

Spatial Modeling of Visual Field Data for Assessing Glaucoma Progression

Brigid D. Betz-Stablein,¹ William H. Morgan,² Philip H. House,² and Martin L. Hazelton¹

PURPOSE. In order to reduce noise and account for spatial correlation, we applied disease mapping techniques to visual field (VF) data. We compared our calculated rates of progression to other established techniques.

METHODS. Conditional autoregressive (CAR) priors, weighted to account for physiologic correlations, were employed to describe spatial and spatiotemporal correlation over the VF. Our model is extended to account for several physiologic features, such as the nerve fibers serving adjacent loci on the VF not mapping to the adjacent optic disc regions, the presence of the blind spot, and large measurement fluctuation. The models were applied to VFs from 194 eyes and fitted within a Bayesian framework using Metropolis-Hastings algorithms.

RESULTS. Our method (SPROG for Spatial PROGression) showed progression in 42% of eyes. Using a clinical reference, our method had the best receiver operating characteristics compared with the point-wise linear regression methods. Because our model intrinsically accounts for the large variation of VF data, by adjusting for spatial correlation, the effects of outliers are minimized, and spurious trends are avoided.

CONCLUSIONS. By using CAR priors, we have modeled the spatial correlation in the eye. Combining this with physiologic information, we are able to provide a novel method for VF analysis. Model diagnostics, sensitivity, and specificity show our model to be apparently superior to current point-wise linear regression methods. (<http://www.anzctr.org.au> number, ACTRN12608000274370.) (*Invest Ophthalmol Vis Sci.* 2013; 54:1544-1553) DOI:10.1167/iovs.12-11226

One of the most important challenges in managing glaucoma is detecting whether the visual field (VF) is progressing. Once a glaucoma patient loses sight the damage is irreversible; however, early detection of progression can lead to treatment that can reduce vision loss.¹ The most common method for monitoring changes in the VF is through Standard Automated Perimetry (SAP).^{2,3} Progression is judged by looking at a series of SAP VF outputs over time. Due to the large amounts of variation in the VF it is difficult to detect true

change from change due to measurement error and random variation.⁴⁻⁶ Many analytical methods exist to help determine progression. Currently, however, there is no consensus on the gold standard for analyzing VF data. This often leads to investigators developing their own methods to suit their particular study.⁷

Ernest et al.⁸ reviewed literature prior to April 2009 and found that 47% of studies using quantitative methods calculated rates of progression, while the remainder dichotomized the outcome. The main methods can be divided into three categories: scoring methods, glaucoma change probability analysis, and point-wise linear regression (PLR) analysis. Other methods investigated include the VF index, neural networks, and spatial filters. A comparison of progression detection methods using a meta-analysis⁷ indicates that methods predicting the highest rates of progression included PROGRESSOR (Medisoft Ltd., Leeds, UK) (a form of PLR), the Threshold Noiseless Trend program, and a qualitative clinical method. The majority of the PLR methods using indices (e.g., mean deviation and VF index methods), and the PLR on threshold values with tighter significance criteria gave lower estimates of progression. However, due to the lack of a gold standard it is also hard to compare methods to determine whether one is better than another. Vesti et al.^{9(p3878)} states that "The ideal method for analyzing VF change should be sensitive, detect progression with few examinations, maintain high specificity, and be resistant to fluctuation," and it has been noted that a method which identifies the rate of progression is of more benefit in determining the future of vision loss, treatment, and thus quality of life in glaucoma patients.¹⁰

We suggest that statistical techniques be employed to model the spatial correlation of the VF data, which in turn will help minimize variation in the VF data. Previously, a spatial filter was introduced to take a weighted average of the surrounding points.¹¹ While this method succeeded in reducing variation, it was also less sensitive,¹² and did not take into account the physiology of the eye, in that the VF locations along the temporal horizontal midline do not map in adjoining locations on the optic disc. Another spatial filter was created to explore spatial correlations in VFs in order to reduce noise. This was based upon the covariances between different points in the eye.¹³ While this method seemed to increase sensitivity and specificity in simulated data,¹³ it performed similarly to PLR when tested on longitudinal data.¹⁴

While these methods attempt to account for the spatial nature of the data, the spatial correlation structure of that data can be modeled more flexibly by applying disease mapping techniques to the VF dataset. These methods cannot be applied blindly to our data since the VF has spatial properties that are not encountered in geographical disease mapping. In particular, the spatial relationships on the VF arise from a somewhat complicated projection from the retina. As mentioned above, points may appear beside each other on the VF although they do not necessarily project along an adjacent pathway to the

From the ¹Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand; and the ²Lions Eye Institute, University of Western Australia, Perth, Australia.

Supported by grants from the National Health and Medical Research Council (1020367, WHM) and a Massey University Doctoral Scholarship (BDB-S).

Submitted for publication October 29, 2012; revised January 8, 2013; accepted January 14, 2013.

Disclosure: **B.D. Betz-Stablein**, None; **W.H. Morgan**, None; **P.H. House**, None; **M.L. Hazelton**, None

Corresponding author: Brigid D. Betz-Stablein, Massey University PN322, Private Bag 11222, Palmerston North, New Zealand 4442; b.d.betz-stablein@massey.ac.nz.

optic disc. Also, the VF includes a blind spot corresponding to the point where the optic nerve enters the eye.¹⁵ In the following sections, we present our methodology, results from fitting our model to data from the Lions Eye Institute trial registry and Vein Pulsation Study in Glaucoma (Lions Eye Institute, University of Western Australia), our discussion, and conclusions.

METHODS

Participants

One hundred and ninety-four series of VFs (1448 fields in total) from 98 subjects were obtained from participants enrolled in the Vein Pulsation Study Trial in Glaucoma (VPSG) and on the Lions Eye Institute trial registry, in Perth, Western Australia. All 98 subjects had a form of open angle glaucoma (OAG). One subject had OAG following successful laser iridotomy for angle closure in one eye, the other eye being blind. Of the remaining 97 subjects, eight eyes had pseudo-exfoliation syndrome and five had pigment dispersion syndrome, with the rest (180 eyes) having POAG. The average age was 66.9 years (SD 9.7), with 34% males and 66% females. Informed consent was obtained for all participants. The trial adhered to the tenets of The Declaration of Helsinki and was approved by the University of Western Australia's ethics committee. The mean follow up time for participants was 2.5 years (range, 0.2–9.4 years), with an average of 7.5 VFs (range, 2–21) taken during this period.

Standard Automated Perimetry

The Humphrey Field Analyzer (Carl Zeiss Meditec Inc., Dublin, CA) measures the differential light sensitivity across the VF. Figure 1 depicts a grid, which contains 54 test locations on the VF. At each cell the subject receives a score ranging between zero to approximately 30 dB on average, where a brighter light correlates with a lower decibel reading. Decibels are measured on the log scale, where 1 log unit is equivalent to 10 dB, and a change in 1 log unit is equivalent to a 10-fold change in light sensitivity (apostilbs). Because 0 dB represents the maximum light (10,000 apostilbs) tested on the Humphrey Field Analyzer (Carl Zeiss Meditec Inc.), the subject's score could theoretically go below zero. We have taken this into account in our statistical methods described later in this section.

The Humphrey Field Analyzer 24-2 program was used with white on white stimulus size III. Fifteen participants from the registry used full threshold SAP, while the 81 participants from the VPSG trial used the Swedish Interactive Threshold Algorithm (SITA). Two participants had a combination of full threshold and SITA exams. SITA is a faster technique than the full threshold program, which maintains similar variability properties; however, threshold estimates are approximately 1-dB higher than the with full threshold technique.¹⁷ Appropriate adjustments were, therefore, made to the observed field thresholds of participants using both techniques. All subjects had undergone at least two VF tests before participating in the trial, thus, minimizing the learning effect. VF output from left eyes were inverted allowing the same modeling as right eyes.

Correlation Structure of Data

Following disease mapping ideology, we assume that points that neighbor each other on the eye will be more similar than points that lie farther apart. Therefore, we expect to see correlation in the spatial pattern of vision loss, reflecting the underlying damage to the optic disc, particularly within the sectors of each VF. In spatiotemporal disease mapping, it is common to use conditional autoregressive (CAR) models^{18,19} for the underlying mean process to describe the dependence between regions sharing a common boundary. For the most part a similar principle applies to the VF measurements, since adjacent grid cells in the VF usually correspond to proximate areas on

the optic disc. However, there is more complexity across the (horizontal) midline. Specifically, at the temporal side of the eye, pairs of adjacent cells with one above the midline and one below that correspond to highly separated pathways across the retina and into the optic disc, such that they can be treated as essentially independent¹⁶ (see dotted line in Fig. 1 separating the superior temporal and inferior temporal sections on the VF). We did consider splitting the temporal sector (Fig. 1) into superior and inferior halves but the degree of separation is not so clear compared with VF nasal fixation.^{20,21} Accordingly, we did not introduce any distinct bias in to the temporal sector as defined by Garway-Heath.¹⁶ Another complication in the retina is the blind spot, which if measured correctly will record values of 0 dB, and, therefore, does not undergo vision loss in the same manner as the rest of the eye.

An adjacency matrix is required to define which points in the VF are neighboring. The denoted adjacencies allow for a weighted average to be generated from the surrounding points. This information is then used by the CAR prior detailed below in the Statistical Modeling subsection (Equations 5, 6). We develop our adjacency matrix for the VF based on physical adjacency in Figure 1 (left), adjusted to represent the physiologic structure of the eye (Fig. 1, right). Because the sectors at the midline on the temporal side are not adjacent on the optic disc, the correlations do not cross the midline on the temporal side (i.e., grid cells above and below are assumed independent). However, it is also known that adjacent points within a sector are more likely to be similar than adjacent points of different sectors. This is because the nerve fibers from the same sector group together on the optic disc. Therefore, we have chosen to weight points that are adjacent and intrasector as 1, points that are adjacent and intersector as 0.3, and points that are nonadjacent as 0. We also remove the two points closest to the blind spot in constructing this adjacency matrix, as the behavior of these points is independent from the other points in the temporal sector, leaving 52 points for analysis.

Statistical Modeling

Having constructed the appropriate adjacency matrix, we are now in a position to develop a model for the vision measurements. We let y_{ik} denote the measurement taken at VF grid location i at time t_k , where the grid cells are numbered by row from top left to bottom right. These are assumed to follow a normal distribution, but to be censored to the left at zero, to account for the idea that a threshold of below zero could be obtained if a higher light intensity, thus, a lower decibel was tested on the Humphrey VF. That is, $y_{ik} = \max(y_{ik}^*, 0)$ where

$$y_{ik}^* | \mu_{ik} \sim \text{Normal}(\mu_{ik}, \sigma_c^2). \quad (1)$$

We implement our model within a Bayesian statistical framework. This requires that model parameters are assigned prior distributions that summarize any available prior information. The variance parameter σ_c^2 describes the magnitude of variation due to measurement error. It is assigned an inverse gamma as follows:

$$\sigma_c^2 \sim \text{InverseGamma}(5.6, 43.2) \quad (2)$$

An informative prior was allocated to σ_c^2 , as we had preexisting information about the measurement error within the VF. A representative subset of 411 full threshold VFs from 34 eyes were examined. The difference in retested loci within each field was used to estimate variation. A mean variance of 9.4 dB (variance 36 dB) was calculated for VF readings at individual loci (y_{ik}). Values for the prior were converted to create an informative inverse gamma prior centered around 9.4 dB with a scale of 36 dB.

The y_{ik}^* are conditionally independent given μ_{ik} . The spatial and temporal dependence in the data are generated by the manner in which we model μ_{ik} . Specifically, we model the spatiotemporal variation in the VF by

$$\mu_{ik} = \alpha + \delta_i + \beta t_k + \eta_i t_k \quad (3)$$

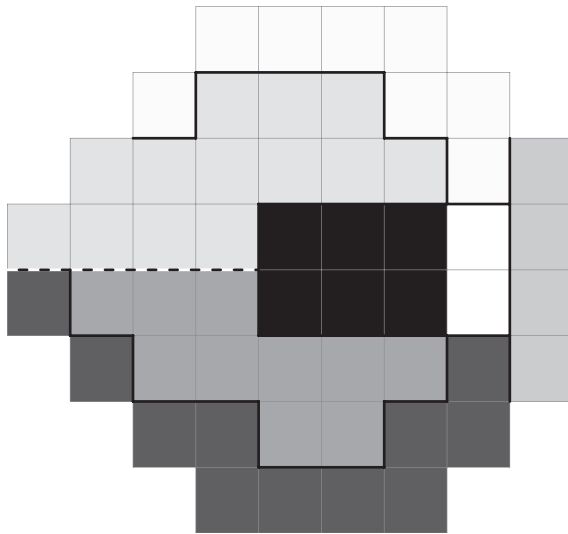


FIGURE 1. Visual field regions mapped to the optic disc.¹⁶

The overall eye mean is represented by α , and the overall temporal trend described by β . CAR prior distributions are employed to account for spatial correlated heterogeneity,²² as described below in Equations 4 to 6. The parameter δ_i adjusts the mean for spatially correlated heterogeneity at each individual locus, while η_i adjusts the trend.

For the δ_i parameter the prior is given by

$$[\delta_i | \delta_j, i \neq j, \sigma_{\delta_i}^2] \sim \text{Normal}(\bar{\delta}_i, \sigma_{\delta_i}^2) \tag{4}$$

where

$$\bar{\delta}_i = \left(\frac{1}{\sum_{j \neq i} w_{ij}} \right) \sum_{j \neq i} w_{ij} \delta_j \tag{5}$$

$$\sigma_{\delta_i}^2 = \left(\frac{\kappa_{\delta}}{\sum_{j \neq i} w_{ij}} \right). \tag{6}$$

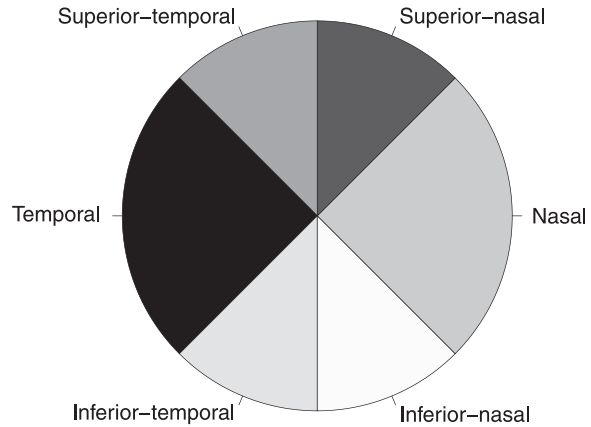
According to our adjacency matrix described earlier, $w_{ij} = 1$ if i, j are adjacent and of the same sector, $w_{ij} = 0.3$ if i, j are adjacent and not of the same sector, and $w_{ij} = 0$ if they are nonadjacent. We employ a vague prior for κ_{δ} ($\kappa_{\delta} \sim \text{InverseGamma}(1/2, e/2)$) where e is a small constant, in our case 0.01.¹⁹ Hence, the conditional posterior for κ_{δ} is given by

$$\kappa_{\delta} \propto \kappa_{\delta}^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\kappa_{\delta}} \left[e + \sum_{i \neq j} w_{ij} (\delta_i - \delta_j)^2 \right] \right\}. \tag{7}$$

A similar type of CAR prior is applied to the space-time effects, δ_i , where η_i and σ_{η}^2 are defined as for δ_i and σ_{δ}^2 . That is

$$[\eta_i | \eta_j, i \neq j, \sigma_{\eta_i}^2] \sim \text{Normal}(\bar{\eta}_i, \sigma_{\eta_i}^2). \tag{8}$$

A normal prior was employed for α ($\alpha \sim \text{Normal}(24, 18)$) where 24 is the mean threshold score and 18 is two times the SD across all loci of all eyes at time point zero. The doubling of the SD was included to reduce the influence of prior information. Time (t_k from Equation 3) is centered at zero with a SD of 0.5. We then assign a Cauchy prior to our trend parameter (or global slope), β . When little is known of the trend an uninformative prior Cauchy (0, 2.5) on a centered dataset with a SD of 0.5 can be used.²³ However, as we know more about the scale we assign $\beta \sim \text{Cauchy}(0, 5)$ instead of a scale 22.5 (2.5 times the SD of y_{ik}).



Model Implementation

Models were fitted in R²⁴ using Markov Chain Monte Carlo (MCMC) methods. MCMC works by drawing simulations of model parameters from a Markov chain whose stationary distribution matches the required posterior distribution.²⁵ The Metropolis-Hastings (MH) algorithm is used to sample values from the Markov chain. A candidate value (θ) is generated from the proposal distribution. The acceptance probability of θ is calculated, and θ is either accepted or rejected. If θ is accepted it replaces the existing value and, thus, the chain moves, if rejected the existing value remains.²⁵ Gibbs sampling is also used and is a special case of the MH algorithm where all candidate values are accepted. Random walk MH, where candidate values are sampled dependent on the current value of the chain, was used for sampling α and β .

We implemented the following component-wise transition MCMC algorithm:

1. Set iteration counter $z = 0$ and initialize $\delta^{(0)}, \alpha^{(0)}, \eta^{(0)}, \beta^{(0)}, \sigma_{\delta}^{(0)}, \sigma_{\eta}^{(0)}, \sigma_{\epsilon}^{(0)}$ as specified in the following steps;
2. Set $z = z + 1$;
3. For i in 1-52, sample $\delta_i^{(z)}$, conditional on $\delta_{1:i-1}^{(z-1)}, \delta_{i+1:52}^{(z-1)}, \alpha^{(z-1)}, \eta^{(z-1)}, \beta^{(z-1)}, \sigma_{\delta}^{(z-1)}, \sigma_{\eta}^{(z-1)}, \sigma_{\epsilon}^{(z-1)}$ using Equation 4;
4. Sample $\alpha^{(z)}$ conditional on $\delta^{(z-1)}, \eta^{(z-1)}, \beta^{(z-1)}, \sigma_{\delta}^{(z-1)}, \sigma_{\eta}^{(z-1)}, \sigma_{\epsilon}^{(z-1)}$ using random walk MH;
5. For i in 1-52, sample $\eta_i^{(z)}$ conditional on $\eta_{1:i-1}^{(z-1)}, \eta_{i+1:52}^{(z-1)}, \delta^{(z-1)}, \alpha^{(z-1)}, \beta^{(z-1)}, \sigma_{\delta}^{(z-1)}, \sigma_{\eta}^{(z-1)}, \sigma_{\epsilon}^{(z-1)}$ using Equation 8;
6. Sample $\beta^{(z)}$ conditional on $\delta^{(z-1)}, \alpha^{(z-1)}, \eta^{(z-1)}, \sigma_{\delta}^{(z-1)}, \sigma_{\eta}^{(z-1)}, \sigma_{\epsilon}^{(z-1)}$ using random walk MH;
7. Sample $\kappa_{\delta}^{(z)}$ using Gibbs sampling (Equation 7) conditional on $\delta^{(z-1)}, \alpha^{(z-1)}, \eta^{(z-1)}, \beta^{(z-1)}, \sigma_{\eta}^{(z-1)}, \sigma_{\epsilon}^{(z-1)}$ and, hence, calculate $\sigma_{\delta}^{(z)}$;
8. Sample $\kappa_{\eta}^{(z)}$ using Gibbs sampling (Equation 7) conditional on $\delta^{(z-1)}, \alpha^{(z-1)}, \eta^{(z-1)}, \beta^{(z-1)}, \sigma_{\delta}^{(z-1)}, \sigma_{\epsilon}^{(z-1)}$ and, hence, calculate $\sigma_{\eta}^{(z)}$;
9. Sample $\sigma_{\epsilon}^{(z)}$ conditional on $\delta^{(z-1)}, \alpha^{(z-1)}, \eta^{(z-1)}, \beta^{(z-1)}, \sigma_{\delta}^{(z-1)}, \sigma_{\eta}^{(z-1)}$ using random walk MH; and
10. Repeat steps 2 through 9 until required number of iterations is completed.

Lack of formal identifiability of α and δ_i , and β and η_i , is handled by replacing α with $\alpha + \sum \delta_i/n$ and recentering so that $\sum \delta_i = 0$ (i.e., $\delta_i \leftarrow \delta_i - \sum \delta_i/n$).²⁶ Parameters $\sum \eta_i$ and β are handled in the same way. Models were run for 100,000 iterations, with burn in period of 12,000, and a thinning factor of 40. Parameter α was initialized at 23, the mean sensitivity of all eyes over all time points. Trend β was initialized at zero. SDs σ_{δ} and σ_{η} were initialized at 5 and 1, respectively, based on

TABLE 1. Summary of PLR Methods Progression Criteria

Method	Progression Criteria
P1	$P \leq 0.001$ for ≥ 1 point ³¹
P2	$P \leq 0.001$ for ≥ 2 points ³¹
P3	$P \leq 0.001$ for ≥ 3 points ³¹
P4	$P \leq 0.001$ for ≥ 4 points ³¹
V1	$P \leq 0.01$ and slope ≤ -1 dB per year for ≥ 2 points over 3 consecutive tests ⁹
V2	$P \leq 0.01$ and slope ≤ -1 dB per year for ≥ 3 points over 3 consecutive tests ⁹
F1	$P \leq 0.1$ and slope ≤ -1 dB for inner points, ≤ -2 dB for points beyond 15° for ≥ 2 points ²⁸
F2	$P \leq 0.1$ and slope ≤ -1 dB for inner points, ≤ -2 dB for points beyond 15° for ≥ 3 points ²⁸
VS	$P \leq 0.05$ and slope ≤ -1 dB per year for ≥ 1 nonedge* point, ≤ -2 dB for edge points, for ≥ 1 point ²⁹
B1	$P \leq 0.05$ and slope ≤ -1 dB per year for ≥ 2 nonedge* points ³⁰

P is the *P* value associated with the slope and points refers to each tested point (*i*) on the VF.

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

our experience with the variability of the VF measurements. Parameter η is initialized at 0, δ_i at y_i at time point 0. Convergence was assessed using the Geweke diagnostic.²⁷ The deviance information criteria (DIC) was calculated and used to test the goodness of fit of our models.²⁵ Like other information criteria it penalizes models, which provide a poor fit and/or are overly complex with superfluous parameters. For each eye, significant progression was defined as a one sided Bayesian *P* value of less than or equal to 0.05 for the overall eye trend, β . We term our method SPROG, for Spatial PROGression, and, henceforth, refer to our model by this name.

PLR Methods

The linear modeling (lm) function in R²⁴ was used to run point-wise linear regression for each of the 52 nonblind points on the 24-2 field. No outliers were removed. Progression was evaluated at the final time point based on the values of the slope and their significance (*P* values). Progression was judged using several criteria as defined as in Table 1. Our results may differ slightly from Fitzke et al., Viswanathan et al., and Birch et al.²⁸⁻³⁰ as PLR was used instead of the PROGRESSOR software.

PLR was then carried out within a Bayesian framework in order to calculate the DIC to test model fit and provide a direct comparison to our new method.

Clinical Judgment of Progression

In order to compare our model with a clinical reference, VFs were evaluated by two independent clinicians using only the VF information. A third clinician evaluated any discrepancies. The expert clinicians were given access to the overview print format of the Humphrey 24-2 fields and Glaucoma Progression Analysis printout and asked to make a judgment as to whether the field was definitely progressing or not.⁵² No specific guidelines or criteria were given to the clinician.

RESULTS

Of the 194 models run, 192 converged within the 100,000 iterations. Acceptance rates for all parameters sampled through MH random walk were close to 50%. The mean DIC for our spatial models was 1719 (range, 310-5940). This is lower than that of the PLR models (2047; range, 583-6421), indicating our model better represents the observed data. Observing the DIC results for models for individual eyes, our method SPROG was preferred to PLR in 97% of cases.

Our model showed 42% of the eyes progressing, with statistically significant global slopes (β), at an average rate of -0.78 (range, -1.76 to -0.27) dB per year. Figure 2 shows the fit of our model by sector, for an individual eye compared with the observed thresholds. The local slopes ($\beta + \eta_i$) for an individual can be observed here through the predicted line. We

can see that the model reduced the effect of overall variation in the eyes by allowing the predicted points to be influenced by their neighboring points. This also means the result is not overly affected by the occurrence of outliers. However, the model is still flexible enough to allow a certain point within a sector to differ if the mean of a certain point in the eye is consistently lower (Fig. 2, temporal sector). The spatial nature of the VF can be seen in Figure 3, where the local slopes ($\beta + \eta_i$) are presented by sector. This eye is progressing at an average rate of -1.48 dB per year. As expected, neighboring points are more similar than points farther apart.

Table 1 summarizes a selection of PLR criteria used to evaluate VF progression. Figure 4 shows the ranking of the PLR methods summarized in Table 1 and our method, SPROG. The methods ranged from predicting 0% to 75% progression. The mean progression for the methods tested was 30%. Our method sits above the mean and ranks eighth out of 12 in terms of the number of eyes it predicts as progressing. Our SPROG method was one of the two closest matches to the clinical reference in terms of overall progression rates, and was by far the closest amongst the methods that did not underdiagnose in comparison to clinical reference. Our model outperforms the other models in that it gives reasonable predictions of both specificity and sensitivity (Table 2). We also investigated using a *P* value of 0.1, and including a slope criteria for our model. We chose the value of -0.5 dB per year as generally any loss less than this is considered clinically

TABLE 2. Sensitivity and Specificity of Our Method Compared with PLR Methods against the Clinical Reference

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

Methods are as described in Table 1 and the Clinical Judgment of Progression subsection in the Methods section. For the SPROG model, we have also looked at the effect of changing the *P* value, as well as introducing a limit on the slope.

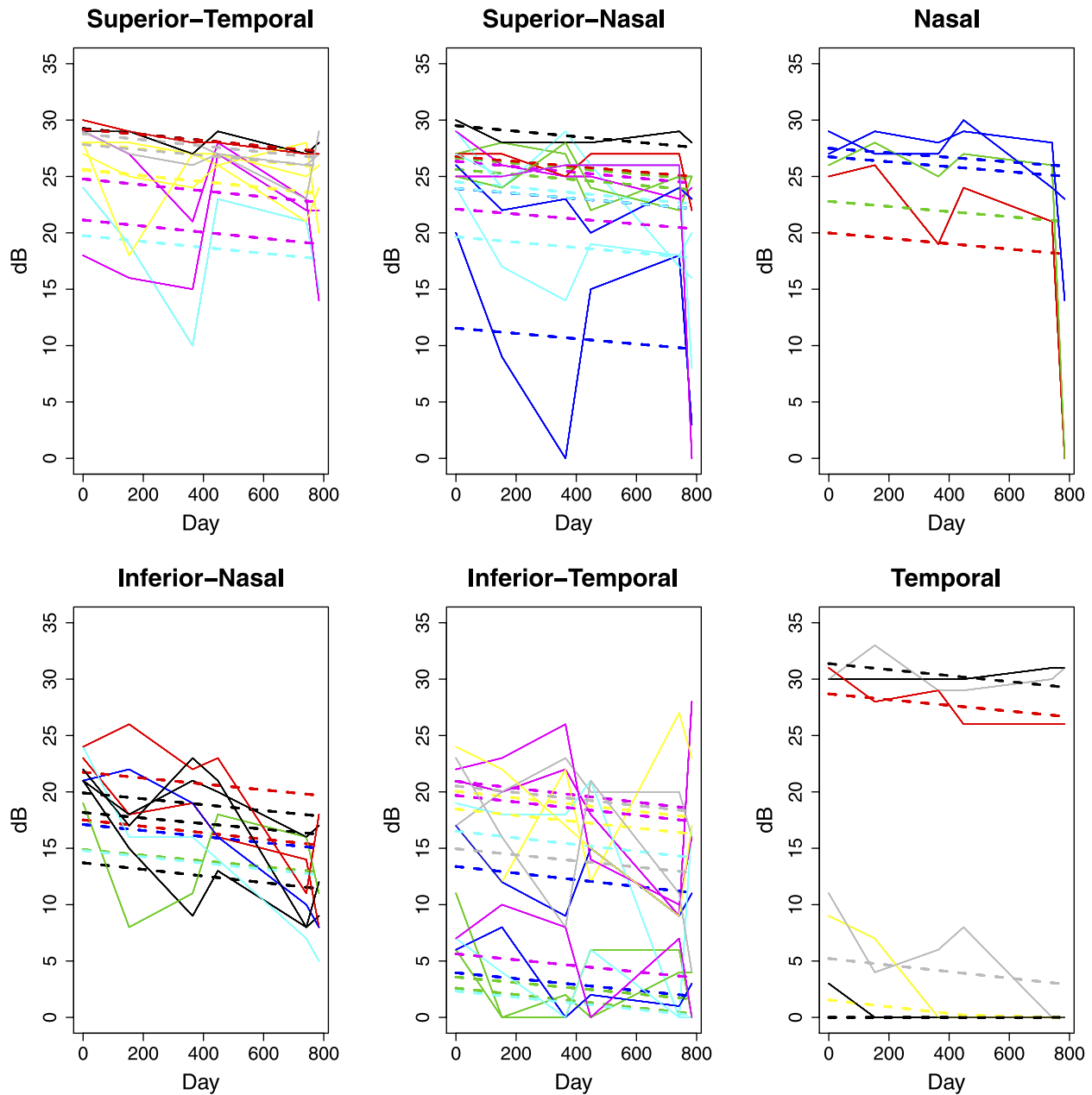


FIGURE 2. Observed versus predicted sensitivities of the 52 analyzed points by sector of the visual field. Observed and predicted lines are color matched for each locus within each sector. *Solid lines* depict observed data and *dashed lines* represent fitted values from the SPROG model.

significant progression. With the changed criteria SPROG still performed best.

Receiver operating characteristic (ROC) curves are shown for SPROG and P1 to P4 (Fig. 5). These are obtained by varying the *P* value thresholds from the standard 0.05 or 0.001. We have not included ROC curves for methods with requirements on the numerical value of the slopes because these are not directly comparable. In particular, when the specificity is zero we will obtain a sensitivity of less than one because the slope conditions will not always be attained. The area under the curve (AUC) values, for the same methods, are presented in the caption of Figure 5. SPROG shows the largest AUC, showing that this method has the best overall ROC properties.

DISCUSSION

Our model provides a statistical method to model VF progression, which takes account of spatial correlation within VFs, while respecting the relationship between the spatial arrangement of the field with the anatomy of the eye. It also provides a method, which is robust to outliers, due to the pooling of strength within and between proximate sectors, and can overcome the large amounts of variation in a VF dataset. Figure 6 illustrates this principle, by presenting the local slopes across the VF map. In (a) the spatial correlation smooths the VF, showing progression in all sectors of the VF. In (b) the PLR method shows erratic estimates of the slopes, which do not support the fact that if one point is progressing, points nearby are more likely to be progressing.³³ These maps can be

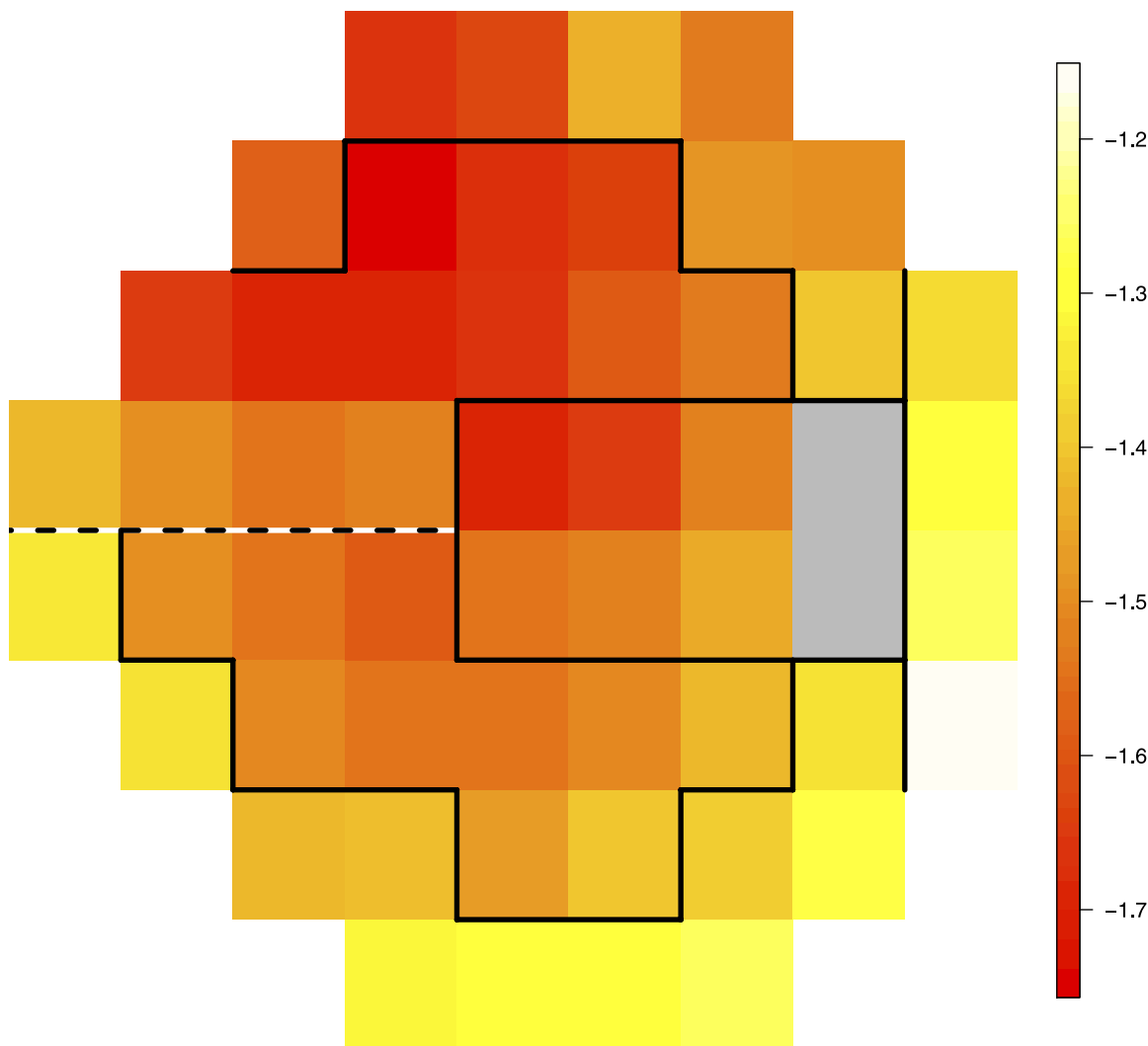


FIGURE 3. Heat map of local slope parameters ($\beta + \eta_i$) by sector for a single progressing eye. Slopes show rates of sensitivity change in decibels per year. Black lines separate the sectors and loci nearest the blind spot are shaded gray.

compared with the observed data seen in Figure 2. SPROG shows a global progression rate (β) of -1.48 dB per year with local slopes ranging from -1.15 to -1.75 . PLR produces a reasonably comparable estimate of the average progression rate (taken as the median of 52 local slopes, so as to control for extreme values) at -2.13 , but gives much greater local variation (range, -7.87 to 4.77). One consequence of the erratic behavior of the PLR method is that it indicates very significant vision improvement in some regions of the eye (where the slopes are positive), an implausible finding that is not seen when using the SPROG methodology.

This model has the potential to provide more reliable results with higher specificity and sensitivity than existing methods. By borrowing information from surrounding regions, our model is able to smooth the occurrence of spikes in the data, while PLR methods struggle with the occurrence of spikes in a dataset. In contrast the PLR methods are highly influenced by outliers and identified eyes as improving that SPROG showed to be stable. Methods P1 to P4 and F1 to F2 even identified an eye as improving, which our method classified these eyes as progressing. Criteria that require three consecutive fields to be significant before classifying progression underestimate true progression. This was noted in our dataset for eyes, which were classified as progressing for two

consecutive points, stable at the next time point and then progressing for the following two time points. Because significant progression was not seen at the middle time point, most likely due to an outlier, this eye was not classified as progressing. For the eye seen in the figures, both the clinician and SPROG identified this eye as progressing, while only 50% of the PLR criteria classified this eye as progressing.

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

Previous spatial modeling has focused on preprocessing the data by applying a spatial filter, where a weighted average is calculated based on the surrounding points.¹¹⁻¹³ PLR was then used for classification of the eye. These methods can be regarded as an approximation to the statistically more principled SPROG method, since all in affect involve averaging

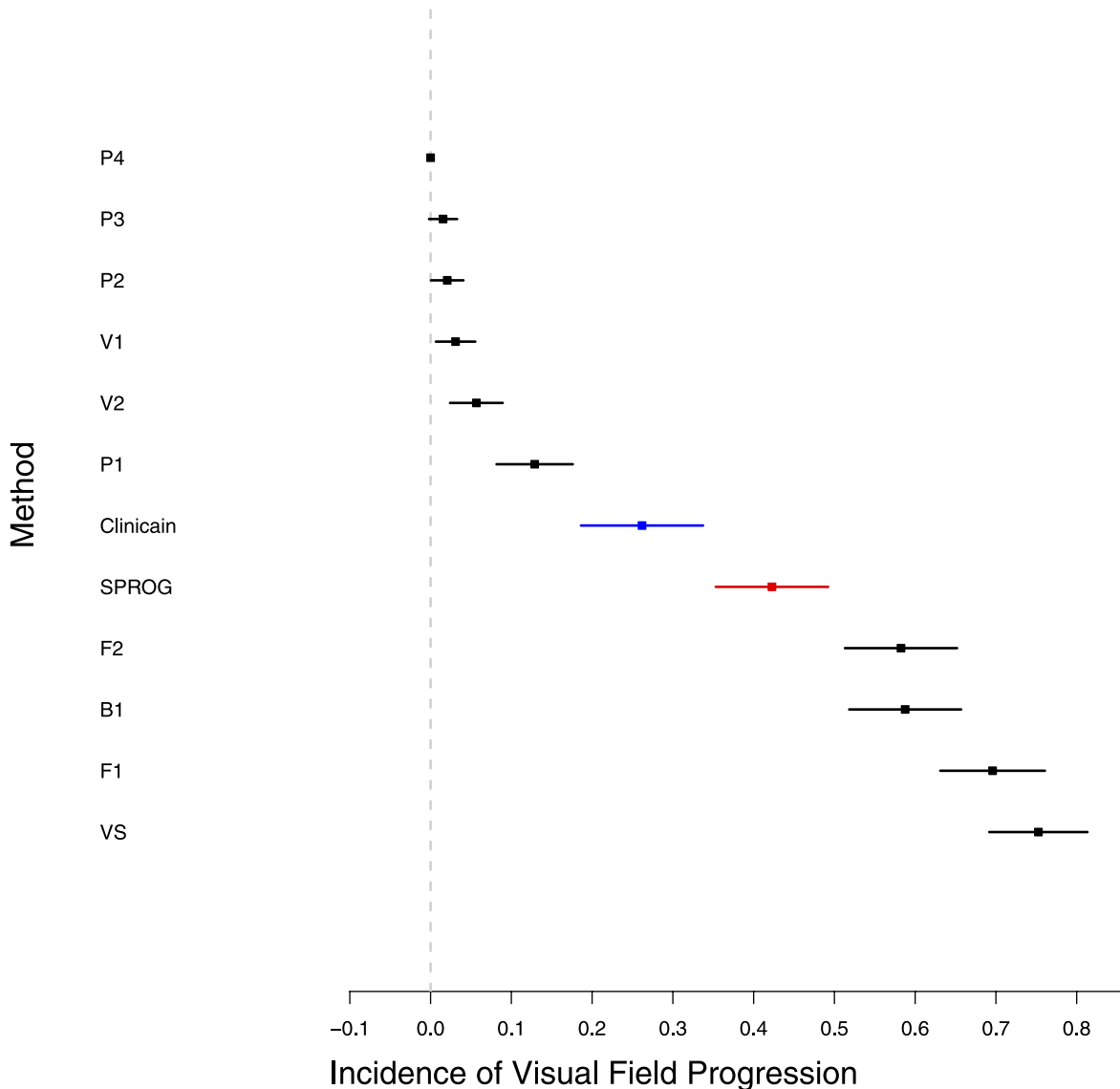


FIGURE 4. A ranking comparing PLR methods and our Bayesian CAR approach to determine Visual Field Progression. Refer to Table 1 for a description of methods. Clinician refers to the method described in the Clinical Judgment of Progression subsection in the Methods section and SPROG is our Spatial Progression model.

over local neighborhoods. However, our model has the additional advantage that these averages are calculated in conjunction with the other model terms (Equation 3) including measurement error, a temporal trend, and a spatial temporal trend. While previous spatial methods are able to account for spatial correlation and, thus, minimize measurement error, their benefits are still confined by the use of PLR for progression classification.

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

significance level should be made so that the null hypothesis has the same chance of being rejected, whether testing an individual case or over multiple cases.³⁴ This is done by dividing the significance level (Table 1) by the number of hypotheses being tested ($n = 52$). Without the correction, when significance is evaluated at the 5% level over 52 points, we would expect two to three points to show significant progression by chance; at 10%, we would expect approximately five points to be significant by chance.

We can compare our model with additional methods using a recent network meta-analysis.⁷ While this does give an idea of high- and low-ranking methods, it does not solve the problem of a lack of a gold standard.⁷ Other methods have to juggle between high sensitivity and high specificity by adjusting the various cutoffs. We propose our method reduces both false positives and false negatives by its novel approach to account for spatial correlation, thus, reducing the effects of high variation, which are largely responsible for false predictions. When comparing our model with the clinical reference, we see that our model is the closest of the conservative methods, as

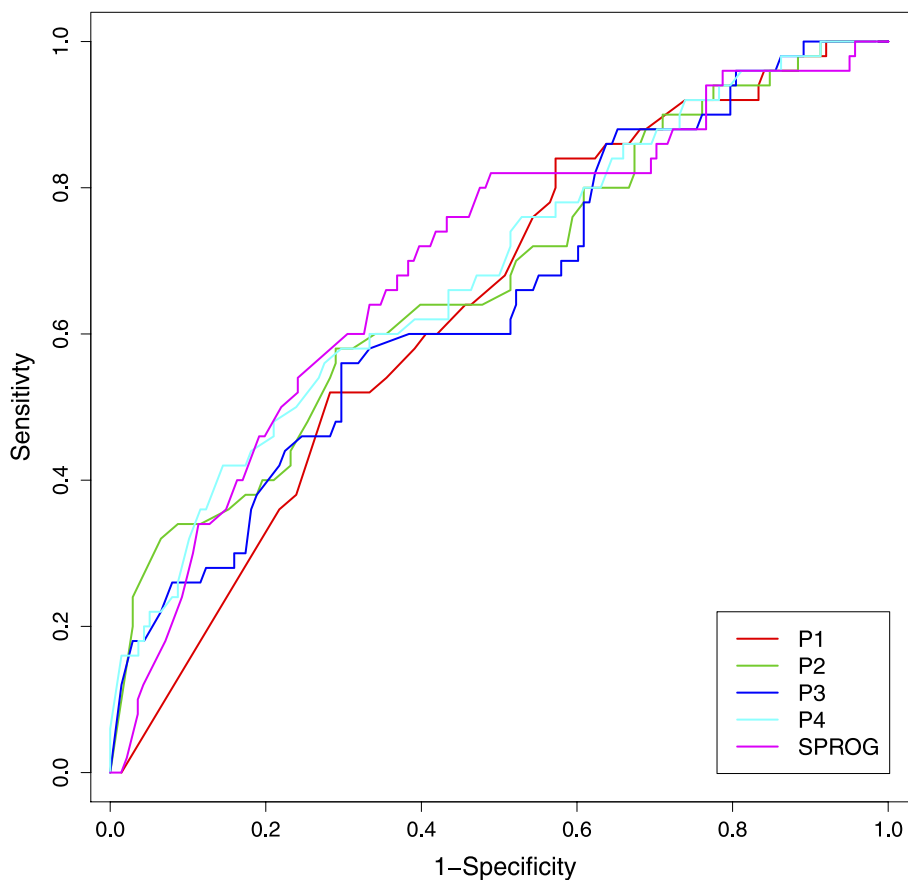


FIGURE 5. ROC curves comparing PLR methods with SPROG. AUC values for methods are as follows: P1 = 0.64, P2 = 0.67, P3 = 0.64, P4 = 0.67, SPROG = 0.69.

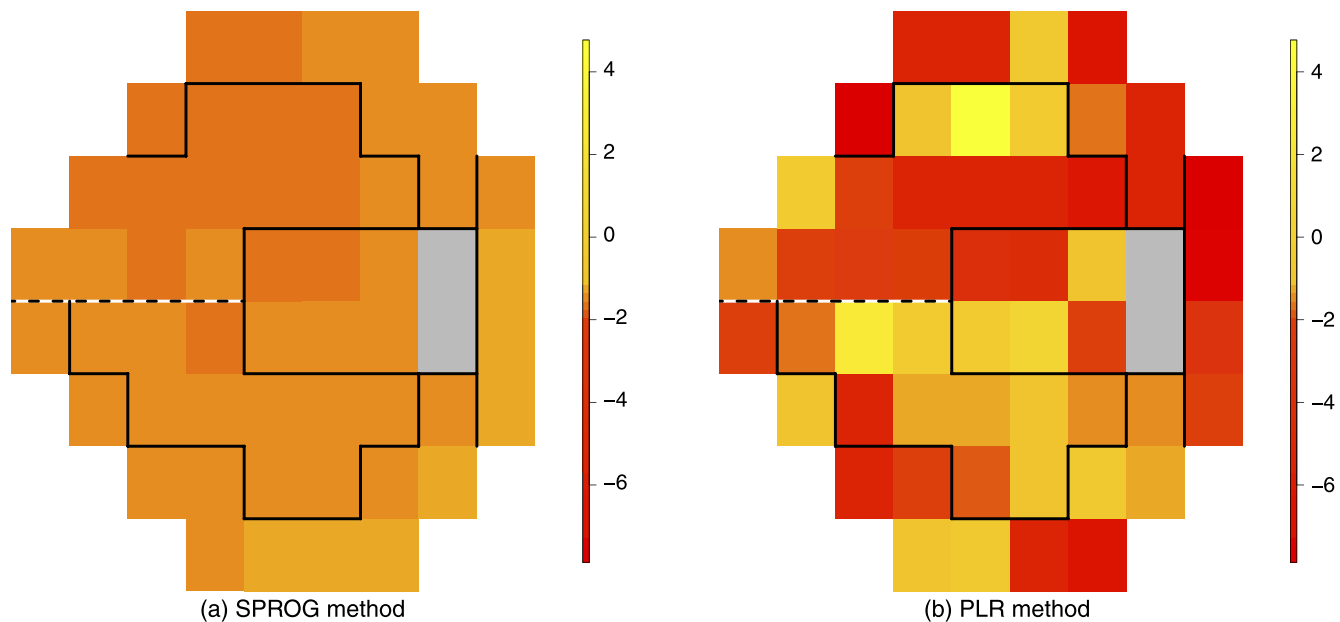


FIGURE 6. Heat maps showing local slope parameters for a single subject over the 52 evaluated points of the visual field. Slopes show rates of sensitivity change in decibels per year. *Black lines* separate the sectors and loci nearest the blind spot are shaded *gray*. (a) Highlights the smoothed results of the SPROG method, while (b) shows the variation of modeling points individually by the PLR method.

well as having the best overall ROC properties. We argue that it is better to slightly overestimate progression than underestimate it, as doing so could lead to a delay in treatment and potential visual loss in a patient.

In addition to providing an overall estimate of whether the eye is progressing, our method also estimates the rate at which the eye is progressing. This provides the clinician with more information and can assist when deciding how aggressively to treat the glaucoma. Furthermore, our model can provide estimated rates of progression for individual loci as well as sub regions of the eye, such as the sectors in Figure 1 or by hemifield. Another advantage of our method is that only two VFs are required to run the model. However, it should be noted that with data from less time points there is considerable uncertainty around parameter values, and the ability to identify marginal cases of progression is reduced. PLR methods require a minimum of three VFs for the model, and many criteria require significant results from three consecutive tests, which means that five VFs must be completed before an eye can be classified as progressing. Given that in clinical practice on average 1.7 VFs tests are taken in a year⁷ this could lead to a long delay before an eye is shown to be progressing.

Measurement error is known to be dependent on the distance from the fixation point in the VF. However, it is also noted that if a patient is highly variable at one point they are likely to be highly variable at another point.⁴ Measurement error is also known to be heterogeneous between fields dependent on glaucoma severity. When choosing how to model the magnitude of measurement error (σ_e), we looked at including this principle based on our repeated measurements data. We investigated the relationship between variance and distance from fixation, and variance, and mean. We found the power of the models to be low, therefore, in keeping to the principle of parsimony, we chose a constant value for measurement error. However, we acknowledge that this is an approach that may benefit from further research.

Another refinement to consider would be how the spatial correlation is modeled. The CAR model uses an adjacency matrix to define spatial correlation. This method of modeling spatial correlation was found to perform better and allow more flexibility than the alternative of accounting for spatial correlation by distance using parametric functions.³⁵ However, the method of modeling spatial correlation is particularly important when considering defects that straddle the sector boundaries defined by Garway-Heath et al.¹⁶ While the CAR model allows for correlation between the sectors, we have placed less weight on correlations between sectors than within. The weighting 0.3 for between sectors was chosen as it gave the best combination of sensitivity and specificity over all the eyes. For eyes with defects straddling the sectors it may be of benefit to use a more even weighting scheme. Investigating modeling spatial correlation via a distance based method may prove worthwhile.

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

Acknowledgments

The authors thank Alex Hewitt for evaluating the discrepant fields. The authors also thank the reviewers for their insightful comments, which greatly enhanced the manuscript.

References

1. Kwon YH, Fingert JH, Kuehn MH, Alward WLM. Primary open-angle glaucoma. *New Engl J Med*. 2009;360:1113-1124.
2. Kerr NM, Chew SS, Eady EK, Gamble GD, Danesh-Meyer HV. Diagnostic accuracy of confrontation visual field analysis. *Neurology*. 2010;74:1184-1190.
3. Atres PH, Chauhan BC. Longitudinal changes in the visual field and optic disc in glaucoma. *Prog Ret Eye Res*. 2005;24:333-354.
4. Heijl A, Lindgren G, Olsson J. Normal variability of static perimetric threshold values across the central visual field. *Arch Ophthalmol*. 1987;105:1544-1549.
5. 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.
6. Walsh TJ. *Visual Fields: Examination and Interpretation*. New York: Oxford University Press; 2011.
7. Ernest PJG, Viechtbauer W, Schouten JSAG, et al. The influence of the assessment of visual field progression in glaucoma: a network meta-analysis. *Acta Ophthalmol*. 2012;90:10-19.
8. Ernest PJG, Viechtbauer W, Schouten JSAG, et al. The evidence base to select a method for assessing glaucomatous visual field progression. *Acta Ophthalmol*. 2012;90:101-108.
9. Vesti E, Johnson CA, Chauhan BC. Comparison of different methods for detecting glaucomatous visual field progression. *Invest Ophthalmol Vis Sci*. 2003;44:3873-3879.
10. Rossetti L, Goni F, Denis P, Bengtsson B, Martinez A, Heijl A. Focusing on glaucoma progression and the clinical importance of progression rate measurement. *Eye*. 2010;24:S1-S7.
11. Fitzke FW, Crabb DP, McNaught AI, Edgar DF, Hitchings RA. Image processing of computerised visual field data. *Br J Ophthalmol*. 1995;79:207-212.
12. 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.
13. Gardiner SK, Crabb DP, Fitzke FW, Hitchings RA. Reducing noise in a suspected glaucomatous visual fields by using a new spatial filter. *Vis Res*. 2004;44:839-848.
14. Strouthidis NG, Scott A, Viswanathan AC, Crabb AP, Garway-Heath DE. Monitoring glaucomatous visual field progression: the effect of a novel spatial filter. *Invest Ophthalmol Vis Sci*. 2007;48:251-257.
15. Strouthidis NG, Scott A, Peter NM, Garway-Heath DE. Optic disk and visual field progression in ocular hypertensive subjects: detection rates, specificity and agreement. *Invest Ophthalmol Vis Sci*. 2010;47:2904-2910.
16. Garway-Heath DE, Poinoosawmy D, Fitzke FW, Hitchings RA. Mapping the visual field to the optic disc in normal tension glaucoma eyes. *Ophthalmology*. 2000;107:1809-1815.
17. Artes PH, Iwase A, Kitazawa Y, Chauhan BC. Properties of perimetric threshold estimates from full threshold, SITA standard and SITA fast strategies. *Invest Ophthalmol Vis Sci*. 2002;42:2654-2659.
18. Banerjee S, Carlin BP, Gelfand AE. *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton, FL: Chapman & Hall CRC Press; 2004.
19. Besag J, York J, Mollie A. Bayesian image restoration with two applications in spatial statistics. *Ann I Stat Math*. 1991;43:1-59.
20. Turpin A, Sampson GP, McKendrick AM. Combining ganglion cell topology and data of patients with glaucoma to determine a structure-function map. *Invest Ophthalmol Vis Sci*. 2009;50:3249-3256.

21. Zhu H, Crabb DP, Schlottmann PG, et al. Predicting visual function from the measurements of retinal nerve fiber layer structure. *Invest Ophthalmol Vis Sci*. 2010;51:5657-5666.
22. Lawson AB, Browne WJ, Vidal Rodeiro CL. *Disease Mapping with WinBUGS and MLwiN*. Chichester, West Sussex, England: Chapman & Hall CRC Press; 2003.
23. Gelman A, Jakulin A, Pittau MG, Su Y. A weakly informative default prior distribution for logistic and other regression models. *Annals of Applied Statistics*. 2008;2:1360-1383.
24. R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2005.
25. Gamerman D, Lopes H. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Boca Raton, FL: Chapman & Hall CRC Press; 2006.
26. Besag J, Kooperberg C. On conditional and intrinsic autoregressions. *Biometrika*. 1995;82:733-746.
27. Geweke J. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In: Bernardo JM, Dawid AP, Smith FM, eds. *Bayesian Statistics 4*. Oxford, UK: Clarendon Press; 1992:169-193.
28. Fitzke FW, Hitchings RA, Poinosawmy D, McNaught AI, Crabb DP. Analysis of visual field progression in glaucoma. *Br J Ophthalmol*. 1996;80:40-48.
29. 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.
30. Birch MK, Wishart PK, O'Donnell NP. Determining progressive visual field loss in serial Humphrey visual fields. *Ophthalmology*. 1995;102:1227-1235.
31. Smith SD, Katz J, Quigley HA. Analysis of progressive change in automated visual fields in glaucoma. *Invest Ophthalmol Vis Sci*. 1996;37:1419-1428.
32. Heijl A, Bengtsson B, Chauhan BC, et al. A comparison of visual field progression criteria of 3 major trials in early manifest glaucoma trial patients. *Ophthalmology*. 2008;115:1557-1565.
33. Heijl A, Lindgren A, Lindgren G. Test-retest variability in glaucomatous visual fields. *Am J Ophthalmol*. 1989;108:130-135.
34. Abdi H. Bonferroni test. In: Salkind NJ, Rasmussen K, eds. *Encyclopedia of Measurement and Statistics*. Thousand Oaks, CA: Sage Publications, Inc.; 2007:104-107.
35. Best N, Richardson S, Thomson A. A comparison of Bayesian spatial models for disease mapping. *Stat Methods Med Res*. 2005;14:35-59.