wild JM, Searle AET, Dengler-Harles M, O'Neill EC. Long-term follow-up of baseline learning and fatigue effects in the automated perimetry of glaucoma and ocular hypertensive patients. *Acta Ophthalmol* 1991;69:210-216.

Wild JM, Hussey MK, Flanagan JG, Trope GE. Pointwise topographical and longitudinal modelling of the visual field in glaucoma. *Invest Ophthalmol Vis Sci* 1993;34:1907-1916.

Wild JM, Hutchings N, Hussey MK, Flanagan JG, Trope GE. Univariate linear regression of pointwise visual field sensitivity against time of follow-up. *Invest Ophthalmol Vis Sci* 1994;35:S4306.

Wild JM, Moss ID, Whitaker D, O'Neill EC. The statistical interpretation of blue-on-yellow visual field loss *Invest Ophthalmol Vis Sci* 1995;36:1398-1410.

Wild JM, Hutchings N, Hussey MK, Flanagan JG, Trope GE. Pointwise univariate linear regression of perimetric sensitivity against follow-up time in glaucoma. *Ophthalmology* 1997;104:808-15.

Wild JM, Cubbidge RP, Pacey IE, Robinson R. Statistical aspects of the normal visual field in short-wavelength automated perimetry *Invest Ophthalmol Vis Sci* 1998;39:54-64.

Wild JM, Pacey IE, Hancock SA, Cunliffe IA. Between-algorithm, between-individual differences in normal perimetric sensitivity: full threshold, FASTPAC, and SITA: Swedish Interactive Threshold Algorithm. *Invest Ophthalmol Vis Sci.* 1999;40:1152-61. (a)

Wild JM, Pacey IE, O'Neill EC, Cunliffe IA. The SITA perimetric threshold algorithms in glaucoma. *Invest. Ophthalmol. Vis. Sci.* 1999;40:1998-2009. (b)

Wirtz MK, Samples JR, Kramer PL, Rust K, Topinka JR, Yount J, Koler RD, Acott TS. Mapping a gene for adult-onset primary open angle glaucoma to chromosome 3q. *Am J Hum Genet* 1997;60:296-304.

Wirtz MK, Samples JR, Rust K, Lie J, Nordling L, Schilling K, Acott TS, Kramer PL. GLC1F, a new primary open-angle glaucoma locus, maps to 7q35-q36. *Arch Ophthalmol* 1999;117:237-41.

Wolfs RCW, Klaver CCW, Ramrattan RS, Van Duijn CM, Hofman A, de Jong PTVM. Genetic risk of primary open glaucoma. *Arch Ophthalmol* 1998;116:1640-45.

Wolfs RCW, Borger PH, Ramrattan RS, Klaver CCW, Hulsman CAA, Hofman A, Vingerling JR, Hitchings RA, De Jong PTVM. Changing views on open-angle glaucoma: definitions and prevalences – the Roterdam study. *Invest Ophthalmol Vis Sci.* 2000;41:3309-21.

Wood JM, Wild JM, Hussey MK, Crews SJ. Serial examination of the normal visual field using Octopus automated projection perimetry evidence for a learning effect. *Acta Ophthalmol* 1987;65:326-33.

Wormald R. Treatment of raised intraocular pressure and prevention of glaucoma. *BMJ* 2003;326:723-724.

Reprints of published work

1. Gardiner SK, Crabb DP. Frequency of testing for detecting visual field progression. *Br J Ophthalmol.* 2002;86:560-564.

2. Gardiner SK, Crabb DP. Examination of different pointwise linear regression methods for determining visual field progression. *Invest. Ophthalmol. Vis. Sci.* 2002;43:1400-1407

3. Gardiner SK, Crabb DP, Fitzke FW, Hitchings RA. Reducing noise in suspected glaucomatous visual fields by using a new spatial filter. *Vision Res* 2003; in print.

CLINICAL SCIENCE

Frequency of testing for detecting visual field progression

S K Gardiner, D P Crabb

Br J Ophthalmol 2002;**86**:560–564

See end of article for authors' affiliations

Correspondence to: Dr D P Crabb, Faculty of Science and Mathematics, Nottingham Trent University, Clifton Campus, Nottingham NG11 8NS, UK; david.crabb@ntu.ac.uk

Accepted for publication 30 January 2002

Aims: To investigate the effect of frequency of testing on the determination of visual field progression using pointwise linear regression (PLR).
Methods: A "virtual eye" was developed to simulate series of sensitivities over time at a given point in the eye. The user can input the actual behaviour of the point (for example, stable or deteriorating steadily), and then a configurable amount of noise is added to produce a realistic series over time. The advantage of this over using patient data is that the actual status of the eye is known. Series were generated using different frequencies of testing, and the diagnosis that would have been made from each series was compared with the true status of the eye. A point was diagnosed as progressing if the regression line for the series showed a deterioration of at least 1 dB per year, significant at the 1% level. From these results, graphs were produced showing the number of points correctly or incorrectly diagnosed as progressing.
Results: With the virtual eye deteriorating at a rate of 2 dB/year, it was found that the point was determined to be progressing quicker when more tests were carried out each year. With a stable virtual eye, it was found that increasing the frequency of testing increased the number of series that were falsely labelled as progressing during the first 3 years of testing.
Conclusions: As the frequency of testing increases, the sensitivity of PLR increases. However, the specificity decreases; possibly meaning more unnecessary changes in treatment. Three tests per year provide a good compromise between sensitivity and specificity.

Visual field results from computerised perimetry are the standard measurement of a patient's visual function and are likely to remain so in future years, especially with the promise of better measurement strategies.[1 2] Perimetry shows the actual effect on the patient's vision, which will vary even among patients with, for example, the same degree of inflated intraocular pressure or damage to the optic nerve. Evaluation of change in a series of visual fields remains an important aspect of monitoring patients for disease progression. New methods, using linear regression analysis of individual sensitivity values against time of follow up have been developed.[3–6] These techniques, known as pointwise linear regression (PLR), have been shown by several studies to compare favourably with other methods for detecting visual field progression in patients with glaucoma.[7–12]

The main purpose of this study was to investigate the effect of frequency of testing on the performance of PLR in detecting gradual sensitivity deterioration. This is particularly important to the clinical management of patients with glaucoma. This type of investigation, like others examining methods for quantifying visual field changes, is hampered by the lack of an external "gold standard" for progression. Computer simulations have been extensively used as an alternative to analysing actual patient data to develop improved perimetric testing strategies[13–16]: they offer a more reproducible and controllable means of examining the behaviour of visual field data. This suggests that they may also be useful in examining the problems of detecting progression in series of visual field data. Hence, this study utilises a computer generated "virtual eye" to produce series of sensitivity values typical of longitudinal visual field data.

METHODS

Virtual eye

The "Virtual eye" simulation program was purpose written using object oriented statistical software (S-PLUS 2000 for Windows, StatSci Europe, MathSoft Inc, Oxford). It generates series of sensitivity values typical of those found at individual test locations within the visual field. The virtual eye allows control over components (parameters) that affect the behaviour of a series of sensitivity values against time:

(1) Length of series (years)
(2) Frequency of observations (fields per year)
(3) Initial level of sensitivity (dB)
(4) Number of simulations (series)
(5) Size of loss (dB per year)
(6) Type of loss (gradually or suddenly deteriorating)
(7) Criteria for progression used by PLR
(8) Noise (variability between and within sensitivity values).

The benefit of having a virtual eye is that the user can specify the actual, noise free behaviour of the eye, and then add noise in later. Thus, knowing that the eye is in fact (for example) stable, it can be seen how many times a correct decision is made when the noisy series are tested. The virtual eye is simplified by considering the behaviour of the sensitivity at one location in the visual field.

Because we know the exact actual sensitivity at each point in the eye (as these are specified), the virtual eye only needs to consider the sensitivity at one point. This assumes that given the actual sensitivity at each point, the noise is independent between points. Note that we are not assuming that actual (noise free) progression is independent between points; but because we have specified this exactly, the noise free behaviour of neighbouring points is irrelevant.

The idea is best explained by example. The results from a simulation can be displayed as individual plots of sensitivity values against time of follow up. In Figure 1, the top row shows the actual sensitivity values as specified for a stable eye. This is defined to be one which shows only the expected age related deterioration of 0.1 dB/year.[17] The lines are those produced by linear regression based on the first three points of the series, then the first four points, and so on; so the graphs

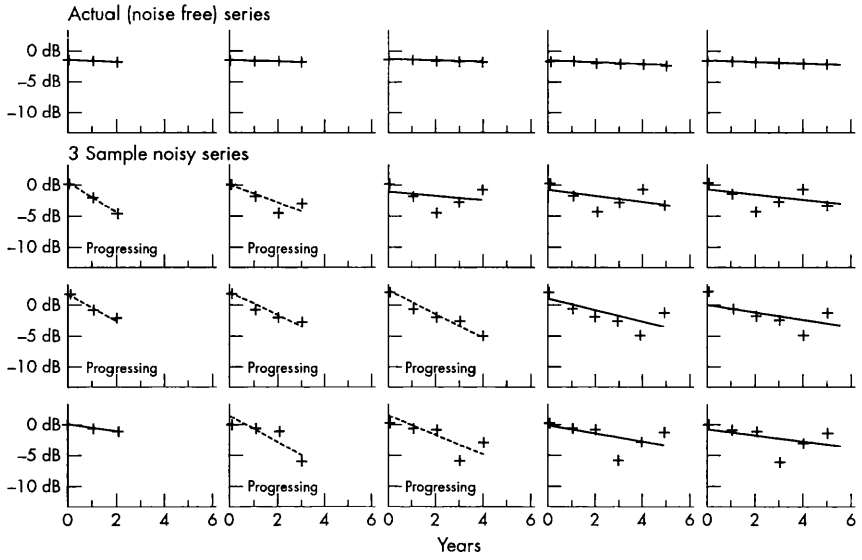


Figure 1 Illustrative series for a stable virtual eye. In the top row, the sensitivity at a point is specified as deteriorating at 0.1 dB/year over 6 years; as the series gets longer (moving from left to right along the series) with the addition of more points, PLR is used to determine whether the point would be flagged as progressing based on the readings so far. The second, third, and fourth rows represent three possible series of artificial sensitivities, after noise has been added in.

show how the results would change as more tests were carried out in subsequent years and added on to the end of the series.

Each point on the graph can be thought of as a threshold deviation (dB) from baseline labelled as 0 dB on the vertical axes. No assumption is made about the actual starting sensitivity. For this illustration it is assumed that the visual field tests are taken at yearly intervals. The criteria for progression are specified as regression slope worse than −1 dB/year and also statistically significant at the 1% level.

Subsequent rows illustrate three different examples of possible artificial series; noise has been added to the actual noise free values in the top row. It is now seen that owing to the effect of the noise, it is possible for the series to be labelled as progressing according to the specified criteria even though we know that the point is actually stable. As time goes on, the diagnosis will get more accurate, with fewer stable series being falsely labelled as progressing.

In Figure 2, the actual sensitivity of the point (as seen in the top row) is deteriorating at a rate of 2 dB/year. This represents approximately 20 times the normal rate of decay,[17] and double the criteria for progression of 1 dB/year we are using for PLR (see above). When noise is added, the point is regularly labelled as being stable, even though we know this not to be the case. This is either because the slope is not sufficiently steep, or because the series is too noisy for the slope to be statistically significant.

In both stable and progressing cases, as the series gets longer, a higher proportion of series will be correctly flagged.

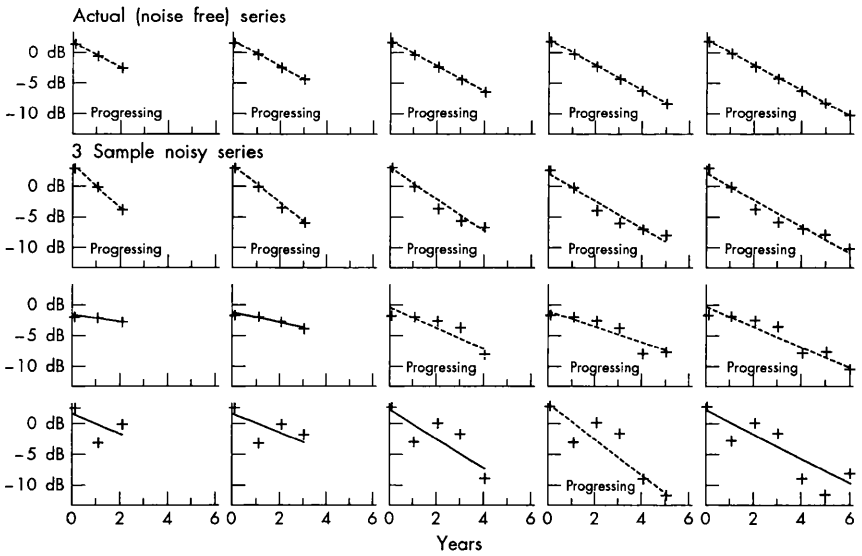


Figure 2 Illustrative series for a progressing virtual eye. In the top row, the sensitivity at a point is specified as deteriorating at 2 dB/year over 6 years; as the series gets longer with the addition of more points, PLR is used to determine whether the point would be flagged as progressing based on the readings so far. This is then done for three sample series of artificial sensitivities, after noise has been added in.

This study seeks to discover whether increasing the frequency of testing will also improve the success rate.

Simulation experiment

Series of actual sensitivity values over a period of 6 years were generated. These series were either designated as "stable," having an age related deterioration of 0.1 dB/year; or "progressing," deteriorating at 2 dB/year. Next, at each test date (determined by the number of tests being carried out each year, which were assumed to be equally spaced through the year), a noisy sensitivity reading was generated by randomly sampling from a normal distribution, with a mean value equal to the calculated actual sensitivity at that time, and a standard deviation of 2 dB. This noise is assumed to be the combination of both short term (within test variability) and long term (between test variability) fluctuation. Moreover, we utilise the simplifying assumption that the amount of noise present in each reading is constant throughout the series. In fact, some patients, and some points within each patient's visual field, will be noisier than others. Also the amount of noise actually increases as the eye deteriorates.[18 19] For example, a point measured at 20 dB will have a higher level of noise than one measured at 30 dB. However, to obtain a fair comparison between different methods, it is essential that precisely the same conditions occur in each case. After all, this possibility is the basic advantage of using a virtual eye in the first place. Therefore in this experiment the same level of noise is used throughout. Henson *et al*[18] demonstrate that noise of approximately 2 dB would be expected if the point was initially at its age related normal level (that is, about 30 dB sensitivity). Heijl *et al*[20] show that for central points whose initial sensitivity equals the age corrected normal sensitivity, 95% of the points will have an intertest variation (the difference between consecutive tests) of between roughly −5 dB and +3 dB, including any actual deterioration over that time. For normally distributed noise, 95% of the distribution falls within 1.96 standard deviations of the mean; and so the standard deviation is approximately 2 dB. However, for lower sensitivities, and (as a result) for points in the periphery, which generally have a lower sensitivity, the amount of noise would be much higher.

The criteria for progression were defined as a slope of at least −1 dB/year, statistically significant at the 1% level. These criteria have been used in several published studies using PLR.[8 9 21–23] The minimum slope guards against significant age related decline, and has also been shown to be related to other methods of pointwise change detection.[3 4] PLR was then carried out on the first three readings to see if they would be flagged as progressing or not. This was then repeated for the first four readings, then five readings and so on. This way it can be seen how the diagnosis (progressing or stable) would change over time as more readings were added to the series. Then, the percentage of correctly diagnosed series (out of 1000) at each point in time was calculated. This experiment was repeated for different numbers r of readings per year.

RESULTS

We compare the number of series of points which would be flagged as progressing when simulated as before, for different numbers r of readings per year. Figure 3A gives stable points, and Figure 3B is points with an actual deterioration of 2 dB/year. They show that, for example, if there is a small local defect of say five points that are actually progressing, then after 3 years at two tests per year the method will flag on average around four of these truly progressing points, and also two or three points elsewhere in the eye that are actually stable.

Clearly, as the number r of tests per year increases, so does the proportion of points labelled as progressing. This confirms the intuitive notion that a progressing eye will be detected quicker with more frequent testing. However, there are also an increased number of early false positives—that is, stable points being labelled as progressing. More frequent testing will result in a higher number of incorrect decisions being made for non-deteriorating patients when the follow up is relatively short. Over the first 2 years of testing, doubling the frequency of testing also doubles the proportion of false positive results—that is, stable points being incorrectly flagged as progressing by PLR. More than 3 years have passed before the lines converge, and this higher error rate disappears, but by this time the sensitivities are also converging, and so the benefits of carrying out more tests per year have vanished. Under the conditions of this experiment, it is never worthwhile carrying out more than three tests per year; the extra expense and inconvenience to the patient of doing so brings no reward.

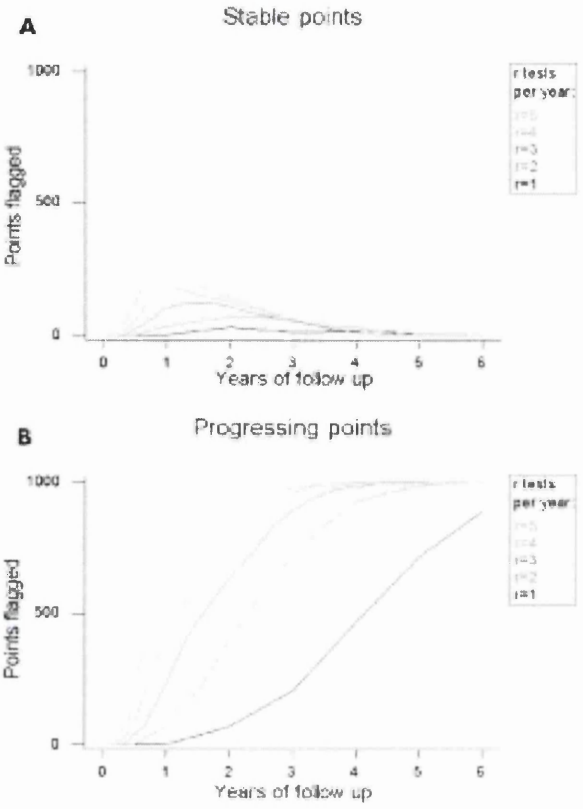


Figure 3 Performance of PLR, for different numbers of tests per year. (A) Each line shows how many actually stable points were flagged as progressing out of 1000 simulated series, when r sensitivity readings were simulated per year. (B) Shows the same for points that were actually deteriorating at 2 dB/year.

DISCUSSION

Computer simulations have been widely used to investigate perimetric testing strategies by modelling the patient response during a perimetric examination.[13–16] The virtual eye simulation described in this study may be useful for exploring the behaviour of visual field progression. Previously we have used a computer simulation similar to the one described here to show that PLR is equally sensitive to detecting both gradual (linear) and sudden (episodic) sensitivity loss.[24] The detection rates, however, were moderate in most cases because of the variability between observations. More recently Spry *et al*[25] have used a computer simulation approach to investigate longitudinal visual field data. This simulation differed from the virtual eye presented here, in that sequences of complete visual fields were generated by interpolating between two

Table 1 The dramatic effect of a relatively small increase in specificity

Prob (given point is FP)	Prob (≥1 point is FP)
3%	79.5%
2%	65.0%
1%	40.7%

"real" measured fields. Simulations allow different test parameters to be analysed using large materials in a short time, and the effect of single variables or parameters can be isolated. The latter is not possible using patient data alone and our plan is to develop this methodology further.

The experiment described in this study has shown that although increasing the number of tests per year speeds up the detection of progression (Fig 3B), it does so at the expense of, initially at least, falsely labelling far more stable points as progressing (Fig 3A); the extra tests are actually making the performance worse. This is because the noise becomes far more significant than the amount of change that may or may not have occurred over the shorter period of time. If the PLR slope based on one test per year for 5 years is k, then the same sensitivities compressed into just 1 year (that is, five tests per year) would give a PLR slope of 5k, even though the eye may not actually be deteriorating any faster.

The first priority is, in many cases, specificity; because this reduces the chances of incorrect changes in clinical management. If, for example, there is a 3% probability that any given stable point will be incorrectly flagged as progressing (which is the level of false positives seen after as much as 3 or 4 years of follow up in Fig 3), then the chance of at least one point out of the 52 test locations in a 24-2 Humphrey visual field being incorrectly flagged is:

Prob(≥1 incorrect point) = 1 − prob (a given point is incorrect)[52]

$= 1 - 0.97^{52}$

= 79.5%

This alarmingly poor performance is improved dramatically by a seemingly small improvement in the specificity at one point (see Table 1).

Hence we have shown that having an increased number of tests per year reduces specificity when the follow up is relatively short. Although the lines in Figure 3 do converge as the series lengthens, a clinical management decision based on a significant slope would be unwise with frequent testing in say the first 2 or 3 years. One strategy is to use one or more confirmation fields which may or may not be part of the actual follow up. This has been done in recent studies using PLR[20 23] and applied to other methods for detecting progression.[26] Of course, this means even more visual field testing to detect progression; and even though this is a sensible approach, ad hoc confirmation fields do not have any exact specificity associated with them. We believe the virtual eye described in this paper could provide more accurate estimates of this specificity, or indeed the diagnostic precision of any visual field progression criteria being used as an outcome measure in investigations and clinical trials.

The second clinical priority is to improve the sensitivity of the test; in other words detecting true visual field progression. From Figure 3B it is clear that the detection of progressing points with just one test carried out per year is extremely poor. Even with two tests per year, points deteriorating at 2 dB/year would only be picked up on in 75% of cases after 3 years; by this time a loss of 6 dB has occurred, which is severe enough that any decent method should have identified it. Clinically, this means that a substantial fraction of visual field locations have to be truly deteriorating before progression can be reliably detected. Therefore, approximately three tests per year, resources permitting, seem to achieve a better success rate at determining which points are progressing and which are not. This supports and extends findings on patient data in glaucoma.[21]

This model provides a much simplified version of reality. The assumption of noise being normally distributed, though commonly used, is unproved and the amount of noise present in readings would typically be larger than the estimates used here[19 20 27–29] and increase as the measured sensitivity at a location decreases.[18 19] As the amount of noise increases, the performance of PLR becomes even worse than in Figure 3. Several years' worth of follow up is needed before satisfactory results can be obtained. Clinically, it is common practice to use confirmation fields to look for points that are persistently progressing.[22 23] The uses of confirmation fields, and different levels of noise have also been tried using our simulation model and although the values for specificity and sensitivity naturally altered, the qualitative results (most importantly the recommendations on ideal frequency of testing) were unchanged. For clarity, the simplest case, without confirmation fields, has been described.

It is also common practice to look for clusters of points that are all progressing, rather than individual points. While this is beyond the scope of this simulation model (because the spatial pattern of deterioration is not clearly known, and varies according to the location in the eye), it is still desirable to have an accurate determination of progression for each of the points in the cluster. The findings, in the present form, may have limited clinical generalisability; but it is our opinion that the conclusions drawn from them are more widely applicable. Non-glaucomatous change (that is, effect of concomitant cataract) is currently indistinguishable for glaucomatous change using PLR and so cannot be included in the virtual eye simulation. Nevertheless, use of simulation enables results to be found which would be extremely hard to achieve using patient data, because for a real world patient, the underlying noise free state of the eye is, as yet, unknown.

Our findings suggest that the current PLR, a popular method for measuring visual field progression[3–12 22 23] still needs improvement. Methods are needed which are more sensitive without compromising specificity; this way, accurate results and faster diagnosis could be obtained with just one or two tests per year over shorter periods of time than at present. This relies primarily on reducing the amount of noise present in the readings.

In conclusion, we believe that three tests per year is a good compromise between sensitivity and specificity. There are certainly large benefits to be had in sensitivity when compared with carrying out just one test per year.

Authors' affiliations

S K Gardiner, D P Crabb, Faculty of Science and Mathematics, The Nottingham Trent University, Nottingham, UK

REFERENCES

1 **Bengtsson B,** Olsson J, Heijl A, *et al.* A new generation of algorithms for computerized threshold perimetry, SITA. *Acta Ophthalmol* 1997;**75**:368–75.
2 **Sekhar GC,** Naduvilath TJ, Lakkai M, *et al.* Sensitivity of Swedish interactive threshold algorithm compared with standard full threshold algorithm in Humphrey visual field testing. *Ophthalmology* 2000;**107**:1303–8.
3 **Fitzke FW,** Hitchings RA, Poinoosawmy D, *et al.* Analysis of visual field progression in glaucoma. *Br J Ophthalmol* 1996;**80**:40–8.
4 **McNaught AI,** Crabb DP, Fitzke FW, *et al.* Visual field progression: comparison of Humphrey Statpac2 and pointwise linear regression. *Graefes Arch Clin Exp Ophthalmol* 1996;**234**:411–18.
5 **O'Brien C,** Schwartz B. The visual field in chronic open angle glaucoma. The rate of change in different regions of the field. *Eye* 1990;**4**:557–62.
6 **O'Brien C,** Schwartz B. Point by point linear regression analysis of automated visual fields in primary open-angle glaucoma. In: Mills RP, ed. *Perimetry update.* Amsterdam: Kugler and Ghedini, 1993:149–52.

7 **Smith SD**, Katz J Quigley HA. Analysis of progressive change in automated visual fields in glaucoma. *Invest Ophthalmol Vis Sci* 1996; **37**:1419–28.
8 **Wild JM**, Hutchings N, Hussey MK, *et al.* Pointwise univariate linear regression of perimetric sensitivity against follow-up time in glaucoma. *Ophthalmology* 1997;**104**:808–15.
9 **Katz J**, Gilbert D, Quigley HA, *et al.* Estimating progression of visual field loss in glaucoma. *Ophthalmology* 1997;**104**:1017–25.
10 **Viswanathan AC**, Fitzke FW, Hitchings RA. Early detection of visual field progression in glaucoma: a comparison of Progressor and Stapac2. *Br J Ophthalmol* 1997;**81**:1037–42.
11 **Nouri-Mahdavi K**, Brigatti L, Weitzman M, *et al.* Comparison of methods to detect visual field progression in glaucoma. *Ophthalmology* 1997;**104**:1228–36.
12 **McNaught AI**, Crabb DP, Fitzke FW, *et al.* Modelling series of visual fields to detect progression in normal-tension glaucoma. *Graefes Arch Clin Exp Ophthalmol* 1995;**233**:750–5.
13 **Shapiro LR**, Johnson CA, Kennedy RL. Kraken: a computer simulation procedure for static, kinetic, supra-threshold and heuristic perimetry. In: Heijl A, ed. *Perimetry update*. Amsterdam: Kugler and Ghedini, 1989:431–8.
14 **Johnson CA**, Chauhan BC, Shapiro LR. Properties of staircase procedures for estimating thresholds in automated perimetry. *Invest Ophthalmol Vis Sci* 1992;**33**:2966–74.
15 **Glass E**, Schaumberger M, Lachenmayr BJ. Simulations for FASTPAC and the standard 4-2 dB full threshold strategy of the Humphrey field analyser. *Invest Ophthalmol Vis Sci* 1995;**36**:1847–54.
16 **Spenceley SE**, Henson DB. Visual field test simulation and error in threshold estimation. *Br J Ophthalmol* 1996;**80**:304–8.
17 **Heijl A**, Lindgren G, Lindgren A. Normal variability of static perimetric threshold values across the central visual field. *Arch Ophthalmol* 1991;**105**:1544–9.
18 **Henson DB**, Chaundry S, Artes PH, *et al.* Response variability in the visual field: comparison of optic neuritis, glaucoma, ocular hypertension, and normal eyes. *Invest Ophthalmol Vis Sci* 2000;**41**:417–21.
19 **Spry PGD**, Johnson CA, McKendrick AM, *et al.* Variability components of standard automated perimetry and frequency-doubling technology perimetry. *Invest Ophthalmol Vis Sci* 2001;**42**:1404–10.
20 **Heijl A**, Lindgren G, Olsson J. Test retest variability in glaucomatous visual fields. *Am J Ophthalmol* 1989;**108**:130–5.
21 **Viswanathan AC**, Hitchings RA, Fitzke FW. How often do patients need visual field tests? *Graefes Arch Clin Exp Ophthalmol* 1997;**235**:563–8.
22 **Membrey W L**, Bunce C, Poinoosawmy DP, *et al.* Glaucoma surgery with or without adjunctive antiproliferatives in normal tension glaucoma: 2 Visual field progression. *Br J Ophthalmol* 2001;**85**:696–701.
23 **Membrey W L**, Bunce C, Poinoosawmy DP, *et al.* Comparison of visual field progression in patients with normal pressure glaucoma between eyes with and without visual field loss that threatens fixation. *Br J Ophthalmol* 2000;**84**:1154–8.
24 **Crabb DP**, Fitzke FW, Hitchings RA. Detecting gradual and sudden sensitivity loss in series of visual fields. In: Wall M, Wild JM, eds. *Perimetry update 1998/1999*. Amsterdam: Kugler, 1999:131–8.
25 **Spry PGD**, Bates AB, Johnson CA, *et al.* Simulation of longitudinal threshold visual field data. *Invest Ophthalmol Vis Sci* 2000;**41**:2192–200.
26 **Leske MC**, Heijl A, Hyman L, *et al.* Early manifest glaucoma trial—design and baseline data. *Ophthalmology* 1999;**106**:2144–53.
27 **Hutchings N**, Wild JM, Hussey MK, *et al.* The long-term fluctuation of the visual field in stable glaucoma. *Invest Ophthalmol Vis Sci* 2000;**41**:3429–36.
28 **Flammer J**, Drance SM, Zulauf M. Differential light threshold—short-term and long-term fluctuation in patients with glaucoma, normal controls, and patients with suspected glaucoma. *Arch Ophthalmol* 1984;**102**:704–6.
29 **Chauhan BC**, Johnson CA. Test-retest variability of frequency-doubling perimetry and conventional perimetry in glaucoma patients and normal subjects. *Invest Ophthalmol Vis Sci* 1999;**40**:648–56.

Want to extend your search?

Cross journal searching

If you can't find what you are looking for in the *British Journal of Ophthalmology* you can extend your search across many of the more than 200 journals available for selection. You can restrict your search to specific subject areas (eg, clinical medicine, basic research), or select specific journals, or search all available titles.

www.bjophthalmol.com

Examination of Different Pointwise Linear Regression Methods for Determining Visual Field Progression

Stuart K. Gardiner and David P. Crabb

PURPOSE. To compare the specificity and sensitivity of several different methods for using pointwise linear regression (PLR) to detect progression (deterioration) in visual fields.

METHODS. First, theoretical results were derived to predict which of the considered PLR methods would be the most specific and hence the least sensitive. Then, a "Virtual Eye" simulation model was developed that simulates series of sensitivity readings for a point over time. The model adds normally distributed noise (estimated from published results) to the sensitivity at each point to produce a series of fields to be analyzed using each method. Stable and deteriorating eyes were simulated, with the latter defined to have a noise-free loss of 2 dB/y at a significant cluster of points over the series.

RESULTS. The most sensitive method tested was to flag a visual field as progressing if it had a point that exhibited a statistically significant slope (at the 1% level) of at least −1 dB/y in the sensitivity. The most specific was a new "Three-Omitting" method that is being proposed, using two confirmation fields in a novel way. Current methods of using confirmation fields to verify a significant slope incorrectly flagged up to twice as many stable eyes as having progressing fields as did our new method.

CONCLUSIONS. Using the new proposed PLR method is recommended in preference to current PLR methods in any applications when a high degree of specificity is the main priority. (*Invest Ophthalmol Vis Sci.* 2002;43:1400-1407)

It has been well documented that one of the biggest clinical challenges when managing glaucoma is identifying whether the visual field is actually deteriorating over time, or remaining stable, with the variation between fields in a series being accounted for by the variability (noise) present in the measurement methods. Naturally, results from a set rule or automated computer program for detecting change will only partially contribute to a clinical management decision for a patient. However, when it comes to comparing, say, two different groups of patients who are undergoing, for example, different treatment regimens, then a benchmark for change is desirable to identify any difference in the amount of progression in the two groups.

Several different methods of determining visual field progression have been proposed, and there is no agreement about which is the best. Indeed, the method that is most suitable for use depends on the purpose to which the results are to be put. When the outcome of an eye's being labeled as progressing would be the patient's undergoing a risky or costly change in clinical management, a method should be used that is known to falsely label very few stable eyes as progressing (i.e., a high degree of specificity). Conversely, when failing to treat a progressing eye would be much more damaging than treating a stable eye, it is better to use a test that will correctly flag progressing eyes quicker (i.e., a more sensitive test). Clearly, the key is which way the method flags eyes in the middle ground that could be judged to be either progressing or stable; normally, tests that are more sensitive are less specific, and vice versa.

Automated visual fields are essentially made up of a grid of numerical sensitivity values, making them amenable to an assortment of statistical analyses. One set of methods for detecting visual field progression relies on estimates of change in summary measures of the field (so-called global indices) such as mean deviation (MD)[1] or visual field defect scores.[2] The advantageous simplicity of these methods is outweighed by the fact that they largely or completely ignore the detailed spatial information contained within computerized field tests, and they are reported to be insensitive to glaucomatous change.[3-5] Methods considering change in parts of the fields or at individual locations are more sensitive to change.[6,7] An example of this is known as glaucoma change probability (GCP), which examines the difference in threshold deviation at individual locations between a given field and baseline test results.[8]

Alternatively, pointwise linear regression (PLR) examines the sensitivity of each test location in the field against time over a patient's series. This provides a measure of the rate of loss (decibels per year) at each test location and a measure of the error associated with this change, summarized by the statistical significance or *P*-value. PLR, fully described elsewhere,[9] has been shown to be clearly more sensitive at identifying visual field loss than monitoring summary measures of the field[5-7] and compares favorably to GCP analysis at detecting and predicting progression.[10,11] PLR has also been found to agree more closely with expert clinical judgment about the status of progression than GCP analysis.[12] Moreover, in a study of untreated glaucomatous eyes, simple linear regression has been shown to perform better than polynomial regression in predicting deterioration of visual fields. The latter merely imitates the noise in the series of readings.[13] Furthermore, PLR has been used to demonstrate the benefits of treatment changes in normal-tension glaucoma,[14-16] and several different research groups have reported on the usefulness of the technique.[5-7,12,17-19]

In spite of the obvious claims of PLR's being a clinically useful tool for examining longitudinal visual field data, there is no consensus on what value of regression slope and *P*-value constitutes progression and whether it should be maintained in subsequent fields. Indeed, the latter idea of "confirmation fields" or "confirmation criteria" has been shown to improve the specificity of other methods for detecting progression in visual fields,[20,21] but has yet to be formally examined for PLR. Yet, such criteria are still used ad hoc to demonstrate the "benefits" of treatment changes in glaucoma.[16] In this study, we examined both the sensitivity and specificity of a selection of the different PLR criteria for confirming progression, and we

From The Nottingham Trent University, Nottingham, United Kingdom.

Submitted for publication September 7, 2001; revised December 27, 2001; accepted January 11, 2002.

Commercial relationships policy: N.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Corresponding author: David P. Crabb, Faculty of Science, The Nottingham Trent University, Clifton Campus, Nottingham, NG11 8NS, United Kingdom; david.crabb@ntu.ac.uk.

Investigative Ophthalmology & Visual Science, May 2002, Vol. 43, No. 5
Copyright © Association for Research in Vision and Ophthalmology

propose an improved method. Studies of deterioration of visual fields are hampered by the lack of a gold standard for progression and complications inherent in using patients' data; and so, in this article we offer a novel approach to this difficulty by comparing the PLR methods theoretically and follow this by using a purpose-written "Virtual Eye" simulation program exploiting newly published estimates of the variability (noise) inherent in visual field data.[22,23] Whether or not PLR is the best way of detecting progression (which is a widely debated question and one that has no firm answer at present), it is a widely used method, and as such any refinements to its methodology are to be welcomed.

Methods

Seven different PLR methods of determining whether a point of an eye is progressing or not, given the sensitivities at n points (equally spaced visual field tests in time) were compared—the purpose being to rank them according to their specificity and sensitivity. We define these as:

1. Standard Criteria: a point is flagged as progressing if it shows a significantly negative slope at the 1% level, together with an observed slope of at least −1 dB/y in sensitivity; this is written as $Z_n = 1$. The significance level is calculated by comparing the slope with the t-distribution, with $(n - 1)$ degrees of freedom. This is the simplest commonly used PLR method in published studies,[5-7,10-12,17-19,24,25] and it is used clinically as an indicator for change in some centers.[26]
2. Two of Two: a point is flagged as progressing if it satisfies the standard criteria and continues to satisfy them after addition of a further observed (confirmation) point: $Z_n = 1$ and $Z_{n+1} = 1$. This has been used by Hitchings et al.[27]
3. Three of Three: a point is flagged if it satisfies the standard criteria and continues to satisfy them after the addition of each of a further two observed (confirmation) points in turn: $Z_n = 1$, $Z_{n+1} = 1$, and $Z_{n+2} = 1$. This approach, applied to a form of GCP analysis, is being used in a clinical trial to evaluate the role of immediate intraocular pressure reduction in glaucoma.[28]
4. Two of Three: a point is flagged if it satisfies the standard criteria and continues to do so after the addition of either the following one or two points: $Z_n = 1$ and either $Z_{n+1} = 1$ or $Z_{n+2} = 1$ (or all three). This has recently been used in a study comparing PLR and the Advanced Glaucoma Intervention Study (AGIS) visual field score.[29]
5. Three of Four: a point is flagged if it satisfies the standard criteria and continues to do so with the addition of two of three successive points to the series: $Z_n = 1$ and at least two of $Z_{n+1} = 1$, $Z_{n+2} = 1$, and $Z_{n+3} = 1$. This method has been used in studies examining the treatment of normal-tension glaucoma.[15,16]

And our two new proposed methods:

6. Two-Omitting: a point is flagged if it satisfies the standard criteria and the slope obtained by adding one confirmation point, but excluding point n also satisfies the criteria. This is written as $Z_n = 1$ and $Z'_{n+1} = 1$.
7. Three-Omitting: a point is flagged if it satisfies the standard criteria, and the two slopes obtained by using points 1 to $(n - 1)$ and either point $(n + 1)$ or $(n + 2)$ both satisfy the criteria: $Z_n = 1$, $Z'_{n+1} = 1$, and $Z'_{n+2} = 1$.

The logic behind these new methods is that if the latest sensitivity reading in the series is significantly worse than preceding readings purely by chance because of the noise, then a confirmatory regression, performed once the following point has been added to the series, will be biased by this low-sensitivity reading, and so omitting the point will give a more conservative estimate of the true rate of progression. As seen in Figure 1, when the point is actually stable, there is less chance of its incorrectly being flagged as progressing—that is, fewer false-positive results are obtained.

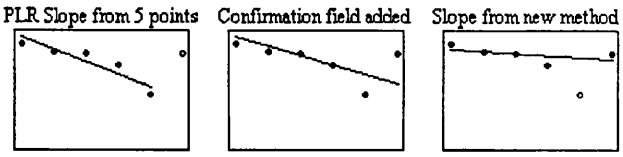


FIGURE 1. Graphs of sensitivities (y-axis) against time. After seeing the PLR slope based on the first five readings (the first graph), a confirmation field is sought to determine whether the series really is progressing (it clearly is not, given the next point). But the PLR slope including the sixth point (the second graph) is still flagged as progressing, because it is biased by the low fifth-point reading. Our Two-Omitting method (the third graph) solves this problem.

Assumptions for the Theoretical Approach

Several simplifying assumptions must be made to make the theoretical problems tractable. First, the noise is assumed to be normal: This is a reasonably accurate assumption but may be less true when the sensitivity is at the extremes of the range of measurements because of the censoring nature of the instrument. In other words, a point with a sensitivity of say 5 dB has positively skewed noise, because no readings below 0 dB are possible, whereas a stimulus seen at low levels, such as 36 dB, has a negatively skewed noise. The mean for this normal distribution is estimated from the readings taken so far. Second, it is assumed that the visual fields for each patient are taken at regularly spaced intervals. Both of these assumptions are in keeping with published work.[18]

Virtual Eye Simulation Model

We simulate sets of readings for a whole visual field in two circumstances:

Stable Eye. A Humphrey 30-2 visual field (Humphrey Instruments, San Leandro, CA) for a Virtual Eye was constructed and is shown in gray-scale form at the top of Figure 2. This left eye has an early-to-moderate inferior arcuate scotoma and some general depression of sensitivity in the superior field. This field has an abnormal Humphrey MD of approximately −4 dB. The individual pointwise sensitivities that make up this field are assumed to be the true physiological data for this eye. These are generally not identical with the measured or estimated data derived in a visual field test. Three typical examples of the latter, where noise has been added to the true visual field, are shown as grayscales at the bottom of Figure 2.

The key to any accurate visual field simulation is how well this noise or variability is estimated. The variability of standard threshold automated perimetry is complex, with it being suggested that it is made up of at least two components: intra- and intertest variability, sometimes referred to as short-term fluctuation (SF) and long-term fluctuation (LF), respectively. Our approach uses one estimate for the noise, dependent on the level of the true thresholded sensitivity according to Henson et al.,[22] and is based on frequency-of-seeing (FOS) data collected in patients and normal subjects at different visual field locations. This study demonstrated that the response variability was well represented by the function: $\log_e(\text{SD}) = A \times \text{sensitivity(dB)} + B$, where the constants A and B are −0.081 and 3.27, respectively. Hence, at each test location of the Virtual Eye, the noise is determined by independent random samples from the normal distribution, with the mean set at the true value for that location, and SD derived from the preceding function. For example, if the true sensitivity was 28 dB, then a measured sensitivity at that point is found by the Monte-Carlo simulated value drawn randomly from a normal distribution with mean 28 ± 2.72 dB (SD). This is repeated across the field to give a simulated noisy field. An entire visual field series for this eye, mimicking n tests per year over t years, $(nt + 1)$ fields in total, can be generated in this manner. Note that, even when the eye is stable, the visual fields at the



FIGURE 2. The effect of adding noise to the actual input sensitivity values for each point in a Virtual Eye. Visual fields are represented as 30-2 grayscales (Humphrey, San Leandro, CA), with darker areas indicating lower sensitivity readings at that point. The three noisy eyes generated are notably different, both from each other and the true physiological values represented by the top grayscale.

beginning and end of the series are not identical, because each has noise.

One extra feature of this stable Virtual Eye is that a normal age-related decline of 0.1 dB per year was subtracted from the true value in all cases. Returning to the numerical example, in which the true sensitivity is 28 dB, this means that a measured sensitivity at that point recorded 5 years after the first test is determined by the simulated value drawn randomly from a normal distribution with mean 27.5 ± 2.84 dB (SD). Note the increase in SD, imitating the established fact that variability is dynamic and increases as sensitivity declines.

Deteriorating Eye. The stable Virtual Eye forms the basis and starting point for the deteriorating eye. A progressive visual field defect is then added to sites in and around the initial inferior arcuate defect. More precisely, this means that six points (with starting sensitivities of 32, 28, 24, 20, 16, and 12 dB) in the visual field are given a rate of loss of 2 dB per year. The consequence of this magnitude of progression on the noise-free eye over a 6-year period is illustrated in Figure 3 (naturally, the simulation subsequently adds noise to each point to generate visual field series).

The visual field at 6 years would have a Humphrey MD of approximately −5 dB. This level of deterioration could be considered clinically moderate, and, with noise added in, it is doubtful that progression would be diagnosed by visual inspection or monitoring changes in global indices. Nevertheless, the cluster of points are decaying at more than 10 to 20 times age-related data,[30,31] and the threat of the progressing defect to the functionally important area of fixation means that it is precisely the type of progression that we suggest any decent pointwise method should detect with reasonable sensitivity and specificity. As with the stable eye, the important feature of this type of simulation is the noise, and this is added to deteriorating points in the visual field as before. For example, the progressing point at −3°, + 3° on the Humphrey visual field starts with a true value of 32 dB. The measurement at this point is imitated by the simulated value drawn randomly from a normal distribution with a mean 32 ± 1.97 dB (SD). After a period of follow-up, say 3 years, the true value at this point is 26 dB and the simulated measurement is arrived at by the value drawn randomly from a Normal distribution with mean 26 ± 3.20 dB. Visual field series for this deteriorating eye can be generated in a manner similar to those for the stable eye.

Another feature of note for both the stable and deteriorating eye is a simple correction for the censored nature of the dB values. That is, if a generated value is below 0 dB or above 36 dB, then it is assigned to be 0 or 36 dB, respectively. Also, the dB values are all rounded to the nearest integer (as they would appear from a Humphrey perimeter). The Virtual Eye simulation program was purpose written in object-orientated statistical software (S-PLUS 2000 for Windows; StatSci Europe, MathSoft Inc., Oxford, UK). The visual fields are represented, stored, and output as numerical arrays on which further analysis can be performed (including the production of Humphrey-type grayscales as shown in Figs. 2 and 3).

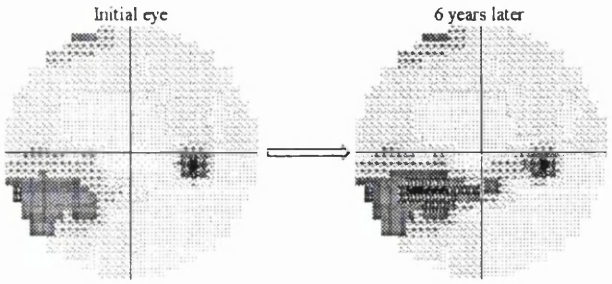


FIGURE 3. The progression in the noise-free eye, as used in the Virtual Eye simulation.

Simulation Experiment

One thousand series from the stable eye and 1000 series from the progressing eye, all with two tests per year over a 6-year follow-up (13 fields in total), were generated. Then PLR was applied to each of the 74 non-blind-spot locations within each 30-2 visual field series sequentially for each test starting from the 4th test (i.e., using the first 1.5 years of readings from the series) and finishing at the 13th test (using all 6 years of readings). Then, each criterion for determining progression was used in turn, and the eye labeled as progressing if one point in the field satisfied that particular criterion. The test at which the eye was first detected as having progressive disease was recorded. For those methods using at least two confirmation fields, it means that the fifth field tested was the earliest that progression could be flagged, and, for those using three confirmation fields, it means the sixth test was the earliest possible. Also note that if, for example, the Two-of-Two criterion is satisfied by using the 9th and 10th fields, then the visual field series was only labeled as progressing at the 10th field tested.

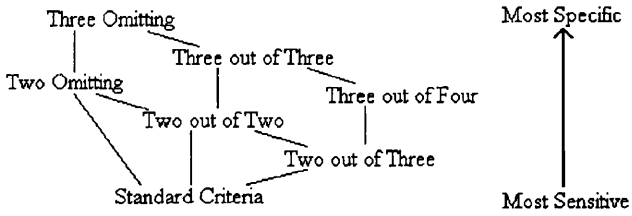


FIGURE 4. Theoretical relative specificity and sensitivity of the different PLR methods.

RESULTS

Theoretical Derivation

Because it is known that a more specific method of determining progression is always less sensitive if the level of noise is unchanged (methods of reducing the amount of noise are a separate issue entirely), it is only necessary to look at the specificity of each method—that is, how many stable eyes are incorrectly flagged as deteriorating. Suppose we have a series of visual fields, taken at times $t_1, t_2, \ldots t_n$. At a given point in the field, the deviations from the age-corrected normal sensitivities are $X_1, X_2, \ldots X_n$. Rather than the raw values, it is chosen to work with the deviations X_t below the age-corrected normal sensitivities. If the expected sensitivity for a patient of this age is S_E, then $X_t = S_E - S_t$. This means that at a stable point, the true (noise-free) deviation remains constant, and the X_t therefore all have the same distribution. If the point is progressing, the X_t increases over time.

According to our first assumption described earlier, we can say that for a stable eye, the X_t is normally distributed about a constant c: $X_t \sim N(c, \sigma^2)$. Note that the assumption is that the actual deviation of the point is constant, but we do not assume that the first or last readings are exactly equal (unlike Spry et al.[18]), because those two readings also have noise, and so the readings are not the same as the actual deviation in the eye. The constant c is estimated by $\hat{c}$—the mean of the readings taken so far—so that the sum of the first $(n - 1)$ readings is $n\hat{c}$. But the PLR slope is unaffected by the addition of a constant to each reading X_t; and so if we define $Y_t = X_t - \hat{c}$, we can perform PLR on the Y_t to produce the same results, where $Y_t \sim N(0, \sigma^2)$ and the sum $Y_1 + Y_2 \cdots + Y_{n-1} = 0$. Thus, the Y_t values are pure noise, because if there were no noise, Y_t would equal 0 for each $i = 1 \cdots n - 1$. We are also assuming that $t_i = i$ —that is, the readings have been taken at equally spaced time intervals.

Now, let β_n be the PLR slope based on the first n readings in the series, and with our new Two-Omitting method, the new PLR slope is θ_{n+1}. We are interested in the case in which X_n has been added to the series and has made the PLR slope significant, and the clinician wants to perform a confirmation test; therefore, we know that $\beta_n > \beta_{n-1}$. But then, according to the work fully described in the Appendix

$$E(\beta_{n+1} \text{ given } Y_1, \ldots Y_n) > E(\theta_{n+1} \text{ given } Y_1, \ldots Y_n)$$

Where $E(Z)$ is the expected (average) value of Z. Therefore in either case, the Two-Omitting method for PLR would be expected to produce a slope closer to the actual slope for a stable eye than the more standard Two-of-Two method. But also

$$E\left(\left|\frac{\theta_{n+1}}{se(\theta_{n+1})}\right| \; Y_1, \ldots Y_n\right) < E\left(\left|\frac{\beta_{n+1}}{se(\beta_{n+1})}\right| \; Y_1, \ldots Y_n\right)$$

Where $se(Z)$ is the standard error of Z. (Again, see Appendix for details). Thus, the slope θ_{n+1} would be expected to be less steep and also less significant than β_{n+1}, proving that our new Two-Omitting method is more specific than the current Two-of-Two method.

Naturally, it is straightforward to extend these two results to prove that the Three-Omitting PLR method is more specific than the Three-of-Three method.

It is obvious (and easily provable) that the Three-of-Four methods flags more points as progressing than Three-of-Three, because any series that satisfies the criteria for the latter method necessarily satisfies the criteria for the former method. We can therefore sum up all the information in Figure 4, where when a line connects two methods, it means that the upper method is known from theoretical results to be more specific than the lower method.

Simulation Model

The graphs in Figures 5 and 6 show the cumulative percentage of visual field series in the Virtual Eye (of 1000) that were flagged as progressing by each time point, using the specified methods. In each graph, the uppermost line flags most eyes as progressing, whether this is incorrect (for a stable eye; this method is then the least specific) or correct (for a deteriorating eye; this method is the most sensitive). For example, Figures 5 and 6 show that, after 3 years of follow-up, the Standard Criteria had already falsely labeled 74.6% of the stable visual field series as progressing. By 6 years, virtually all the stable eyes had been falsely labeled as progressing. In contrast, by 6 years, the Three-Omitting method had falsely labeled just 11.6% of the stable eyes as having progressive disease. The sensitivity (proportion of correctly labeled progressing eyes) of the Standard Criteria and the Three-Omitting methods after 6 years were 97.1% and 65.7%, respectively (a significant but much smaller difference).

Of those methods that use at least two confirmation fields (i.e., the three lower lines in Fig. 6), the Three-Omitting

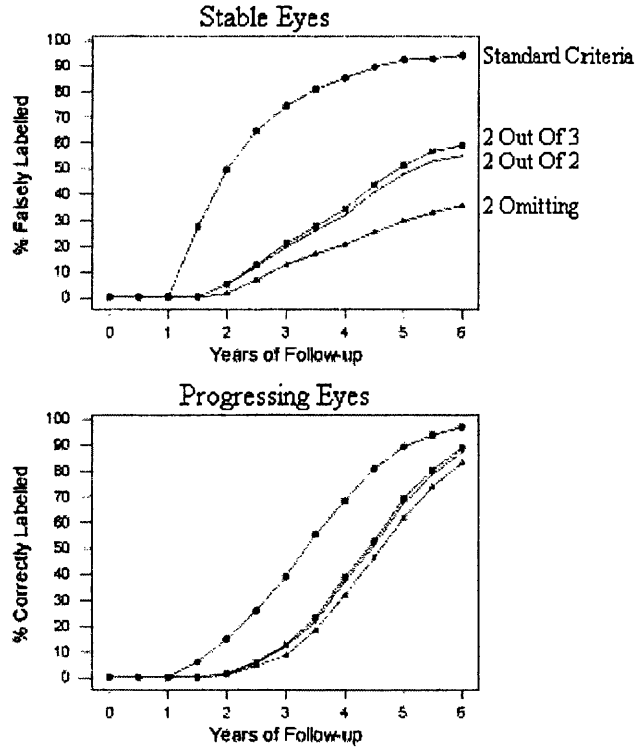


FIGURE 5. The relative specificity and sensitivity (respectively) of four PLR methods, as determined by the Virtual Eye simulation.

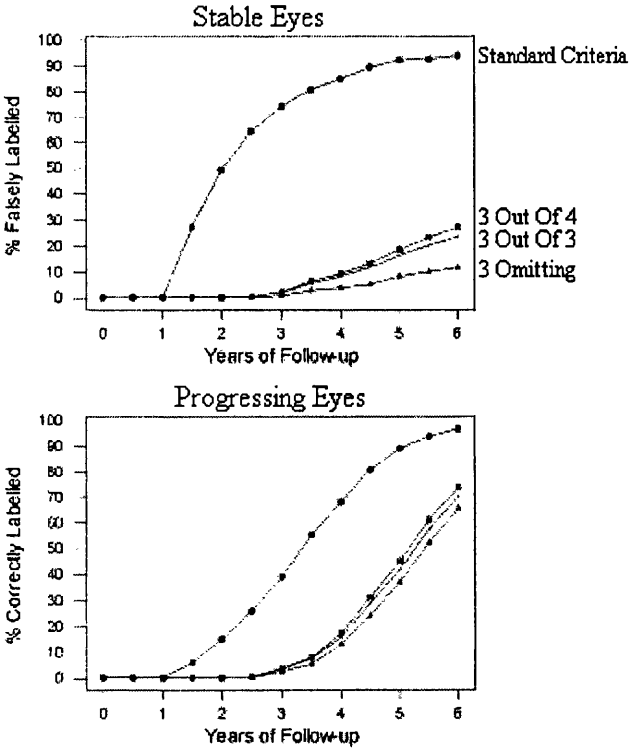


FIGURE 6. The relative specificity and sensitivity (respectively) of the remaining PLR methods (plus the Standard Criteria method again, for reference), as determined by the Virtual Eye simulation.

method affords a significant improvement in specificity, whereas the accompanying loss in sensitivity is much less marked.

Thus, it is clear that, as predicted theoretically, our new Two-Omitting and Three-Omitting methods are the most specific of those using one and two confirmation fields. The currently used methods can incorrectly flag twice as many stable eyes as having progressive fields. These two new methods improve the specificity of PLR, without having such a severe negative effect on the sensitivity.

DISCUSSION

When choosing a method for determining whether a visual field is progressing or stable, the specificity and sensitivity both have to be taken into account. PLR has previously been shown to be a sensitive technique for detecting progression of deterioration in visual fields in patients with glaucoma.[5-7,11,12] However, invoking so many hypothesis tests of the individual rates of loss (slopes) several times in a follow-up period using standard interpretation of the statistical type 1 errors (*P*-values) means that specificity is difficult to estimate and control.[32] Confirmation tests or fields require more examinations of patients but afford better specificity. Of the seven PLR methods we have examined in this article, the Standard Criteria method was the most sensitive (it successfully detected progressing eyes quicker), whereas the Three-Omitting method was the most specific (it wrongly identified the fewest stable eyes as progressing). This work has also highlighted that fulfilling the Standard Criteria for PLR at just one point in the visual field without any confirmation criteria is clinically unreliable as a means of diagnosing progression, because of its absurd levels of specificity.

Attempting to develop specific methods for detecting visual field progression has several benefits. There are 74 non-blind-spot points in the 30-2 Humphrey field, and therefore if, for example, the probability of each point's being flagged as progressing is 3%, then the probability of at least one point in the eye's being incorrectly flagged by the Standard Criteria method is

$$P(\text{at least 1 false positive}) = 1 - P(\text{no false positive points})$$

$$= 1 - P(\text{given point is not false positive})^{74}$$

$$= 1 - 0.97^{74} = 89.5\%$$

This alarmingly poor performance is improved dramatically by a seemingly small improvement in the specificity at one point, as seen in Table 1.

Thus, there are substantial gains in specificity to be had by using our new proposed *Omitting* PLR methods. The other priority is to improve the sensitivity of the test. All the methods using more than one confirmation field exhibit only modest levels of sensitivity. Of course, this would improve if the rate of loss were greater or more widespread than that described by the progressing defect ascribed to our deteriorating Virtual Eye. It would also improve if the amount of noise in the measurements decreased. New perimetric testing strategies such as the Swedish interactive test algorithm (SITA; Humphrey)[33] are primarily designed to offer reduced testing time, rather than directly reducing this noise. Alternative types of perimetry, exhibiting less noise, may be better equipped for detecting visual field progression.[23,34]

One of the reasons that there is little consensus about which method of visual field detection offers the best specificity and sensitivity in diagnosing true deterioration is the lack of an independent gold standard. Agreement on progression between expert clinicians is poor[35,36] or moderate[12,37] at best, and this precludes it from being a good arbiter of different methods. In this article we offer an alternative approach. We have separated the diagnostic precision of each PLR criterion under examination by using statistical theory. This demonstrated that our newly proposed techniques offer higher levels of specificity, which may appear surprising because they use less of the data (PLR is applied to one field less than is available). A computer simulation of a stable and deteriorating Virtual Eye, which shows the improved specificity of the new techniques, confirms this result.

Computer simulations are used as an alternative to analyzing the actual data of patients to examine data acquisition in automated perimetry.[33,38-41] It seems natural, therefore, to use them on the difficult problem of assessing visual field series. Spry et al.[18] designed a simulation of progressive and

TABLE 1. The Dramatic Effect of a Relatively Small Increase in Specificity

Prob (Given Point FP)	Prob (≥1 FP in 24-2 Field)	Prob (≥1 FP in 30-2 Field)
3	79.5	89.5
2	65.0	77.6
1	40.7	52.5

A small probability of an individual point's being falsely labeled as progressing (FP) leads to a high probability of at least one point in the eye's (52 points for a 24-2 Humphrey field, 74 points for a 30-2 Humphrey field) being FP. Data are percentages. Prob, probability.

stable sequences of data by interpolating between two "real" measured fields. Their model had several parameters controlling the noise in the data, including values for SF, LF, and eccentricity-related fluctuation. Only one measure of noise, directly related to the pointwise sensitivity, was used in our Virtual Eye, based on that reported by Henson et al.[22] The latter reported the noise (measured by accurate FOS techniques) to be independent of stimulus eccentricity, and therefore we have not included a parameter for this in our model. It must be stipulated, however, that the estimates of noise used in our Virtual Eye are unlikely to be perfect. For example, the noise is expected to have a component based on the spatial configuration of a defect within a visual field, and work is under way to estimate this.

Statistically speaking, there is very little justification for classifying a point within a visual field as either stable or deteriorating. The rationale behind such a distinction is that if a defect containing one or more points is progressing, then treatment is needed. A clinical decision of whether to treat is based on much more complex factors than whether the slope satisfies one strict criterion. Although we have used a significant slope of −1 dB/y to indicate progression, in accordance with published studies, this figure is, in essence, arbitrary, as is describing a typical progressing point as being one with a deterioration of −2 dB/y. Any software used clinically for the analysis of visual field data should allow users to alter the level of slope required for a point to be flagged, according to their clinical judgment based on other factors. Some software has this feature.[9] Also, when comparing two treatments, for example, it may be better to compare the distribution of slopes of points, rather than to compare perfunctorily the proportion of slopes that satisfy such criteria.

This study has been confined to a consideration of different PLR methods. No comparison is made against nonlinear or non-pointwise methods. There is disagreement about whether PLR really is the best method of determining progression, but it is a commonly examined method,[6,7,9-19,24-27] and, as such, any potential refinements to the method should be considered. Moreover, PLR may be suitable for examining any deepening of an existing defect, but its effectiveness in examining enlargement of defects is more open to question. Potential disadvantages of PLR are not addressed by the present study. However, the conclusion that omitting techniques may be of benefit to regression methods has wider applications than the limited set of conditions used herein. For example, different levels of noise, different defect sizes, and different rates of progression would all affect the quantitative results in Figures 5 and 6, but they would not affect the qualitative comparisons between the methods. The simulation should be viewed as an example, which supports the theoretical comparisons between the methods.

Another way of evaluating the new method would be to use real data from patients. However, when the actual state of the eye is unknown (as in the case of real data), there is currently no gold standard against which to judge different techniques. A further advantage of the techniques used herein is that they are not limited to glaucomatous eyes. Thus, the methods developed and the principles behind confirmation techniques could be applied in other situations in related fields. Nevertheless, before any methods are recommended for widespread use, their usefulness and practicality would have to be tested in a clinical situation.

In the clinical setting, the interpretation of any method remains a subjective matter, but using the approach described in this article provides decent estimates of sensitivity and specificity for particular methods, arming the clinician to make more informed management decisions for patients than with other PLR methods. It is very clear from these results how much specificity is gained by adopting a confirmation approach. We have demonstrated that (when the criteria of −1 dB/y and a 1% significance level are fixed) any gain in specificity when the PLR method is changed will be accompanied by a loss of sensitivity. But because of the disproportionate effect of a small change in specificity at each point, as demonstrated in our results, and because progressing eyes may often form a small proportion of the study population, seeking out specific methods is generally more important than sensitivity. For this reason, we suggest considering the new Three-Omitting method as an alternative to current confirmation methods applied to PLR.

References

1. Heijl A, Lindgren G, Olsson J. A package for the statistical analysis of visual fields. *Doc Ophthalmol Proc Ser.* 1987;49:153-168.
2. AGIS Investigators. The Advanced Glaucoma Intervention Study (AGIS). 2: visual field test scoring. *Ophthalmology.* 1994;101: 1445-1455.
3. Chauhan BC, Drance SM, Douglas GR. The use of visual field indices in detecting changes in the visual field in glaucoma. *Invest Ophthalmol Vis Sci.* 1990;31:512-20.
4. Katz J, Congdon N, Friedman DS. Methodological variations in estimating apparent progressive visual field loss in clinical trials of glaucoma treatment. *Arch Ophthalmol.* 1999;117:1137-1142.
5. Birch MK, Wishart PK, O'Donnell N. Determining progressive field loss. In: Mills RP, Wall M, eds. *Perimetry Update 1994/1995.* Amsterdam: Kugler & Ghedini; 1995:31-36.
6. Smith SD, Katz J, Quigley HA. Analysis of progressive change in automated visual fields in glaucoma. *Invest Ophthalmol Vis Sci.* 1996;37:1419-1428.
7. Katz J, Gilbert D, Quigley HA, Sommer A. Estimating progression of visual field loss in glaucoma. *Ophthalmology.* 1997;104:1017-1025.
8. Heijl A, Lindgren G, Lindgren A, et al. Extended empirical statistical package for evaluation of single and multiple fields in glaucoma: Statpac 2. In: Mills RP, Heijl A, eds. *Perimetry Update 1990/1991.* Amsterdam: Kugler & Ghedini; 1991:303-315.
9. Fitzke FW, Hitchings RA, Poinoosawmy DP, McNaught AI, Crabb DP. Analysis of visual field progression in glaucoma. *Br J Ophthalmol.* 1996;80:40-48.
10. McNaught AI, Crabb DP, Fitzke FW, Hitchings RA. Visual Field Progression: comparison of Humphrey Statpac2 and pointwise linear regression analysis. *Graefes Arch Clin Exp Ophthalmol* 1996;234:411-418.
11. Viswanathan AC, Fitzke FW, Hitchings RA. Early detection of visual field progression in glaucoma: a comparison of Progressor and Statpac2. *Br J Ophthalmol.* 1997;81:1037-1042.
12. Nouri-Mahdavi K, Brigatti L, Weitzman M, Caprioli J. Comparison of methods to detect visual field progression in glaucoma. *Ophthalmology.* 1997;104:1228-1236.
13. McNaught AI, Crabb DP, Fitzke FW, Hitchings RA. Modelling series of visual fields to detect progression in normal-tension glaucoma. *Graefes Arch Clin Exp Ophthalmol.* 1995;233:750-755.
14. Bhandari A, Crabb DP, Poinoosawmy D, Fitzke FW, Hitchings RA, Noureddin BN. Effect of surgery on visual field progression in normal tension glaucoma. *Ophthalmology.* 1997;104:1131-1137.
15. Membrey WL, Poinoosawmy DP, Bunce C, Fitzke FW, Hitchings RA. Comparison of visual field progression in patients with normal pressure glaucoma between eyes with and without visual field loss that threatens fixation. *Br J Ophthalmol.* 2000;84:1154-1158.
16. Membrey WL, Bunce C, Poinoosawmy DP, Fitzke FW, Hitchings RA. Glaucoma surgery with or without adjunctive antiproliferatives in normal tension glaucoma. 2: visual field progression. *Br J Ophthalmol.* 2001;85:696-701.
17. Wild JM, Hutchings N, Hussey MK, Flanagan JG, Trope GE. Pointwise univariate linear regression of perimetric sensitivity against follow-up time in glaucoma. *Ophthalmology.* 1997;104:808-815.
18. Spry PGD, Bates AB, Johnson CA, Chauhan BC. Simulation of longitudinal threshold visual field data. *Invest Ophthalmol Vis Sci.* 2000;41:2192-2200.

19. Hoffman DC, Westcott MC, Gaasterland D, Caprioli J. Comparison of methods to detect glaucomatous visual field progression [ARVO abstract]. *Invest Ophthalmol Vis Sci.* 2001;42(4):S152. Abstract nr 815.
20. Schulzer M, The Normal Tension Glaucoma Study Group. Errors in the diagnosis of visual field progression in normal tension glaucoma. *Ophthalmology.* 1994;101:1589-1595.
21. Chauhan BC, House PH, McCormick TA, LeBlanc RP. Comparison of conventional and high-pass resolution perimetry in a prospective study of patients with glaucoma and healthy controls. *Arch Ophthalmol.* 1999;117:24-33.
22. Henson DB, Chaundry S, Artes PH, Faragher EB, Ansons A. Response variability in the visual field: comparison of optic neuritis, glaucoma, ocular hypertension, and normal eyes. *Invest Ophthalmol Vis Sci.* 2000;41:417-421.
23. Spry PGD, Johnson CA, McKendrick AM, Turpin A. Variability components of standard automated perimetry and frequency-doubling technology perimetry. *Invest Ophthalmol Vis Sci.* 2001;42:1404-1410.
24. Viswanathan AC, Hitchings RA, Fitzke FW. How often do patients need visual field tests? *Graefes Arch Clin Exp Ophthalmol.* 1997;235:563-568.
25. Viswanathan AC, McNaught AI, Poinoosawmy D, et al. Severity and stability of glaucoma: patient perception compared with objective measurement. *Arch Ophthalmol.* 1999;117:450-454.
26. Kamal D, Hitchings R. Normal tension glaucoma: a practical approach. *Br J Ophthalmol.* 1998;82:835-840.
27. Hitchings RA, Migdal CS, Wormald R, Poinooswamy D, Fitzke FW. The primary treatment trial: changes in the visual field analysis by computer-assisted perimetry. *Eye.* 1994;8:117-120.
28. Leske MC, Heijl A, Hyman L, Bengtsson B. Early manifest glaucoma trial: design and baseline data. *Ophthalmology.* 1999;106:2144-2153.
29. Westcott MC, Hoffman D, Gaasterland D, Viswanathan AC, Caprioli J. Agreement between pointwise linear regression and AGIS score change in identifying visual field deterioration in glaucoma [ARVO abstract]. *Invest Ophthalmol Vis Sci.* 2001;42(4):S560. Abstract nr 3005.
30. Heijl A, Lindgren G, Lindgren A. Normal variability of static perimetric threshold values across the central visual field. *Arch Ophthalmol.* 1991;105:1544-1549.
31. Spry PGD, Johnson CA. Senescent changes of the normal visual field: an age-old problem. *Optom Vis Sci.* 2001;78:436-441.
32. Perneger TV. What's wrong with Bonferroni adjustments. *BMJ.* 1998;316:1236-1238.
33. Bengtsson B, Olsson J, Heijl A, et al. A new generation of algorithms for computerized threshold perimetry, SITA. *Acta Ophthalmol.* 1997;75:368-375.
34. Johnson CA, Cioffi G, Van Buskirk. Frequency doubling technology using a 24-2 stimulus presentation pattern. *Optom Vis Sci.* 1999;76:571-581.
35. Werner EB, Bishop KI, Koelle J, et al. A comparison of experienced clinical observers and statistical tests in detection of progressive visual field loss in glaucoma using automated perimetry. *Arch Ophthalmol.* 1988;106:619-23.
36. Viswanathan AC, Crabb DP, Westcott MC, McNaught AI, Fitzke FW, Hitchings RA. Inter-observer agreement on visual field progression in glaucoma: A comparison of methods [ARVO abstract]. *Invest Ophthalmol Vis Sci.* 1997;38(4):S571. Abstract nr 2662.
37. Chauhan BC, Drance SM, LeBlanc RP, Lieberman MF, Mills RP, Werner EB. Technique for determining glaucomatous visual field progression by using animation graphics. *Am J Ophthalmol.* 1994;118:485-91.
38. Shapiro LR, Johnson CA, Kennedy RL. KRAKEN: A computer simulation procedure for static, kinetic, supra-threshold and heuristic perimetry. In: Heijl A, ed. *Perimetry Update.* Amsterdam: Kugler and Ghedini; 1989:431-438.
39. Johnson CA, Chauhan BC, Shapiro LR. Properties of staircase procedures for estimating thresholds in automated perimetry. *Invest Ophthalmol Vis Sci.* 1992;33:2966-2974.
40. Glass E, Schaumberger M, Lachenmayr BJ. Simulations for FASTPAC and the standard 4-2 dB full threshold strategy of the Humphrey field analyser. *Invest Ophthalmol Vis Sci.* 1995;36:1847-1854.
41. Spenceley SE, Henson DB. Visual field test simulation and error in threshold estimation. *Br J Ophthalmol.* 1996;80:304-308.

APPENDIX

When the first $(n-1)$ readings have been taken, the PLR slope is β_{n-1}. We are interested in the effect that X_n (and hence, Y_n) has on β_n. For a pointwise linear least-squares regression, the slope is calculated by the equation

$$\left(\sum t_i^2 - \frac{(\sum t_i)^2}{n}\right)\beta_n = \sum t_i X_i - \frac{\sum t_i}{n}\sum X_i$$

Because $t_i = i$ in this case, and using the substitution $Y_i = X_i - c$, we can simplify this to

$$\frac{1}{12}n(n+1)(n-1)\beta_n = \sum iY_i - \frac{n+1}{2}\sum Y_i$$

Next, using our new Two-Omitting method, we can calculate the new PLR slope θ_{n+1} from the equation

$$\frac{1}{12n}(n^2-1)(n^2+12)\theta_{n+1} = \sum_{i=1}^{n-1} iY_i + (n+1)Y_{n+1} - \left(\frac{n+1}{2}+\frac{1}{n}\right)\left(\sum_{i=1}^{n-1} Y_i + Y_{n+1}\right)$$

We are interested in the case in which X_n has been added to the series, and has made the PLR slope significant, and the clinician wants to perform a confirmation test. Therefore, we know that $\beta_n > \beta_{n-1}$ or, equivalently,

$$\frac{(n-2)(n-1)}{2}Y_n > 3\sum_{i=1}^{n-1} iY_i$$

From our definition of Y_i, given the first $(n-1)$ readings, we know that the expected values $E(Y_n) = E(nY_n) = E(Y_{n+1}) = E[(n+1)Y_{n+1}] = 0$, and remembering that $\sum_{i=1}^{n-1} Y_i = 0$, we see that when $\sum_{i=1}^{n-1} iY_i \geq 0$

$$E(\beta_{n+1} \text{ given } Y_1, \ldots Y_n) = \frac{12}{n(n+1)(n+2)}\left(\sum_{i=1}^{n} iY_i - \frac{n+2}{2}\sum_{i=1}^{n} Y_i\right)$$

$$= \frac{12}{n(n+1)(n+2)}\left(\sum_{i=1}^{n-1} iY_i - \frac{n+2}{2}\sum_{i=1}^{n-1} Y_i + \left(n - \frac{n+2}{2}\right)Y_n\right)$$

$$= \frac{12}{n(n^2-1)}\left(\frac{n-1}{n+2}\sum_{i=1}^{n-1} iY_i + \frac{(n-2)(n-1)}{2(n+2)}Y_n\right)$$

$$> \frac{12}{n(n^2-1)}\left(\frac{n-1}{n+2}\sum_{i=1}^{n-1} iY_i + \frac{3}{n+2}\sum_{i=1}^{n-1} iY_i\right) = \frac{12n}{n^2(n^2-1)}\sum_{i=1}^{n-1} iY_i$$

$$\geq \frac{12n}{(n^2+12)(n^2-1)}\sum_{i=1}^{n-1} iY_i = E(\theta_{n+1}|Y_1, \ldots Y_n)$$

Also, because $\beta_n > 0$ (because the PLR slope satisfies the progression criteria), if $\sum_{i=1}^{n-1} iY_i < 0$ then $[(n-1)/2]\ Y_n > -\sum_{i=1}^{n-1} iY_i$ and in this case,

$$E(\beta_{n+1} \text{ given } Y_1, \ldots Y_n) = \frac{12}{n(n+1)(n+2)}\left(\sum_{i=1}^{n} iY_i - \frac{n+2}{2}\sum_{i=1}^{n} Y_i\right)$$

$$= \frac{12}{n(n+1)(n+2)}\left(\sum_{i=1}^{n-1} iY_i - \frac{n+2}{2}\sum_{i=1}^{n-1} Y_i + \left(n - \frac{n+2}{2}\right)Y_n\right)$$

$$= \frac{12}{n(n^2-1)(n+2)}\left((n-1)\sum_{i=1}^{n-1} iY_i + \frac{(n-2)(n-1)}{2}Y_n\right)$$

$$> \frac{12}{n(n^2-1)(n+2)}\left((n-1)\sum_{i=1}^{n-1} iY_i - (n-2)\sum_{i=1}^{n-1} iY_i\right)$$

$$= \frac{12n}{n^2(n^2-1)(n+2)}\sum_{i=1}^{n-1} iY_i \geq \frac{12n}{(n^2+12)(n^2-1)}\sum_{i=1}^{n-1} iY_i$$

$$= E(\theta_{n+1}|Y_1, \ldots Y_n)$$

Therefore, in either case, the Two-Omitting method for PLR would be expected to produce a slope closer to the actual slope for a stable eye than would the more standard Two-of-Two method.

The advantage to assuming the Y_i values to be normally distributed is that any linear combination of them is also normal. In particular, β_{n+1} and θ_{n+1} are, with variance

$$\text{Var}(\beta_{n+1}) = \frac{12\sigma^2}{n(n+1)(n+2)} \text{ and Var}(\theta_{n+1}) = \frac{12n\sigma^2}{(n^2+12)(n^2-1)}$$

where σ^2 is the variance of each Y_i (assumed earlier to be constant). And so for testing the significance of the PLR slopes

$$E\left(\left|\frac{\theta_{n+1}}{se(\theta_{n+1})}\right| |Y_1, \ldots Y_n\right) = \sqrt{\left(\frac{12n}{(n^2+12)(n^2-1)\sigma^2}\right)}\left|\sum_{i=1}^{n-1} iY_i\right|$$

$$< \sqrt{\left(\frac{12(n+1)(n+2)}{n(n-1)^2(n-2)^2\sigma^2}\right)}\left|\sum_{i=1}^{n-1} iY_i\right|$$

$$= \sqrt{\left(\frac{12}{n(n+1)(n+2)\sigma^2}\right)}\left|\sum_{i=1}^{n-1} iY_i + \frac{2}{(n-1)(n-2)}3n\sum_{i=1}^{n-1} iY_i\right|$$

$$< \sqrt{\left(\frac{12}{n(n+1)(n+2)\sigma^2}\right)}\left|\sum_{i=1}^{n-1} iY_i + nY_n\right|$$

$$= E\left(\left|\frac{\beta_{n+1}}{se(\beta_{n+1})}\right| |Y_1, \ldots Y_n\right)$$

Thus, the slope θ_{n+1} would be expected to be less steep and also less significant than β_{n+1}, which shows that our new Two-Omitting method is more specific than the current Two-of-Two method.

PERGAMON

Available online at www.sciencedirect.com
SCIENCE DIRECT®

Vision Research xxx (2003) xxx–xxx

Vision Research

www.elsevier.com/locate/visres

Reducing noise in suspected glaucomatous visual fields by using a new spatial filter ☆

Stuart K. Gardiner [a], David P. Crabb [a,*], Fred W. Fitzke [b], Roger A. Hitchings [c]

[a] *Faculty of Science, Nottingham Trent University, Clifton Campus, Nottingham NG11 8NS, UK*
[b] *Institute of Ophthalmology, London, UK*
[c] *Glaucoma Unit, Moorfields Eye Hospital, London, UK*

Received 11 December 2002; received in revised form 8 May 2003

Abstract

Visual field testing with automated perimetry is hampered by the amount of noise present in the readings. Here, we derive a physiologically accurate spatial filter to be applied to the data after patient examination. The filter was tested by a Virtual Eye computer simulation. By simulating series of stable fields it was shown that specificity of determining visual field changes was improved; while simulating progressing fields (based on a map of the optic nerve head) it was shown that sensitivity was also improved. The filter appears to reduce the noise in glaucomatous visual field data and may be clinically useful.
© 2003 Published by Elsevier Ltd.

Keywords: Visual field; Perimetry; Glaucoma; Optic nerve head; Statistics

1. Introduction

Reliable detection of glaucomatous visual field defects, and follow-up of current defects to determine whether they are spreading or deepening, is crucial for correct management of patients. Yet this is extremely difficult given the inaccuracy of threshold perimetry (in terms of the high inter-test and intra-test variability) and the various components of variability (or noise) associated with the perimetric process (Chauhan & House, 1991; Flammer, Drance, & Zulauf, 1984; Spenceley & Henson, 1996; Spry & Johnson, 2002; Wild, Searle, Dengler-Harles, & O'Neill, 1991). Therefore, it is sought to find ways of reducing the noise present in the readings. Reducing this noise is particularly relevant when detecting visual field progression (Spry & Johnson, 2002): true change in a glaucomatous visual field has to be larger than the noise before it becomes statistically distinguishable (Chauhan & Johnson, 1999; Heijl, Lindgren, & Olsson, 1989). Improved methods of data acquisition (Bengtsson, Olsson, Heijl, & Rootzén, 1997) have tended to focus on reducing the time of perimetric examination rather than making the measurement more accurate (Artes, Iwase, Ohno, Kitazawa, & Chauhan, 2002; Wild, Pacey, O'Neill, & Cunliffe, 1999). Carrying out more readings during the testing procedure is also clearly undesirable. Far better would be to process the data in such a way that the noise would be reduced, without any additional testing time. One way of doing this relies on exploiting the relationships between the actual sensitivities of different points; in essence, if one point has a reduced sensitivity, then its neighbours are more likely to also have reduced sensitivities. This principle points towards spatial filtering of the data as a possible solution.

Spatial filtering is a widely used image processing technique used to improve the quality of digital information. Spatial filtering applied to perimetric threshold sensitivity values using a Gaussian filter (fully described elsewhere Crabb, Edgar, Fitzke, McNaught, & Wynn, 1995; Fitzke, Crabb, McNaught, Edgar, & Hitchings, 1995, and illustrated in Fig. 1) has been show to reduce test–retest variability and reduce measurement noise (Crabb, Fitzke, McNaught, Edgar, & Hitchings, 1997). Moreover, benefits from post-test filtering of the data

☆ This work has not been published elsewhere and is not under review with another journal. If published in *Vision Research* it will not be reprinted elsewhere in any language in the same form without the consent of the publisher, who holds the copyright.

* Corresponding author. Tel.: +44-115-848-3275; fax: +44-115-848-6690.
E-mail address: david.crabb@ntu.ac.uk (D.P. Crabb).

0042-6989/$ - see front matter © 2003 Published by Elsevier Ltd.
doi:10.1016/S0042-6989(03)00474-7

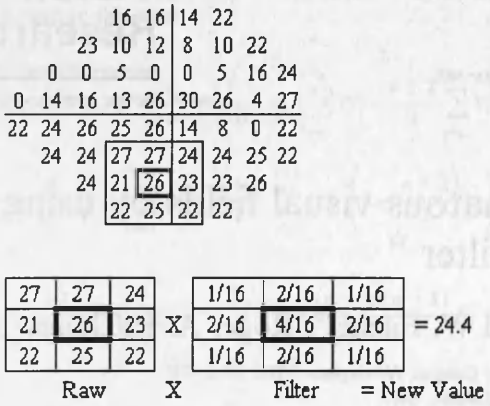


Fig. 1. How the new sensitivity is calculated by the Gaussian filter. In this method, the raw sensitivity value is replaced by one derived from a linear combination of the sensitivities at the nine points in a square centred on the point of interest. This is repeated for each point in the field in turn, each time looking at the points in a square surrounding the point of interest.

are accrued without any extra patient testing or alteration to the perimetric process.

However, there are problems with the basic Gaussian filter when applied to perimetric data. The visual field test locations, unlike say pixel values in digital images, are not physiologically linked in a 'grid-like' fashion as indicated by the matrix of values on the visual field chart and the Gaussian filter takes no account of the actual anatomical structure of the optic disc. For example, glaucomatous defects are generally limited to either the superior hemifield or the inferior hemifield; they rarely cross the horizontal meridian that passes through the fovea (Lachenmayr & Vivell, 1993). Yet the Gaussian filtered value at a point adjacent to this meridian is based in part on values from the opposite hemifield. There is general (albeit vague) agreement about the shape of the nerve fibre layer, and that the shapes of glaucomatous defects follow the nerves; the Gaussian filter ignores this information. It has been reported that applying the Gaussian filter to data may reduce the proportion of false positives (i.e. the number of healthy points flagged as being part of a defect), but it also reduces the detection of true positives, blurring out true defects (Spry, Johnson, Bates, Turpin, & Chauhan, 2002). This study aims to derive a filter taking into account the true structure of the retinal nerve fibre layer, that will reduce the noise without obscuring true defects.

2. Methods

The work is based on a visual fields database consisting of patients seen by the Moorfields Eye Hospital Glaucoma Service in London, UK. This means that the filter is based on, and designed for use in, a tertiary glaucoma referral service. It would be feasible to utilise the methods below to produce a filter for different situations, such as population screening events or general ophthalmic clinical settings; however the process would be hindered by the increased proportion of non-glaucomatous defects. It would be expected that the results would be extremely similar, although obviously that cannot be definitively assumed without testing. The database contains 98,821 visual fields, taken from 14,675 individual suspected glaucomatous patients. Fields were measured using the Humphrey Visual Field Analyser (Humphrey Instruments Inc., Dublin, California, USA). The data goes back as far as 1985; it consists of both 30-2 fields and more recently 24-2 fields (all standard white-on-white, full threshold tests), although only complete 24-2 fields (the non-edge points in 30-2 fields) were used. This database, consisting of around 5 million individual threshold values, provides an excellent resource giving a comprehensive and representative cross-section of all tests carried out at glaucoma clinics. The database was cleaned to remove duplicated visual fields; and each patient given a unique ID number. This way, the data was completely anonymised. Visual fields from right eyes were transposed, and their mirror images used instead as 'left' eyes.

2.1. Deriving the filter

When deriving the new filter, no patients were excluded. It was sought to use as representative a sample as possible of all the patients entering the clinic, on an "intention-to-treat" basis. A filter derived purely on data from glaucomatous patients may not be accurate for healthy eyes, and vice-versa; so all available data was used in the belief that this gives added weight to our results. This means that the results are truly representative of the average test carried out in a glaucoma clinic.

It is important to note that unreliable fields, or fields from patients presenting at the glaucoma clinic who turn out to have other pathologies (e.g. pituitary tumours, myelinated nerve fibres etc.), were not removed from the database prior to analysis. This is because we seek to represent the typical test carried out at a glaucoma clinic, and so these patients should not be ignored. The principle of "intention-to-treat" is well founded, and is very commonly used throughout the scientific and statistical literature. A further important distinction should be drawn between using every available test from each patient (making the results representative of the average visual field test carried out in a glaucoma clinic) and using one randomly selected field from each patient (which would make the results representative of the average patient seen by a glaucoma clinic). Since patients with true glaucoma would be expected to have more fields in the database than patients with other

pathologies, and since it is more important to deal with patients with glaucoma who make repeated visits to the clinic, the choice was made to use all the fields from each patient.

First, the covariances and correlations between sensitivities at each pair of points in the eye were found. For the covariance between sensitivities at points A and B,

$$\mathrm{Cov}(S_A, S_B) = E(S_A S_B) - E(S_A)E(S_B) \quad (1)$$

where $E(S_A)$ is the expected (mean) value of the sensitivity at A, etc. Now, suppose we can accurately predict the sensitivity at point A by a linear combination of the sensitivities at all the other points

$$\widehat{S}_A = k_1 S_1 + k_2 S_2 + k_3 S_3 + \cdots \quad (2)$$

for some constants $k_1, k_2, \ldots$; then from Eq. (1),

$$\mathrm{Cov}(S_A, S_B) = \sum_i k_i \mathrm{Cov}(S_i, S_B) \quad (3)$$

We choose to perform regressions on the covariances in Eq. (3) rather than on the basic sensitivities in Eq. (2). This is because an apparent relationship between two points, X and Y say, may actually be the product of relationships between each of these points and some third point Z; looking at covariances takes account of this second-order of complexity. Before performing multiple regressions to find these coefficients k_i for each point A, a few constraints were placed on the regressions. Firstly, each k_i had to be non-negative. Secondly, the expected predicted value must equal the expected raw value for the filter to make sense; so

$$E(\widehat{S}_A) = \sum_i k_i E(S_i) = E(S_A)$$

Finally, if the correlation between the sensitivities at point A and point i was low (less than 0.7), k_i was set to equal zero, as those points were considered to be unrelated (this was done mainly for computational reasons. In tests on small samples with fewer points it had no effect on the outcome; the regression still identified the same points as being good predictors with associated high values of k_i. In fact a small minority of the k_i were constrained to be zero because of this rule; far more were found to be zero after the regressions were carried out).

So, for each point A, regressions were performed on these series of equations (one equation for each point B) to produce constants k_i, and hence an algorithm for predicting the sensitivity at point A based on those at points elsewhere in the visual field.

The next stage is to produce the filtered value S_A^f for the sensitivity at point A; which should be a combination of the predicted sensitivity and the raw sensitivity at that point.

$$S_A^f = c\widehat{S}_A + (1-c)S_A \quad (4)$$

The value of c in Eq. (4) will vary according to the position in the eye; some points are more predictable than others, and these points should have a correspondingly high value of c, hence giving more weight to the predicted value. (Note that in the Gaussian filter, $c = 12/16$ for all points not adjacent to either the edge of the field or the blind spot, as shown in Fig. 1). Further, the more points that are used in the prediction (i.e. the greater the number of non-zero k_i), the more accurate this predicted value would be. Now, if the predictions were entirely accurate, the variance of the filtered value would be

$$\begin{aligned}\mathrm{Var}(S_A^f) &= \mathrm{Var}(c\widehat{S}_A + (1-c)S_A) \\ &= \mathrm{Var}\left(c\sum_i k_i S_i + (1-c)S_A\right) \\ &= \left(c^2\sum_i k_i^2 + (1-c)^2\right)\sigma^2 \end{aligned} \quad (5)$$

where σ^2 is the variance of the noise at each point, assumed for this to be constant throughout the eye. Now, we want to choose c to minimise this variability in the filtered value from Eq. (5), and so reduce the noise as much as possible. Mathematically, the minimum of a function occurs when its differential equals zero. So, differentiating

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}c}(\mathrm{Var}(S_A^f)) = \left(2c\sum_i k_i^2 - 2(1-c)\right)\sigma^2 = 0 \\ &\left(\sum_i k_i^2 + 1\right)\hat{c} = 1 \\ &\hat{c} = \left(\sum_i k_i^2 + 1\right)^{-1}\end{aligned} \quad (6)$$

So, we have a prescription from Eq. (6) for choosing c if the predicted values were completely accurate. However, this is clearly not the case; indeed, the predictions at some points will be closer to the raw values than elsewhere in the visual field. So, the correlations Corr_A between the predicted and raw values at each point A were found by calculating the predicted sensitivity at point A for each visual field in the database; so $\mathrm{Corr}_A = \mathrm{Corr}(S_A, \widehat{S}_A)$. This then provides a measure of the predictability of point A; the higher this correlation is, the more accurate the predicted sensitivity and so the more weight should be given to it in the filter. Now, $\hat{c}$ was found in Eq. (6) based on the assumption that the predicted values were 100% accurate; so in the final filter, the weighting given to the predicted sensitivity in the filter, is given by $c = \hat{c} \times \mathrm{Corr}_A$. So, having found the k_i for point A by multiple regressions (as described above), the final filtering algorithm for point A is

$$S_A^f = \frac{\mathrm{Corr}_A}{\sum_i k_i^2 + 1}\left(\sum_i k_i S_i\right) + \left(1 - \frac{\mathrm{Corr}_A}{\sum_i k_i^2 + 1}\right) S_A \quad (7)$$

Obviously, the k_i differ according to which point in the visual field is being considered (point A); and so the filter is, in effect, a matrix containing the coefficients for filtering each point in the field, coefficients which vary from point to point.

2.2. Testing the filter

Visual fields are typically followed over a number of years to determine whether a suspected glaucomatous defect is progressing or not. It is proposed that this feature could provide an indication of the effectiveness of the filter. Indeed, the Gaussian filter has previously been shown to improve the consistency over time of visual fields (Crabb et al., 1997), which is one important aspect of reducing the noise.

Significantly, however, it has previously been reported that although the Gaussian filter may reduce noise, it also reduces the signal (Spry et al., 2002), i.e. methods which use filtering will be less successful at detecting localised defects than methods based on the raw data. It was decided to investigate this by looking at the effect of filtering on simulated visual field series, using pointwise linear regression (PLR) as the testing tool. The simulated data was generated by a previously developed "Virtual Eye", fully described elsewhere (Gardiner & Crabb, 2002a, 2002b). Briefly, this produces series of visual fields with known properties (baseline sensitivity values, length of series, frequency of testing, rates of loss) and adds noise by sampling randomly from a normal distribution, mean zero and with a standard deviation which increases (hence increasing the amount of noise) as the underlying sensitivity decreases, following a result from Henson, Chaundry, Artes, Faragher, and Ansons (2000). Using this "Virtual Eye" simulation, series of six fields (from annual testing over a period of 5 years) were simulated based on input noise-free visual fields; the simulation adds random noise to each of the six fields in the series. It was sought to make these noise-free series as simple as possible, to simplify and clarify the results obtained. So, each point in the field was assigned as one of

1. *Stable*: the sensitivity remains constant at 30 dB throughout the series, before noise is added.
2. *Defective*: the noise-free sensitivity is reduced by 2 dB per year, from 30 to 20 dB over the 5 years.
3. *Border*: the noise-free sensitivity is reduced by 1 dB per year, from 30 to 25 dB.

This process was first carried out on a 'stable' visual field; i.e. one where all 54 points in the visual field (including the two points coinciding with the blind spot) were *Stable*. Next, localised defects were chosen with reference to the map of the physiological optic nerve head (ONH) locations for each visual field test point produced by Garway-Heath, Poinoosawmy, Fitzke, and Hitchings (2000). The rationale for using this map is that localised defects are more likely to occur in clusters determined by the anatomical entry position of nerve fibre bundles into the optic disc. This rationale partly satisfies the need for determining localised defects separate from one based on perimetric data alone, avoiding the use of purely arbitrary and subjective defects. Using the map, a 'Central Point' was chosen, and its location in the ONH noted. This point and all other points located within ±5 degrees of the Central Point in the ONH were assigned as being *Defective* points with a glaucomatous loss of 2 dB per year. Further points between 5 and 10 degrees away from the Central Point (in terms of ONH location) were assigned as being *Border* points, with a glaucomatous loss of 1 dB per year. Points elsewhere in the eye were assumed to be Stable. Each of the 54 points in the standard visual field was considered in turn as the Central Point, with the exceptions of the two points forming the blind spot. Anderson's criteria (Anderson & Patella, 1999) define a clinically significant progressing localised defect as being a cluster of three or more points that have sensitivities occurring in fewer that 5% of the normal population, at least one of which must occur in less that 1% of the population, according to the pattern deviation probability plot. Hence, localised defects generated in this way which were found to consist of fewer than three adjacent points, and therefore not constituting an identifiable defect according to Anderson's criteria, were separated out and will be commented further upon later (this comprised six 'defects' consisting of a solitary point, and six 'defects' of just two points). Some examples of defects generated in the manner described are shown in Fig. 2.

Finally, the process was carried out on a field with a non-glaucomatous defect (as shown in Fig. 3), consisting of three Defective points with one Border point at each end. This was done in the hope that a filter designed to identify glaucomatous defects will not pick up defects which would not occur in glaucoma. Such a defect may be genuine, caused by for example a neurological disease; but since the purpose of this work is to more readily identify the existence and progression of glaucomatous defects, and it is not designed for use when other suspected conditions are being looked for, it is desirable that such non-glaucomatous defects be blurred out.

Each input series was simulated 5000 times. After each run, points were tested to see whether they would be flagged as progressing using a widely used PLR criterion. PLR is fully described elsewhere (Fitzke, Hitchings, Poinoosawmy, McNaught, & Crabb, 1996; Katz,

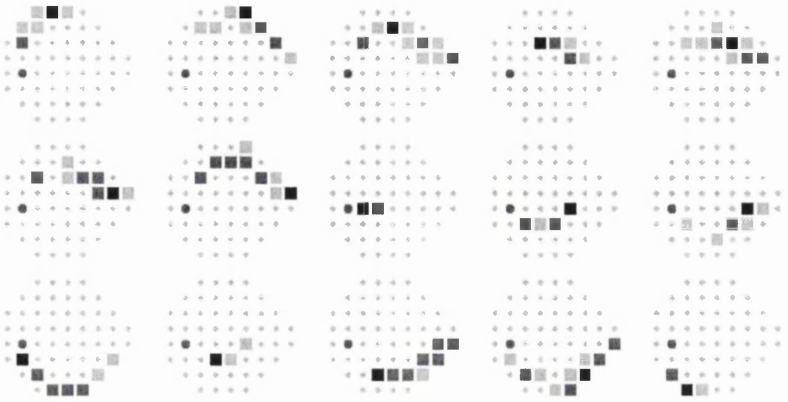


Fig. 2. The shapes of some of the 52 generated defects tested, superimposed onto a Humphrey 24-2 visual field. The black square in each case is the Central Point. Dark grey squares represent the Defective points (within 5 degrees of the Central Point in the ONH); light grey squares represent the Border points (between 6 and 10 degrees away from the Central Point in the ONH). The grey circle represents the location of the physiological blind spot.

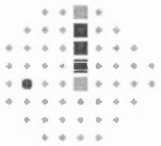


Fig. 3. The shape of the non-glaucomatous defect used, superimposed onto a Humphrey 24-2 visual field. As before, the grey circle represents the physiological blind spot.

Gilbert, Quigley, & Sommer, 1997; Smith, Katz, & Quigley, 1996; Wild, Hutchings, Hussey, Flanagan, & Trope, 1997) and is a form of change analysis whereby each individual field location is tested to find its pointwise rate of deterioration (slope) and a measure of statistical significance. The criteria for a progressive point used in these simulation experiments were a deterioration of at least 1 dB per year, statistically significant at the 1% level. This criteria has been used in published studies (Membrey, Bunce, Poinoosawmy, Fitzke, & Hitchings, 2001; Nouri-Mahdavi, Brigatti, Weitzman, & Caprioli, 1997; Viswanathan et al., 1999) and is examined in detail elsewhere (Gardiner & Crabb, 2002b). This analysis was repeated after filtering the noisy simulated series using each of the Gaussian and our new filter. So for each of the 5000 runs for each input series, the points were tested three times. In this way, the sample probabilities of the points being flagged as progressing based on each of the Raw, Gaussian and Filtered data were calculated. In the case of the input localised defects, the probabilities of the Central Point being (correctly) flagged as progressing in each case will hereafter be referred to as the three Detection Rates for that defect.

The computational derivation of the filter, the virtual eye simulation and additional statistical analyses were carried out using purpose written programs developed in S-PLUS for Windows (Insightful Corporation, Seattle, USA.).

3. Results

3.1. Shape of the filter

The filter conforms to the accepted physiological shape of the retinal nerve fibre layer. Fig. 4 shows the points that are used to filter a few Central Points (i.e. those points with a non-zero k_i when the relevant Central Point is 'point A' in the derivation above). If a point is connected to the Central Point by a line, it indicates that that point is a predictor for the Central Point; the thicker the line, the larger the effect it has on the prediction (i.e. the higher k_i is for that point). The remaining contribution to the filtered value comes from the Central Point itself. It is seen that predictors are not necessarily neighbours of the Central Point (as they would be if the Gaussian filter was being considered), but they follow the expected arcs. As seen in Fig. 1, this diagram for the Gaussian filter would simply connect each Central Point with all of its immediate neighbours.

Further, points on opposite sides of the horizontal meridian only turn out to be significant predictors on one occasion. This is very promising, since the derivation of the filter at no point took account of the relative

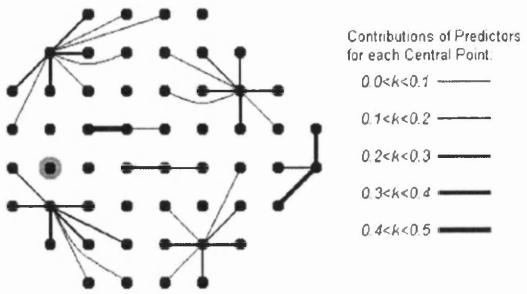


Fig. 4. An illustration of the shape of the filter. The grey area represents the blind spot.

positions of point in the visual field. The only location where a point does turn out to be a significant predictor of a point on the opposite side of the horizontal meridian is at the nasal step. This observation is interesting in itself, since nasal step defects are seen to cross the horizontal meridian in clinical practice. One possible explanation is the anatomical notion that some nerve fibres from upper and lower hemiretinas interdigitate at the temporal end of the horizontal raphe (Sakai, Kuniyoshi, Tsuzuki, Makoto, & Kawamura, 1987; Vrabec, 1966); alternatively, the phenomenon could be caused by local effects from the glaucomatous process (e.g. excitotoxicity) or a testing artefact. Also, the filter as derived above always assigns a sensitivity of zero to the point labelled in grey in Fig. 4, which corresponds to the blind spot (10–15 degrees temporal of the fovea, just below the horizontal meridian); again, this was not predetermined by the method of deriving the filtering algorithm.

3.2. Effect of the filter

Clearly, the noisier the initial (raw) visual field is, the more obvious to the naked eye will be the difference caused by filtering. Finding suitable visual fields which are noisy yet certainly have localised defects is near impossible. The effects of the filter are frequently subtler; although improvements undetectable when simply viewing greyscales of the data may have much larger effects over time when a computer analysis, such as PLR (Fitzke et al., 1996) or the Glaucoma Change Probability (Heijl et al., 1991), is used. However, there are cases when the benefits are clear for all to see. Fig. 5 shows a 24-2 greyscale from an eye, taken from a textbook example (Budenz, 1997), noted as having moderate to early glaucoma. There is a scotoma superior to the nasal horizontal meridian partially obscured by the noise present throughout the field. Filtering emphasises this defect, at the same time as removing most of the random noise elsewhere in the field. In contrast, the Gaussian filter almost entirely blurs out this defect. Note also that the new filter leaves the blind spot intact; the Gaussian filter may or may not do so, depending on the implementation.

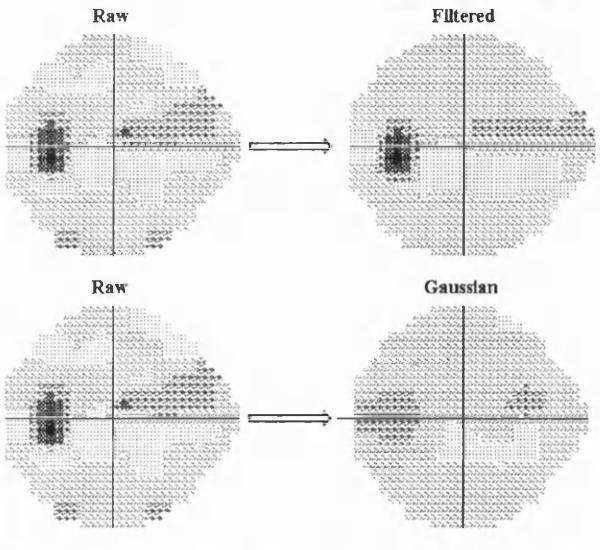


Fig. 5. The effect of the new filter on a visual field exhibiting a scotoma in the superior hemifield, compared with the effect of applying the Gaussian filter to the same data.

3.3. Virtual eye simulation

When the stable eye was tested (i.e. one where, before the addition of noise, all the points remain at a sensitivity of 30 dB throughout, as described above), the proportion of points still being flagged as progressing (all of which are therefore false positives) fell dramatically from 0.59% to 0.04% upon applying our new filter. This is further evidence that the filter is reducing the noise.

For each of the artificial localised defects based on the ONH structure (illustrated in Fig. 2), the percentage of the 5000 runs of the simulation which resulted in the Central Point satisfying the standard criteria for progression under PLR (i.e. a slope of at least −1 dB per year, statistically significant at the 1% level) was calculated. This percentage was found after the same 5000 noisy series had been filtered using the Gaussian filter (marked as 'Gaussian'), and using the new filter (marked as 'Filtered'), as well as before filtering (marked as 'Raw'). The results are shown in Fig. 6. The detection rates for the Central Points vary considerably after filtering, since they depend in part on the status of neighbouring points; without any filtering, the detection rates are consistently around 20% for the Central and Defective points (which are deteriorating at 2 dB per year) and around 8% for the Border points (which are deteriorating at 1 dB per year). It is seen that in almost all cases, the percentage of Central Points correctly identified as progressing increased significantly after filtering. Further, the new filter is performing much better than the Gaussian filter in this regard; it was common for defects to be blurred out by the Gaussian filter. The mean detection rate for the localised defects was 19.9% for the unfiltered, raw data; rising to 29.9% after application of the Gaussian filter, and 38.0% after using our new filter.

Of the 54 possible Central Points in the central visual field, 14 resulted in defects consisting of fewer than three adjacent points and so did not satisfy the criteria of Anderson and Patella (1999) (comprising the two points coinciding with the physiological blind spot, six solitary point 'defects', and six two point 'defects'). Because they did not produce realistic glaucomatous defects (according to this definition), results from these 14 Central

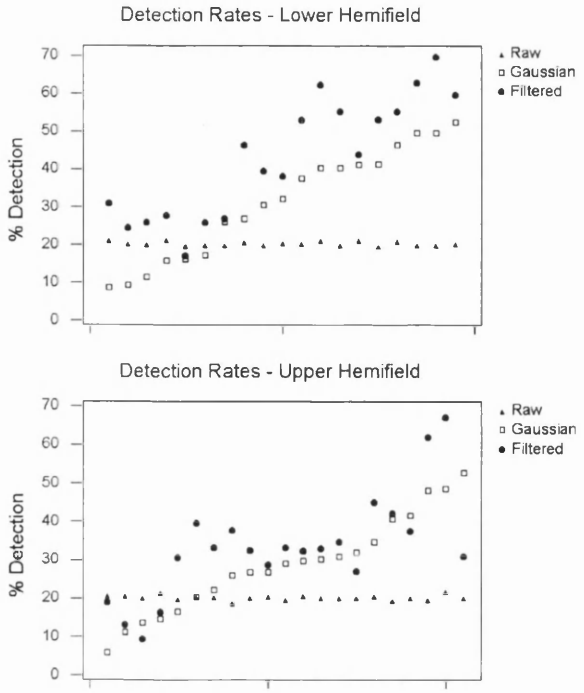


Fig. 6. Percentage of Central Points correctly flagged as progressing for each of the tested defects. Each column contains three symbols, representing the detection rates for the raw data, the data after application of the Gaussian filter, and the rate after application of our new filter; there is one column per point tested. The points have been split into the two hemifields. The points are in each case shown in order of increasing detection rate after application of the Gaussian filter (rather than in order of location), to make the figure clearer to understand.

Points are not included in Fig. 6. However, this does not necessarily mean that these small defects were not picked up. Indeed, the two-point localised defect in the centre of the field next to the blind spot, illustrated by the central field in Fig. 2, resulted in the detection rate increasing from 19.9% to 37.1% after filtering. Certainly, as with any type of spatial filtering, if a defect consists of only one solitary point its detection rate will decrease. However, one deteriorating location may not constitute clinical glaucomatous progression.

Of more concern are the five Central Points for which the detection rate in Fig. 6 appears to be reduced by filtering. These five defects are those shown in Fig. 7. In each case, the apparent reduction in the detection rate caused by filtering is explained by the differences in shape between the defect being tested (generated from the ONH map) and the shape of the filter at that point which is shown in the bottom row of Fig. 7 (generated from visual field data). A large proportion of the filtered value for the Central Point is based on points not appearing in the defect; over 50% for defects A, D and E, and between 40% and 50% for defects B and C. These five isolated cases appear to behave in this way because of the method used for generating simulated defects, and are not necessarily reason to doubt the usefulness of the filter. Furthermore, in the first three cases, the overall detection rate taken over all the points which are deteriorating at 2 dB per year (the Central Point, represented by the black square, and the Defective points represented by dark grey squares), rather than just the Central Point, has actually improved. With defect A, the average detection rate has in fact increased from 19.9% to 29.1% after filtering (Gaussian: 22.3%); although the Central Point itself is being flagged as progressing less frequently, the rest of the defect is more likely to be picked up, and so the overall detection rate has still improved. Similarly for defect B, where the average has increased from 19.6% to 21.5% (Gaussian: 23.6%); and for defect C, where the average has increased from 20.1% to 31.5% (Gaussian: 26.9%).

The non-glaucomatous defect shown in Fig. 3 was tested in the same way. Averaged over the three Defective points, the filter reduced the percentage of runs of the simulation which resulted in points being flagged as progressing from 19.7% to 10.0%; in effect blurring out non-glaucomatous defects, as hoped. The Gaussian filter did not have this beneficial effect; in fact it increased the proportion flagged from 19.7% to 29.6%.

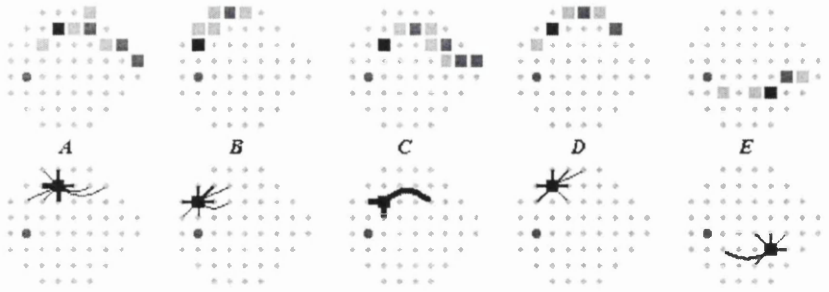


Fig. 7. The five generated localised defects whose detection rate appears to be reduced by filtering in Fig. 6, using the same symbols as in Fig. 2. The lower line shows the points involved in the filtering algorithm for that particular Central Point; the thicker the line, the more effect the point has on the filtered value for the Central Point.

4. Discussion

The principle of filtering real-world data is long established as a tool for reducing noise. Indeed, even in a standard measurement of the visual field with a Humphrey perimeter, double determinations of the sensitivity are carried out at certain points in the field; and while this is done principally to provide an estimate of the variability present in the readings, it also provides a basic form of filtering, as clinicians may base their judgment on the mean of these double determinations. The semi-Bayesian approach of using the prior expected sensitivity at a point to predict it before it is measured is also utilised in the SITA algorithm (Bengtsson et al., 1997). Spatial filtering, in the form of a Gaussian filter, has also been applied to visual field data with the benefit of reducing between test variability and improving the predictability of future visual field changes (Crabb et al., 1995, 1997; Fitzke et al., 1995). However, a recent study using a computer simulation of progressive glaucomatous visual field loss (Spry et al., 2002) concluded that Gaussian filtering does not offer a consistent benefit over the analysis of raw visual field data, and in some instances significantly inhibits the ability to detect small, gradual progressive field changes. In support of these findings, our testing of the Gaussian filter on our selection of localised defects also showed that, in some cases, it blurred out smaller progressive defects. With our new improved filtering algorithm, the detection of progressing points not only did not worsen, but it actually improved significantly, even in cases of small clusters of progressing points. The new filter also offers dramatic improvement in specificity when compared with unfiltered fields (reducing the false positive rate from 0.59% to 0.04%), vital for pointwise methods for detecting progression. Indeed, a seemingly small improvement in specificity may be more important than a larger improvement in sensitivity (Gardiner & Crabb, 2002b) (although here we are in the happy position of improving both!).

For any new method to become widespread in clinical use, it must be widely tested and demonstrated to provide a notable improvement on the current methodology. Opinions will naturally differ on what criteria are necessary for such a new method to be adopted. In this paper, a simulation method has been used to test the filter. Simulation, which is becoming an increasingly utilised tool in visual field problems (Gardiner & Crabb, 2002a, 2002b; Glass, Schaumberger, & Lachenmayr, 1995; Johnson, Chauhan, & Shapiro, 1992; Spenceley & Henson, 1996; Spry, Bates, Johnson, & Chauhan, 2000; Spry et al., 2002; Vesti, Spry, Chauhan, & Johnson, 2002), was used to test not only whether the noise was being reduced by filtering, but also whether true localised defects were being blurred out. The details of the simulation used, and its benefits and disadvantages, are discussed elsewhere (Gardiner & Crabb, 2002a, 2002b). With any simulation, the results should be viewed qualitatively rather than quantitatively, due to uncertainty over such issues as the amount of noise used etc. Nevertheless, these experiments have clearly shown that defects are emphasised by the new filter, whilst there are less false positives from stable points and non-glaucomatous defects.

The use of a map of the ONH for generating localised defects (Garway-Heath et al., 2000) has the notable advantage that it is not wholly based on perimetric results, thus making it suitable for independently selecting progressive defects. The rare instances of progressive defects where the new filter did not emphasise progression may tell us more about that lack of relationship between functional and structural change; or equally, may be a result of correlations forged as part of the testing strategy of the perimeter itself in establishing thresholds. Some correlation between points, whether contiguous or not, may well be the result of the testing strategy and not of the physiological connections. It was interesting to note that the perimetric data alone indicated correlation between the two nasal points either side of the horizontal meridian which is often noted clinically (Sakai et al., 1987; Vrabec, 1966). The effect of this is that the range of defects generated may not be entirely realistic or comprehensive. Further, it means that the detection rate being reduced by filtering in the cases of two of the tested defects (defects D and E in Fig. 7) need not be considered a problem with the filter; more, it may point to a difference between the structural and functional maps of the eye at those points. It is perfectly believable, and in fact to be expected, that there will be a few points in the eye where the results from this testing technique would not be perfect even if the filter were. This is a limitation of using this method for choosing realistic localised defects; yet this is preferable to choosing defects in a completely subjective manner based on common beliefs about visual fields, which would be biased towards the expected shape.

Naturally, any further testing is to be welcomed. It would be desirable to also test whether or not localised defects are blurred out by filtering by looking at the real patient data. This is hampered by the lack of a definition in the literature of what constitutes an expected defect; progressing or otherwise (hence why our simulation uses shapes of the localised defects generated by looking at a physiological map of the ONH). One possibility which is currently being considered is to look at the predictive power of trend analysis methods, in the hope that this power would increase once the data had been filtered; however to do this would require a clean, independent database, and so it has not been carried out to date. Despite these inherent problems, and although the results of simulation should not be underestimated, they still need to be confirmed by clinical observations.

It should be emphasised that the filter described here is designed solely for suspected glaucoma patients. It is based on data from a glaucoma clinic, and so resembles the shapes of glaucomatous defects. This means that it is unsuitable for use when other conditions are suspected; neurological defects, for example, will be blurred out by the use of this filter, as demonstrated by the effect of the filter on the defect shown in Fig. 3. Of course, when the issue is whether the patient has glaucoma or not, or whether their glaucoma is progressing, it is entirely desirable that non-glaucomatous defects be blurred out.

Even without further testing, there is sufficient evidence presented here to say that the new filter shows clinical potential, especially since it requires no changing of visual field testing or extra patient test time. It is to be hoped that the filter technique described here could in the future become a widely accepted tool in glaucoma clinics.

Acknowledgements

The authors would like to thank David Garway-Heath and Ananth Viswanathan of the Glaucoma Unit, Moorfields Eye Hospital, London for advice during the preparation of this manuscript.

References

Anderson, D. R., & Patella, V. M. (1999). *Automated static perimetry* (2nd ed.). St. Louis, MO: Mosby, pp. 147–159.

Artes, P. H., Iwase, A., Ohno, Y., Kitazawa, Y., & Chauhan, B. C. (2002). Properties of perimetric threshold estimates from Full Threshold, SITA Standard, and SITA Fast strategies. *Investigative Ophthalmology and Visual Science., 43*, 2654–2659.

Bengtsson, B., Olsson, J., Heijl, A., & Rootzén, H. (1997). A new generation of algorithms for computerized threshold perimetry SITA. *Acta Ophthalmologica Scandinavica, 75*, 368–375.

Budenz, D. L. (1997). *Atlas of visual fields*. Philadelphia, PA: Lippincott-Raven, pp. 146–155.

Chauhan, B. C., & House, P. H. (1991). Intratest variability in conventional and high-pass resolution perimetry. *Ophthalmology, 98*, 79–83.

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

Crabb, D. P., Edgar, D. F., Fitzke, F. W., McNaught, A. I., & Wynn, H. P. (1995). New approach to estimating the variability in visual field data using an image processing technique. *British Journal of Ophthalmology, 79*, 213–217.

Crabb, D. P., Fitzke, F. W., McNaught, A. I., Edgar, D. F., & Hitchings, R. A. (1997). Improving the prediction of visual field progression in glaucoma using spatial processing. *Ophthalmology, 104*, 517–524.

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

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

Flammer, J., Drance, S. M., & Zulauf, M. (1984). Differential light threshold: short- and long-term fluctuation in patients with glaucoma, normal controls, and patients with suspected glaucoma. *Archives of Ophthalmology, 102*, 704–706.

Gardiner, S. K., & Crabb, D. P. (2002a). Frequency of testing for detecting visual field progression. *British Journal of Ophthalmology, 86*, 560–564.

Gardiner, S. K., & Crabb, D. P. (2002b). Examination of different pointwise linear regression methods for determining visual field progression. *Investigative Ophthalmology and Visual Science, 43*, 1400–1407.

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

Glass, E., Schaumberger, M., & Lachenmayr, B. J. (1995). Simulations for FASTPAC and the standard 4-2 dB full threshold strategy of the Humphrey field analyser. *Investigative Ophthalmology and Visual Science, 36*, 1847–1854.

Heijl, A., Lindgren, G., Lindgren, A., Olsson, J., Asman, P., Myers, S., & Patella, M. (1991). Extended empirical statistical package for evaluation of single and multiple fields in glaucoma: Statpac 2. In R. P. Mills, & A. Heijl (Eds.), *Perimetry update 1990/1991* (pp. 303–315). Amsterdam: Kugler & Ghedini.

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

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

Johnson, C. A., Chauhan, B. C., & Shapiro, L. R. (1992). Properties of staircase procedures for estimating thresholds in automated perimetry. *Investigative Ophthalmology and Visual Science, 33*, 2966–2974.

Katz, J., Gilbert, D., Quigley, H. A., & Sommer, A. (1997). Estimating progression of visual field loss in glaucoma. *Ophthalmology, 104*, 1017–1025.

Lachenmayr, B. J., & Vivell, P. M. O. (1993). *Perimetry and its clinical correlations*. Stuttgart: Georg Thieme Verlag.

Membrey, W. L., Bunce, C., Poinoosawmy, D. P., Fitzke, F. W., & Hitchings, R. A. (2001). Glaucoma surgery with or without adjunctive antiproliferatives in normal tension glaucoma: 2. Visual field progression. *British Journal of Ophthalmology, 85*, 696–701.

Nouri-Mahdavi, K., Brigatti, L., Weitzman, M., & Caprioli, J. (1997). Comparison of methods to detect visual field progression in glaucoma. *Ophthalmology, 104*, 1228–1236.

Sakai, T., Kuniyoshi, S., Tsuzuki, K., Makoto, U., & Kawamura, Y. (1987). Temporal raphe of the retinal nerve fiber layer revealed by medullated fibers. *Japanese Journal of Ophthalmology, 31*, 655–658.

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

Spenceley, S. E., & Henson, D. B. (1996). Visual field test simulation and error in threshold estimation. *British Journal of Ophthalmology, 80*, 304–308.

Spry, P. G. D., Bates, A. B., Johnson, C. A., & Chauhan, B. C. (2000). Simulation of longitudinal threshold visual field data. *Investigative Ophthalmology and Visual Science, 41*, 2192–2200.

Spry, P. G. D., & Johnson, C. A. (2002). Identification of progressive glaucomatous visual field loss. *Survey of Ophthalmology, 47*, 158–173.

Spry, P. G. D., Johnson, C. A., Bates, A. B., Turpin, A., & Chauhan, B. C. (2002). Spatial and temporal processing of threshold data for

detection of progressive glaucomatous visual field loss. *Archives of Ophthalmology, 120*, 173–180.

Vesti, E., Spry, P. G. D., Chauhan, B. C., & Johnson, C. A. (2002). Sensitivity differences between real-patient and computer-simulated visual fields. *Journal of Glaucoma, 11*, 35–45.

Viswanathan, A. C., McNaught, A. I., Poinoosawmy, D., Fontana, L., Crabb, D., Fitzke, F., & Hitchings, R. (1999). Severity and stability of glaucoma: patient perception compared with objective measurement. *Archives of Ophthalmology, 117*, 450–454.

Vrabec, F. (1966). The temporal raphe of the human retina. *American Journal of Ophthalmology, 62*, 926–938.

Wild, J. M., Hutchings, N., Hussey, M. K., Flanagan, J. G., & Trope G. E. (1997). Pointwise univariate linear regression of perimetric sensitivity against follow-up time in glaucoma. *Ophthalmology 104*, 808–815.

Wild, J. M., Pacey, I. E., O'Neill, E. C., & Cunliffe, I. A. (1999). The SITA perimetric threshold algorithms in glaucoma. *Investigative Ophthalmology and Visual Science, 40*, 1998–2009.

Wild, J. M., Searle, A. E. T., Dengler-Harles, M., & O'Neill, E C (1991). Long-term follow-up of baseline learning and fatigue effects in the automated perimetry of glaucoma and ocular hypertensive patients. *Acta Ophthalmologica Scandinavica, 69*, 210–216.

UNCORRECTED PROOF