UCLA
UCLA Electronic Theses and Dissertations

Title
Bayesian Spatial Longitudinal Modeling for Local Rates of Glaucoma Progression

Permalink
https://escholarship.org/uc/item/578495rj

Author
Su, Erica

Publication Date
2023

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA

Los Angeles

Bayesian Spatial Longitudinal Modeling for Local Rates of Glaucoma Progression

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Biostatistics

by

Erica Su

2023

© Copyright by

Erica Su

2023

ABSTRACT OF THE DISSERTATION

Bayesian Spatial Longitudinal Modeling for Local Rates of Glaucoma Progression

by

Erica Su
Doctor of Philosophy in Biostatistics
University of California, Los Angeles, 2023
Professor Robert E. Weiss, Co-Chair
Professor Andrew J. Holbrook, Co-Chair

Timely detection of glaucoma progression is imperative to identify eyes for treatment and prevent further loss of vision. Methods to diagnose and monitor progression include routine use of structural and functional tests at multiple locations across the macula, the central part of the retina. Visual field (VF) testing provides functional measures of sensitivity to light while optical coherence tomography imaging gives structural thickness measurements of macular layers. In current practice, physicians assess progression by modeling functional or structural changes over time using simple linear regression (SLR) for each subject-location combination separately. This dissertation motivates and develops Bayesian hierarchical spatial longitudinal models to analyze structural and functional data from multiple subjects, borrowing strength across subjects and locations, to better detect glaucoma progression and predict future observations for individual subjects.

Chapter 1 gives an overview of the study objectives and summarizes the contributions of this dissertation. Chapter 2 presents a novel Bayesian hierarchical spatial longitudinal (HSL) model and compares its performance in estimating macular rates of structural change to the performances of SLR and a conditional autoregressive model. Notably in the simulation study, the HSL model is

more than three times as efficient as SLR in estimating local rates of change. To more explicitly model the spatial correlation in intercepts, slopes, and residual standard deviations (SD), Chapter 3 proposes a Bayesian hierarchical model with spatially varying random coefficients and visit effects. A comparison of the model to several nested models lacking different model components demonstrates the benefit of incorporating spatially varying visit effects in improving model fit and reducing prediction error. Chapter 4 extends the spatially varying coefficients approach to model VF data. This model simultaneously accounts for censoring and heteroskedasticity, which are inherent qualities of VF data, and the spatial structure in the data. Chapter 4 highlights the importance of using Gaussian processes with nonstationary covariance functions to better model spatial irregularities in the subject-level intercepts and slopes.

The dissertation of Erica Su is approved.

Donatello Telesca

Kouros Nouri-Mahdavi

Andrew J. Holbrook, Committee Co-Chair

Robert E. Weiss, Committee Co-Chair

University of California, Los Angeles

2023

To my parents and brother,

for their endless love and support.

TABLE OF CONTENTS

1 Introduction **1**

1.1 Overview 1

1.2 A Bayesian hierarchical spatial longitudinal model for structural data 3

1.3 A Bayesian spatially varying hierarchical random effects model for structural data . . 4

1.4 A Bayesian hierarchical spatially varying coefficients model for functional data . . . 4

1.5 Outline of dissertation 5

2 A Bayesian hierarchical spatial longitudinal model improves estimation of local macular rates of change in glaucomatous eyes **6**

2.1 Introduction 6

2.2 Methods 8

2.2.1 Macular OCT imaging 8

2.2.2 Data management and statistical modeling 9

2.2.3 Simulation study 12

2.3 Results 14

2.3.1 Simulation results 14

2.3.2 Analysis of AGPS data 18

2.4 Discussion 25

2.5 Supplementary material 30

2.5.1 Outlier removal algorithm 30

2.5.2 Bayesian hierarchical spatial longitudinal model specification 31

2.5.3 Bayesian conditional autoregressive model specification 34

2.5.4 Supplementary tables 37
2.5.5 Supplementary figures 50

3 A spatially varying hierarchical random effects model for longitudinal macular structural data in glaucoma patients 61

3.1 Introduction 61
3.2 Ganglion cell complex data 64
3.2.1 Macular optical coherence tomography 64
3.2.2 Advanced Glaucoma Progression Study 65
3.2.3 Data exploration 66
3.2.4 Modeling goals 71
3.3 Methods 73
3.3.1 Gaussian processes 73
3.3.2 Multivariate Gaussian processes 74
3.3.3 Model specification for a spatially varying hierarchical random effects with visit effects model 75
3.3.4 Priors 77
3.3.5 Computation and inference 78
3.3.6 Model comparison 79
3.4 Advanced Glaucoma Progression Study 80
3.5 Discussion 85
3.6 Supplementary material 87
3.6.1 Convergence assessment of the SHREVE model 87
3.6.2 Additional results for AGPS analysis 88

4 A hierarchical spatially varying coefficients model for longitudinal visual field data in glaucoma subjects 91

4.1 Introduction 91

4.2 Visual field data 94

4.2.1 Advanced Glaucoma Progression Study 94

4.2.2 Visual field threshold sensitivity 94

4.2.3 Heteroskedasticity 95

4.2.4 Spatial structure of test locations 96

4.2.5 Modeling criteria 97

4.3 Methods 98

4.3.1 Gaussian processes 98

4.3.2 Nonstationary covariance functions 99

4.3.3 Single eye model 100

4.3.4 Model specification for a hierarchical spatially varying longitudinal visual field model 102

4.3.5 Computation and inference 103

4.4 Advanced Glaucoma Progression Study 104

4.4.1 Single eye model results 104

4.4.2 HSLVF results 105

4.5 Discussion 109

Bibliography 112

LIST OF FIGURES

2.1 Histograms of mean squared error (MSE) (upper left), squared bias (upper right), and estimator variance (lower left) for patient-superpixel slopes comparing simple linear regression (SLR), conditional autoregressive (CAR), and hierarchical spatial longitudinal (HSL) models pooled across all superpixels from the simulation study. Mean squared error is squared bias plus estimator variance; the x-axis values are directly comparable in all three plots. Lower MSE indicates better model performance. Counts of large values omitted due to truncating the x-axis at 1: MSE (HSL: 43; CAR: 241; SLR: 500), squared bias (HSL: 31; CAR: 120; SLR: 0), and estimator variance (HSL: 0; CAR: 8; SLR: 492). 17

2.2 Heat map of the average root mean squared error (RMSE) ratio for simple linear regression (SLR) over hierarchical spatial longitudinal (HSL) or conditional autoregressive (CAR) over HSL for intercepts (top) and slopes (bottom) by superpixel location from the simulation study. HSL outperforms SLR and CAR in all superpixels as ratios greater than one favor HSL over the alternative model. 19

2.3 Heat map of the percentage of significant negative slopes (top) and significant positive slopes (bottom) detected by simple linear regression (SLR), conditional autoregressive (CAR), and hierarchical spatial longitudinal (HSL) models when the true slope is negative in the simulation study. 20

2.4 Heat map of the median (10th, 90th percentile) slope SD ratio as a function of superpixel location for the patient cohort data. The ratio was defined as either conditional autoregressive (CAR) over hierarchical spatial longitudinal (HSL) or simple linear regression (SLR) over HSL. Ratios greater than 1 indicate better performance for HSL. On average, HSL outperforms SLR at all superpixel locations. 23

2.5 Scatter plots of slope posterior means from simple linear regression model (SLR, y-axis) against those from the hierarchical spatial longitudinal model (HSL, x-axis) in each superpixel for the patient cohort data (right eye format). The SLR posterior means are much more variable than the HSL means. There is noticeable shrinkage towards the population mean in the HSL estimates in the peripheral superior and temporal regions. Each plot is square with its own axes with the x- and y- axes having the same range. The red dashed line represents the $x = y$ diagonal. 24

2.6 Histograms of mean squared error (MSE) (upper left), squared bias (upper right), and estimator variance (lower left) for patient-superpixel intercepts comparing simple linear regression (SLR), conditional autoregressive (CAR), and hierarchical spatial longitudinal (HSL) models pooled across all superpixels from the simulation study. Mean squared error is squared bias plus estimator variance; the x-axis values are directly comparable in all three plots. Lower MSE indicates better model performance. Counts of large values omitted due to truncating the x-axis at 4: MSE (HSL: 75; CAR: 271; SLR: 377), squared bias (HSL: 12; CAR: 68; SLR: 0), and estimator variance (HSL: 4; CAR: 25; SLR: 368). 50

2.7 Heat map of the average estimator standard deviations (SD) of intercepts (top) and slopes (bottom) by superpixel location for hierarchical spatial longitudinal (HSL) (left), conditional autoregressive (CAR) (middle), and simple linear regression (SLR) (right) models from the simulation study. 51

2.8 Heat map of the average root mean squared error (RMSE) of intercepts (top) and slopes (bottom) by superpixel location for hierarchical spatial longitudinal (HSL) (left), conditional autoregressive (CAR) (middle), and simple linear regression (SLR) (right) models from the simulation study. 52

2.9 Heat map of the average 95% Credible Interval Length of intercepts (top) and slopes (bottom) by superpixel location for hierarchical spatial longitudinal (HSL) (left), conditional autoregressive (CAR) (middle), and simple linear regression (SLR) (right) models from the simulation study. 53

2.10 Heat map of the 95% coverage probability minus 0.95 of intercepts (top) and slopes (bottom) by superpixel location for hierarchical spatial longitudinal (HSL) (left), conditional autoregressive (CAR) (middle), and simple linear regression (SLR) (right) models from the simulation study. 54

2.11 Heat map of the percentage of significant negative slopes (top) and significant positive slopes (bottom) detected by hierarchical spatial longitudinal (HSL) (left), conditional autoregressive (CAR) (middle), and simple linear regression (SLR) (right) models when the true slope is positive from the simulation study. 55

2.12 Scatter plots of slope posterior means from conditional autoregressive model (CAR, y-axis) against those from the hierarchical spatial longitudinal model (HSL, x-axis) in each superpixel for the patient cohort data. The CAR posterior means are more variable than the HSL means in temporal superpixels. There is noticeable shrinkage towards the population mean in the HSL estimates in the temporal regions. Each plot is square with its own axes with the x- and y- axes having the same range. The red dashed line represents the $x = y$ diagonal. 56

2.13 Scatter plots of slope posterior means from simple linear regression model (SLR, y-axis) against those from the hierarchical spatial longitudinal model (HSL, x-axis) in each superpixel for the patient cohort data. The SLR posterior means are much more variable than the HSL means. There is noticeable shrinkage towards the population mean in the HSL estimates in the peripheral superior and temporal regions. Each plot is square with the x- and y- axes having the same range across all 49 superpixels. The red dashed line represents the $x = y$ diagonal. 57

2.14 Scatter plots of slope posterior means from conditional autoregressive model (CAR, y-axis) against those from the hierarchical spatial longitudinal model (HSL, x-axis) in each superpixel for the patient cohort data. The CAR posterior means are more variable than the HSL means in temporal superpixels. There is noticeable shrinkage towards the population mean in the HSL estimates in the temporal regions. Each plot is square with the x- and y- axes having the same range across all 49 superpixels. The red dashed line represents the $x = y$ diagonal. 58

2.15 The proportion of significant negative slopes for each superpixel detected by hierarchical spatial longitudinal (HSL), conditional autoregressive (CAR), and simple linear regression (SLR) for the data from the patient cohort. The asterisk on the CAR bar indicates a significant p-value from McNemar's tests comparing proportions between HSL and CAR and similarly for the asterisk on the SLR bar for the comparison between HSL and SLR. 59

2.16 The proportion of significant positive slopes for each superpixel detected by hierarchical spatial longitudinal (HSL), conditional autoregressive (CAR), and simple linear regression (SLR) for the data from the patient cohort. The asterisk on the CAR bar indicates a significant p-value from McNemar's tests comparing proportions between HSL and CAR and similarly for the asterisk on the SLR bar for the comparison between HSL and SLR. 60

3.1 Visualization of the 8×8 grid of superpixels and labels from the Spectralis posterior pole algorithm. The inner 36 superpixels included in the analysis are shaded in gray and delineated with thicker lines. Superpixels are shown in right eye orientation where rows 1-4 are located in the *superior hemiretina* and rows 5-8 are located in the *inferior hemiretina*; the temple and nose are to the left and right, respectively. Superpixels labels are row number 1-8, a dot, then column number 1-8. The black dot indicates the foveal center for visual orientation. 66

3.2 Profile plots of ganglion cell complex (GCC) thickness measurements for 10 subjects across 36 superpixels against follow-up time in years since baseline visit. Each color represents a different subject. These profiles illustrate the variability in baseline GCC thickness across the 10 subjects within superpixels, with a range within a superpixel of up to 84 μm. The average baseline thicknesses over subjects vary across superpixels, generally increasing from the temporal to nasal regions (left to right). 67

3.3 Heat maps of ganglion cell complex (GCC) thickness measurements (μm) across 8 visits for 4 subjects for all 36 superpixels (top left 2.2 to bottom right 7.7). Each row is a different subject. The follow-up time of each visit is labeled at the top of each block. All maps share a common color scale for comparison. GCC measurements are highly correlated within subjects over time, illustrated by similar color patterns over time. The color patterns also highlight the spatial correlation between locations. GCC measurements are highly variable across subjects, as seen by the difference in color shades. Over time, the third row subject has noticeable thinning in many superpixels while the other subjects are more stable in comparison. 69

3.4 Heat maps of (a) estimated slopes (μm/year) and (b) residual standard deviations (SD) (μm) for the same 4 subjects as in Figure 3.3 using simple linear regressions of ganglion cell complex (GCC) thicknesses on time since baseline in each superpixel. Each column is a different subject. Estimated slopes appear spatially correlated within subjects. Subject 3 has particularly steep negative slopes in the upper half of the eye, while Subjects 1 and 2 have more stable slopes across superpixels. The estimated residual SDs vary within subject by superpixel location. Subjects 1 and 4 have more uniform residual SDs across locations while Subjects 2 and 3 have some superpixels with much higher residual SDs. 70

3.5 Empirical residual profile plots (superpixel mean subtracted from ganglion cell complex (GCC) thickness) for a single subject across 36 superpixels. There is an increase at four years for many superpixel locations suggesting visit-specific spatially correlated errors. . 72

3.6 Plate diagram of the proposed model. Blue nodes are latent variables, red nodes are observed variables, gray nodes are deterministic nodes, GP stands for *Gaussian process*, and MGP stands for *multivariate Gaussian process*. Plates are used to group variables repeated together over subjects, time, and space, where $i = 1,\ldots,N$ indexes subjects, $j = 1,\ldots,J_i$ indexes subject i's visits, and $k = 1,\ldots,K$ indexes superpixel locations. . . . 76

3.7 Comparison of predicted observations and model fit from the SHREVE model and simple linear regression (SLR) after holding out the last observation at 3.6 years follow-up of this subject. The gray line plots the raw data, the red line is the posterior mean fitted line from the SHREVE model without adding in the visit effects, and the blue line shows the fitted line from SLR. The SHREVE model is able to better estimate slopes and predict the last observation in noisy superpixels like 2.3 and 5.6 than SLR. 82

3.8 Posterior mean (line) and 95% pointwise credible intervals (colored bands) of correlation as a function of distance h between superpixels for subject-specific intercepts, slopes, and log residual SDs from the SHREVE (Visit Effects) and SHRE (No Visit Effects) models. The correlations decay faster in the SHREVE model with shorter lengthscales for slopes and log residual SDs. The dashed line indicates where the correlation is $\exp(-1)$ and the distance between superpixels is equal to the lengthscale in the exponential kernel. 84

3.9 Heat map of the log residual standard deviations (SD) comparing the SHREVE (Visit Effects) and SHRE (No Visit Effects) models. The values shown are the posterior mean (posterior SD) across the 36 superpixels. The log residual SDs from the SHREVE model are uniformly reduced across all superpixels compared to those from the SHRE model. The white dot is the fovea. 85

3.10 Bar charts of the proportion of significant negative slopes detected by the SHREVE model and simple linear regression (SLR) across the 36 superpixels. The difference ($\Delta = \text{SHREVE} - \text{SLR}$) in proportion is labeled at the top of each subplot. Across all locations, the SHREVE model detects a higher proportion of significant negative slopes (21.4% vs 18.0%) than SLR. 86

3.11 Plots of efficiency per iteration of the bulk effective sample size (ESS) on the left and potential scale reduction factor $\widehat{R}$ on the right for the 7 parameters with the largest $\widehat{R}$ from the SHREVE model. The bulk ESS increases linearly with increasing iterations while $\widehat{R}$ decreases exponentially with increasing iterations. The bulk ESSs were all greater than 100 per chain and $\widehat{R}$ were all less than 1.01. 88

3.12 Bar charts of the proportion of significant positive slopes detected by the SHREVE model and simple linear regression (SLR) across the 36 superpixels. The difference ($\Delta = \text{SHREVE} - \text{SLR}$) in proportion is labeled at the top of each subplot. Across all locations, the SHREVE model detects a lower proportion of significant positive slopes (3.1% vs 4.3%) than SLR. 90

4.1 Scatter plot of intercept against log residual standard deviation (SD) posterior means from simple linear regression of sensitivity in decibels (dB) against time from each subject-location combination. The blue line is a lowess curve. We leave out 4 points with intercepts greater than 35 dB and 450 points with constant 0 dB sensitivities over time. 96

4.2 Plot of the 10-2 visual field test locations before (left) and after (right) accounting for ganglion cell displacement. The distances are measured in degrees from the center. We analyze and visualize data in right eye orientation, where SN is superior nasal, ST is superior temporal, IN is inferior nasal, and IT is inferior temporal. 97

4.3 Scatter plot of (a) intercept posterior means in decibels (dB) and (b) slope posterior means (dB / year) from the Single Eye Stationary (SES, y-axis) vs. Single Eye Nonstationary (SENS, x-axis) models. The red dashed line is the $x = y$ diagonal. Some subject-location intercepts and slopes that differ greatly between the models are highlighted by the lighter colored dots. 105

4.4 Comparison of intercepts, slopes, and model fit between the Single Eye Stationary (SES) and Single Eye Nonstationary (SENS) models on two subjects. Heat maps of intercept posterior means in decibels (dB) across the 68 locations for (a) Subject 1 and (d) Subject 2. Heat maps of slope posterior means (dB / year) across the 68 locations for (b) Subject 1 and (e) Subject 2. Posterior mean fitted lines from the SES (red) and SENS (blue) models for the (c) superior hemiretina of Subject 1 and (f) inferior hemiretina of Subject 2. Posterior mean fitted lines from the SENS model follow the raw data more closely than SES in multiple locations. 106

4.5 Heat maps of posterior means of population-level (a) intercepts in decibels (dB) and (b) slopes (dB / year) across the 68 test locations from the HSLVF model. The asterisk indicates a significant negative slope, where we define significance when the upper bound of the 95% credible interval is less than 0. We use a significantly negative slope to indicate progression. 109

4.6 Heat maps of posterior means of the spatially varying standard deviation (SD) parameters $\lambda_{\beta 0k}$ and $\lambda_{\beta 1k}$ for the nonstationary Matérn covariance functions of the subject-level (a) intercepts in decibels (dB) and (b) slopes (dB / year), respectively, in the HSLVF model. There is more variability in random intercepts and slopes in the superior hemiretina than inferior. 110

LIST OF TABLES

2.1 Averages over all patient-superpixel intercepts and slopes of root mean squared error (RMSE), 95% credible interval length (CrIL), and 95% credible interval (CrI) coverage probability of intercepts and slopes for simple linear regression (SLR), conditional autoregressive (CAR), and hierarchical spatial longitudinal (HSL) models and the averages of the CAR/HSL and SLR/HSL ratios from the simulation study. Lower RMSE and 95% CrIL indicates better model performance and ratios greater than one favor HSL over CAR or SLR. The HSL model outperforms SLR for RMSE and 95% CrIL for both intercepts and slopes. CAR does substantially worse on 95% CrI coverage probability compared to HSL and SLR. 14

2.2 Summary of the grand mean (10th, 90th percentile) of posterior means and posterior standard deviations (SD) for intercepts and for slopes for all 5,419 patient-superpixel combinations for simple linear regression (SLR), conditional autoregressive (CAR), and hierarchical spatial longitudinal (HSL) models for the data from the patient cohort. The difference in posterior means for intercepts and slopes is defined as SLR minus HSL or CAR minus HSL. The ratio in posterior SDs is defined as SLR over HSL or CAR over HSL and we report the median (10th, 90th percentile) ratio of posterior SDs. 22

2.3 Clinical and demographic characteristics of study sample. SD = standard deviation; MD = mean deviation; IQR = interquartile range; GCC = ganglion cell complex. 37

2.4 Summary of the grand mean (10th, 90th percentile) of posterior means, posterior standard deviations (SD)s, estimator SDs, 95% Credible Interval Lengths (CrIL)s, root mean squared error (RMSE), and 95% Credible Interval (CrI) coverage probability for intercepts and for slopes for all 5,439 patient-superpixel combinations for simple linear regression (SLR), conditional autoregressive (CAR), and hierarchical spatial longitudinal (HSL) models for the simulation study. The difference in posterior means for intercepts and slopes is defined as CAR or SLR minus HSL. All ratios are defined as CAR or SLR over HSL and we report the median (10th, 90th percentile) ratios. One patient had 4 repeated measures and infinite posterior variance and was omitted from the posterior SD calculations, which represents a slight advantage on the scale for SLR over HSL. . . 38

2.5 The proportion of significant negative and positive slopes detected by simple linear regression (SLR), conditional autoregressive (CAR), and hierarchical spatial longitudinal (HSL) models by superpixel (SP) location from the simulation study. The proportions are given first for all patient-superpixel slopes, second when the true slope is negative, and third when the true slope is positive. The first row is the global summary across all superpixels. 40

2.6 The proportion of significant negative and positive slopes detected by simple linear regression (SLR), conditional autoregressive (CAR), and hierarchical spatial longitudinal (HSL) models by superpixel (SP) location from the simulation study. The proportions are given first for all patient-superpixel slopes, second when the true slope is negative, and third when the true slope is positive. The first row is the global summary across all superpixels. 42

2.7 The proportion of significant negative and positive slopes detected by simple linear regression (SLR), conditional autoregressive (CAR), and hierarchical spatial longitudinal (HSL) models by superpixel (SP) location from the simulation study. The proportions are given first for all patient-superpixel slopes, second when the true slope is negative, and third when the true slope is positive. The first row is the global summary across all superpixels. 44

2.8 The mean and percentile range (10th, 90th) of posterior means and standard deviations (SD) of slopes by superpixel (SP) location from the analysis of data from the patient cohort for simple linear regression (SLR), conditional autoregressive (CAR), and hierarchical spatial longitudinal (HSL) models. The first row is the global summary across all superpixels. 46

2.9 The mean and percentile range (10th, 90th) of posterior means and standard deviations (SD) of slopes by superpixel (SP) location from the analysis of data from the patient cohort for simple linear regression (SLR), conditional autoregressive (CAR), and hierarchical spatial longitudinal (HSL) models. The first row is the global summary across all superpixels. 48

3.1 Model fit comparison with widely applicable information criterion (WAIC), approximate leave-one-out cross-validation with Pareto Smoothed Importance Sampling (LOO), and mean squared prediction error (MSPE) of predictions. For predictions, we hold out 7 randomly sampled observations y_{iJ_ik} at the last visit of each of 110 AGPS subjects and 6 observations from one subject. Models with visit effects perform better than models without visit effects. SLR performs noticeably worse compared to the hierarchical models. The smallest WAIC, LOO, and MSPE values are bolded. 81

3.2 Posterior mean and 95% credible interval (CrI) for global parameters and subject-level multivariate Gaussian process (MGP) parameters comparing the SHREVE and SHRE models. 83

3.3 The mean minimum/maximum bulk effective sample size (ESS), tail ESS, and potential scale reduction factor $\widehat{R}$ for the SHREVE model broken down by parameter types. For convergence, $\widehat{R}$ should be < 1.01 and ESS should be > 100 per chain. The first column gives the parameter types: Hyperparameters include global parameters and multivariate Gaussian process hyperparameters; Population-level includes the population-level intercepts, slopes, and log residual standard deviations (SD) across all locations; Intercepts, Slopes, and Log Residual SDs include the subject-level intercepts, slopes, and log residual SDs across all locations, respectively; and Visit Effects include the visit effects across all locations. The second column gives the number of parameters summarized. . . 89

3.4 Posterior mean and 95% credible interval (CrI) for population-level multivariate Gaussian process (MGP) parameters for the SHREVE and SHRE models. SD stands for standard deviation. 89

4.1 The mean minimum/maximum bulk effective sample size (ESS), tail ESS, and potential scale reduction factor $\widehat{R}$ for the HSLVF model broken down by parameter types. For convergence, $\widehat{R}$ should be < 1.01 and ESS should be > 100 per chain. The first column gives the parameter types: Hyperparameters include global parameters and Gaussian process hyperparameters; Population-level includes the population-level intercepts, slopes, and log residual standard deviations (SD) across all locations; Intercepts, Slopes, and Log Residual SDs include the subject-level intercepts, slopes, and log residual SDs across all locations, respectively; and Log SD Process includes the spatially varying log SD for the subject-level intercepts and slopes across all locations. The second column gives the number of parameters summarized. 107

4.2 Posterior mean and 95% credible interval (CrI) for global parameters and Gaussian process (GP) parameters for the HSLVF model. SD stands for standard deviation. 108

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisors Robert Weiss and Andrew Holbrook. Without either of them, this work would not have been possible. I have been incredibly lucky to have advisors who are so supportive of me and dedicated to my professional development. I would also like to thank Kouros Nouri-Mahdavi and Donatello Telesca for their valuable time and agreeing to be on my committee. I thank Kouros for sharing his expertise on glaucoma research and for our meaningful collaborations. I want to thank all the members of the Glaucoma Advanced Imaging Laboratory. Thank you, Vahid, for answering all of my questions.

I would like to thank my friends for keeping me sane and getting me through graduate school. I thank my brother for all the fun things we do. I thank my parents for their love, encouragement, and support.

This work was supported by an NIH R01 grant (R01-EY029792) and used computational and storage services associated with the Hoffman2 Shared Cluster provided by UCLA Institute for Digital Research and Education's Research Technology Group.

VITA

2011 B.S. (Biomedical Engineering: Premedical), University of California, Irvine

2020 M.S. (Biostatistics), University of California, Los Angeles

2016–present Graduate Student Researcher, Department of Biostatistics, University of California, Los Angeles

SELECTED PUBLICATIONS

Shi, L., Mohammadi, M., Mohammadzadeh, V., **Su, E.**, Weiss, R. E., Caprioli, J., & Nouri-Mahdavi, K. (2023). Comparing rates of change in moderate to advanced glaucoma: retinal nerve fiber layer vs. Bruch's membrane opening-minimum rim width. *American Journal of Ophthalmology*.

Mohammadzadeh, V., **Su, E.**, Mohammadi, M., Law, S. K., Coleman, A. L., Caprioli, J., Weiss, R. E., & Nouri-Mahdavi, K. (2023). Association of blood pressure with rates of macular ganglion cell complex thinning in patients with glaucoma. *JAMA Ophthalmology*, 141(3), 251-257.

Mohammadi, M., **Su, E.**, Chew, L., Mohammadzadeh, V., Caprioli, J., Weiss, R. E., & Nouri-Mahdavi, K. (2023). Comparison of ganglion cell layer and inner plexiform layer rates of change in suspected and established glaucoma. *American Journal of Ophthalmology*, 249, 12-20.

Mohammadzadeh, V., **Su, E.**, Shi, L., Coleman, A. L., Law, S. K., Caprioli, J., Weiss, R. E., & Nouri-Mahdavi, K. (2022). Multivariate longitudinal modeling of macular ganglion cell com-

plex: spatiotemporal correlations and patterns of longitudinal change. *Ophthalmology Science*, 2(3), 100187.

Mohammadzadeh, V., **Su, E.**, Rabiolo, A., Shi, L., Zadeh, S. H., Law, S. K., Coleman, A., Caprioli, J., Weiss, R. E., & Nouri-Mahdavi, K. (2022). Ganglion cell complex: the optimal measure for detection of structural progression in the macula. *American Journal of Ophthalmology*, 237, 71-82.

Asarnow, R. F., Newman, N., Weiss, R. E., & **Su, E.** (2021). Association of attention-deficit/hyperactivity disorder diagnoses with pediatric traumatic brain injury: A meta-analysis. *JAMA Pediatrics*, 175(10), 1009-1016.

Su, M., Huynh, V., Bronsert, M., **Su, E.**, Goode, J., Lock, A., Banden, S., Ahrendt, G., Afghahi, A., Arruda, J., & Tevis, S. (2021). Longitudinal risk management for patients with increased risk for breast cancer. *Journal of Surgical Research*, 266, 421-429.

Mohammadzadeh, V., **Su, E.**, Zadeh, S. H., Law, S. K., Coleman, A. L., Caprioli, J., Weiss, R. E., & Nouri-Mahdavi, K. (2021). Estimating ganglion cell complex rates of change with Bayesian hierarchical models. *Translational Vision Science & Technology*, 10(4), 15.

DeVon, H. A., Daya, M. R., Knight, E., Brecht, M. L., **Su, E.**, Zègre-Hemsey, J., Mirzaei, S., Frisch, S., & Rosenfeld, A. G. (2020). Unusual fatigue and failure to utilize EMS are associated with prolonged prehospital delay for suspected acute coronary syndrome. *Critical Pathways in Cardiology*, 19(4), 206.

Li, M. J., **Su, E.**, Garland, W. H., Oksuzyan, S., Lee, S. J., Kao, U. H., Weiss, R.E., & Shoptaw, S. J. (2020). Trajectories of viral suppression in people living with HIV receiving coordinated care: differences by comorbidities. *Journal of Acquired Immune Deficiency Syndromes*, 84(4), 387.

CHAPTER 1

Introduction

This dissertation presents novel methodology to study and analyze data from patients with moderately advanced to severe glaucoma. We provide a brief introduction to essential topics for ease of understanding and give a summary of our contributions.

1.1 Overview

Glaucoma is a chronic and progressive disease that damages the optic nerve and is the second leading cause of blindness worldwide (Kingman, 2004). Glaucoma poses a significant worldwide public health burden. The number of people affected by glaucoma was estimated to be approximately 76 million in 2020 and is predicted to reach more than 110 million by 2040 (Allison et al., 2020; Tham et al., 2014). There is currently no cure for glaucoma and vision loss due to glaucoma is irreversible. Detection of disease progression is crucial to identify eyes for treatment and prevent vision from worsening.

The visual field (VF) refers to the field of vision or the entire space that is visible when the eyes are focused on a single point. Light entering the eye is absorbed by photoreceptor cells that activate ganglion cells located in the retina, the back of the eye. The ganglion cells form the nerve fiber layer, where the visual stimulus is transmitted to the optic nerve and relayed to the brain. Glaucoma leads to ganglion cell death and loss of optic nerve fiber. Typically as the disease progresses, patients experience loss in peripheral vision and eventually central vision (Weinreb et al., 2014).

Currently, methods to diagnose and monitor progression in more advanced glaucoma focus on

assessing structural or functional changes in the macula, the central area of the retina. The macula contains the highest density of photoreceptors and is responsible for sharp, central vision. Optical coherence tomography (OCT) is an imaging modality commonly used to obtain structural (thickness) measurements from the optic nerve head, retinal nerve fiber layer, and the macula. Routine macular OCT imaging enables ophthalmologists to monitor the thickness of macular layers and detect thinning indicative of structural damage over time. Alternatively, VF testing provides functional measures by assessing the patient's response to light sensitivities and serves as the gold standard for glaucoma detection.

Throughout this dissertation, we develop models for, and analyze data, from the Advanced Glaucoma Progression Study (AGPS) (Mohammadzadeh et al., 2021, 2022a,b). We obtain macular OCT measurements in a grid of 8×8 superpixel locations and VF measurements in a grid of 68 test locations. In current practice, physicians assess progression by modeling functional (VF loss) or structural (thinning of macular layers) changes over time using repeated simple linear regression (SLR) for each subject-location combination (Gardiner and Crabb, 2002; Nouri-Mahdavi et al., 2007; Tatham and Medeiros, 2017; Thompson et al., 2020a). These methods do not take into account the hierarchical structure, repeated measures within subjects, or spatial correlation of the data. Since there is no single, established method to evaluate rates of progression, there is a need for more complex models to properly analyze spatio-temporal structural or functional data.

Recently, some methods accounting for the hierarchical structure or spatial dependence have been proposed to model functional VF data. Betz-Stablein et al. (2013) propose a model using conditional autoregressive (CAR) priors to account for spatial correlations between test locations and VF clusters corresponding to sectors of the optic nerve head (Garway-Heath et al., 2000). Berchuck et al. (2019) introduce a model to account for localized spatio-temporal smoothing, allowing the spatial correlation structure to adapt over time. Montesano et al. (2021) develop a model to account for the hierarchical structure (cluster and location levels) and heteroskedasticity. These recent developments account for spatial structure but are fit separately for each eye, effectively ignoring population information. Bryan et al. (2017) describe a two-stage approach to fit a hierarchical model

taking subject, eye, hemifield (one half of the VF), and location levels into account, but this work ignores spatial correlations. Despite these recent advances, there is still a need for methods accounting for the hierarchical structure of the data, spatial correlations, and heteroskedasticity simultaneously. Moreover, studies of VF rarely account for inherent censoring and instead treat the censoring cutoff value as the actual measurement.

This dissertation provides methodology for ophthalmologists to better detect local structural and functional rates of change for individual subjects. To this end, we propose Bayesian hierarchical models to analyze data from all AGPS subjects simultaneously, borrowing information across subjects to reduce uncertainty in estimating individual slopes. The Bayesian framework provides a means to quantify the uncertainty in estimation, as subject-specific slopes are estimated parameters in the model. Location-level parameters summarize population characteristics at each location. Our models are appropriately tailored to the characteristics of structural and functional data to better estimate rates of progression.

1.2 A Bayesian hierarchical spatial longitudinal model for structural data

In Chapter 2, we model longitudinal macular ganglion cell complex (GCC) thickness measurements from 111 patients in the AGPS. We propose a novel Bayesian hierarchical spatial longitudinal (HSL) model to estimate local macular GCC rates of change over 49 superpixel locations simultaneously. The model includes global, superpixel-specific, patient-specific, and superpixel-patient specific parameters. To account for spatial variation, we employ semi-parametric spatial effects for patient-specific intercepts and slopes and visit effects. Visit effects account for spatially correlated measurement errors affecting all superpixels on the same visit.

To assess estimation performance of the HSL model on AGPS data, we compare superpixel-patient specific estimates and posterior variances from HSL to those from SLR and a Bayesian CAR model. Comparison to SLR is important as SLR is often used in practice. We implement a CAR model inspired by Betz-Stablein et al. (2013) for further comparison to a model more complex than

SLR. While we fit data from all patients together in HSL, SLR is fit separately for each superpixel-patient combination and the CAR model is fit separately for data from each eye (patient). Since we do not know the true rates of progression, we perform a simulation study to better compare the accuracy of the slope estimates between HSL, SLR, and CAR. In particular, we show that HSL estimates slopes more than twice as efficiently as CAR and three times as efficiently as SLR.

1.3 A Bayesian spatially varying hierarchical random effects model for structural data

In Chapter 3, we consider an improved spatial model for longitudinal GCC measurements. To better account for the spatial structure of measurement locations, we model superpixel-level and subject-specific parameters and visit effects as spatially varying coefficients with Gaussian process priors. Healthier eyes with more thickness at baseline have more potential for progression and faster rates of change (Rabiolo et al., 2020). In some macular layers, measurement variability increases at very low thicknesses (Miraftabi et al., 2016). We explicitly model the correlations between intercepts, slopes, and residual standard deviations (SD) to capture these relationships, further reducing uncertainty in slope estimates. To assess the performance of the proposed model on AGPS data, we compare the model to several nested models lacking different model components and to SLR. The proposed model with spatially varying visit effects vastly outperforms models without visit effects and SLR in terms of model fit and reduces error in predicting future observations.

1.4 A Bayesian hierarchical spatially varying coefficients model for functional data

In Chapter 4, we extend the spatially varying coefficients model for longitudinal VF measurements. VF sensitivities are reported in decibels (dB) of attenuation, ranging from 50 dB to 0 dB, with 0 dB being the brightest stimulus. There is inherent censoring and heteroskedasticity in VF data. Due

to the limitations of the measurement scale, VF sensitivities are censored at 0 dB, the maximal, brightest stimulus. Measurement variability increases with level of glaucoma damage. Due to the complex nature of VF data, analysis of VF progression is difficult. Our proposed model fits data from all AGPS patients while accounting for censoring, heteroskedasticity, and the spatial structure of test locations. To better accommodate spatial irregularities, we model the subject-level intercepts and slopes using Gaussian process priors with nonstationary covariance functions. We illustrate how spatially varying SD processes better capture the variability in subject-specific intercepts and slopes across locations. Our inclusion of data from the entire cohort and appropriate spatial modeling provide a means to better estimate subject-specific rates of change.

1.5 Outline of dissertation

We present the statistical models and detailed results, as described in Sections 1.2, 1.3, and 1.4, in Chapters 2, 3, and 4, respectively. Each chapter can be read as a standalone paper, each with its own introduction and notation. Some information may be repeated between chapters.

CHAPTER 2

A Bayesian hierarchical spatial longitudinal model improves estimation of local macular rates of change in glaucomatous eyes

2.1 Introduction

Detection of change over time is often important for the proper treatment of chronic diseases and is crucial to the management of glaucoma. Glaucoma is particularly suited to applying statistical and machine learning models to detect disease and to identify its progression, since most disease-related outcome measures are quantifiable (Thompson et al., 2020b; Liu et al., 2013; Abu et al., 2020; Mohammadzadeh et al., 2022a; Kass et al., 2002; Gordon et al., 2002). Relevant outcomes include the optic disc size, neuroretinal rim, retinal nerve fiber layer, or macular thickness measures along with a host of quantitative variables related to visual fields.

Global, regional, or local structural and functional measures have frequently been used as metrics to detect differences of a specified amount from baseline, commonly called event analyses, or to estimate rates of change, also known as trend analyses (Rao et al., 2013; Wu and Medeiros, 2018; Rabiolo et al., 2020). Longitudinal analyses of local structural or functional measures frequently rely on repeated simple linear regression (SLR) of such measures from single eyes against time to estimate eye-specific rates of change at macular superpixels, optic disc or retinal nerve fiber layer sectors or visual field test locations or clusters (Rabiolo et al., 2020; Mohammadzadeh et al., 2020b; Medeiros et al., 2012; Mansouri et al., 2011; Gracitelli et al., 2014). While SLR is easy to do and seems intuitive, there are multiple issues with this approach, which while occasionally acknowledged, have mostly been ignored (Montesano et al., 2021; Swaminathan et al., 2022; Zhu et al.,

2014b). Simple linear regression models ignore the spatial relationships among local structural (sectors, superpixels) or functional (visual field clusters or locations) measures. Another shortcoming of SLR is that valuable population information from the cohort is not used to refine estimated rates of change. Similarly, there is no formal way to specify the correlation of baseline measurements (the intercepts) with slopes (rates of change). This is an important issue as baseline measurements influence the magnitude of rates of change in an OCT sector or superpixel (Rabiolo et al., 2020). Accounting for baseline structural measurements can lead to a reduction of the estimated variability for rates of change (slopes). A few prior studies have proposed linear mixed or hierarchical models to address some of the above shortcomings (Montesano et al., 2021; Swaminathan et al., 2022; Russell et al., 2012b; Medeiros et al., 2011). For example, Montesano et al. (2021) applied a hierarchical linear mixed model to 24-2 visual field data to estimate global and local rates of change in individual eyes. To account for spatial correlation, Betz-Stablein et al. (2013) developed a model using conditional autoregressive (CAR) priors on intercepts and slopes for individual-level 24-2 visual field data. However, in either case, no attempt was made to include the cohort's population data to help estimate individual eye parameters. The cohort's data has information on the distribution of possible slopes and using this information as a prior results in more accurate slope estimates. Such models are challenging to specify correctly and run and can require a significant investment of time and computer CPU power.

Our team has developed several versions of a novel Bayesian hierarchical spatial longitudinal (HSL) model to improve estimation of local macular thickness rates of change in a prospective cohort of glaucoma patients (Mohammadzadeh et al., 2022a, 2021). This ongoing project aims to provide a longitudinal framework to estimate global and most importantly local rates of change more precisely across the macula within individual eyes while at the same time overcoming the inadequacies of the SLR approach. Each successive model improves on the previous version with additional modeling features. The current model adds spatially correlated subject specific random effects and spatially correlated visit effects. The goal of the current work is to compare the performance of the latest HSL model to that of SLR and a CAR model. We hypothesized that the HSL model would provide

more accurate estimates of local macular rates of change thanks to reduced variance for those rates of change compared to SLR and more appropriate shrinkage towards the population averages by location than CAR.

2.2 Methods

One hundred eleven eyes (111 patients) from the Advanced Glaucoma Progression Study (AGPS) were included in this study. The AGPS is an ongoing longitudinal study at the University of California Los Angeles. University's Institutional Review Board (IRB) approval was obtained for this study. The study adhered to the tenets of the Declaration of Helsinki and conformed to Health Insurance Portability and Accountability Act policies. All patients provided written informed consent at the time of enrollment in the study. Inclusion criteria for enrolled eyes were: a) clinical diagnosis of primary open-angle glaucoma, pseudoexfoliative glaucoma, pigmentary glaucoma, or primary angle-closure glaucoma; b) evidence of either central damage on 24-2 visual field (VF), defined as two or more points within the central 10° with $p < 0.05$ on the pattern deviation plot or VF mean deviation (MD) worse than -6 dB. Exclusion criteria consisted of baseline age less than 40 years or greater than 80 years, best-corrected visual acuity $< 20/50$, refractive error exceeding 8 diopters of sphere or 3 diopters of cylinder, or significant retinal or neurological disease affecting OCT measurements. Eyes with ≥ 4 optical coherence tomography (OCT) scans and ≥ 2 years of follow-up were included. We analyzed observations up to 4.25 years after baseline. Data from visits less than 0.2 years after a previous visit were omitted (Mohammadzadeh et al., 2021).

2.2.1 Macular OCT imaging

Macular volume scans were obtained with a Spectralis spectral-domain OCT (Heidelberg Engineering, Heidelberg, Germany). The Posterior Pole Algorithm of the Spectralis OCT acquires $30° \times 25°$ volume scans of the macula (61 B-scans spaced approximately 120 μm apart) centered on the fovea and repeated 9-11 times to reduce speckle noise. Proprietary software of Spectralis OCT, the Glau-

coma Module Premium Edition, was used to automatically segment individual retinal layers before data export. Images were reviewed for segmentation errors and image artifacts. Segmentation errors were manually rectified with the OCT device's built-in software. After segmentation, the individual layer thickness measurements are provided as 8×8 arrays of $3° \times 3°$ superpixels for the central $24° \times 24°$ region centered on the fovea. Due to substantial measurement noise in the peripheral macular regions, we only included the ganglion cell complex (GCC) thickness within 7×7 arrays of superpixels from Spectralis OCT macular volume scans after excluding the most inferior row and nasal column of superpixels. The GCC thickness was calculated by adding the thickness measurements of the retinal nerve fiber layer, ganglion cell layer, and inner plexiform layer.

2.2.2 Data management and statistical modeling

Our methods have been described previously (Mohammadzadeh et al., 2021). We identified potential outliers as observations with very large increases or decreases between consecutive measurements and removed observations that produced the greatest reduction in the sum of absolute consecutive differences resulting in removing approximately 0.5% of observations as outliers. We removed all observations for a single person in a single superpixel if we identified 2 or more outliers in that superpixel (Mohammadzadeh et al., 2021). We then fit a Bayesian normal hierarchical random effects model using the JAGS package in R (R2jags) (R Core Team, 2021; Su and Yajima, 2021). The current version of our Bayesian HSL model includes (a) a macula-wide global intercept and slope, (b) superpixel-specific intercept and slope deviations from the global intercept and slope (superpixel-level random effects), (c) macula-wide patient-specific intercept and slope effects (global patient random effects), (d) patient-superpixel specific intercept and slope random effects, and (e) macula-wide visit effects. Letting y_{ijk} denote a single observation of GCC thickness (μm) for patient i at time t_{ij} in superpixel k, the model is

$$y_{ijk} = \mu_0 + \mu_1 t_{ij} + \alpha_{0k} + \alpha_{1k} t_{ij} + \beta_{0ik} + \beta_{1ik} t_{ij} + \phi_{0k} \mathrm{PC}_{0i} + \phi_{1k} \mathrm{PC}_{1i} t_{ij} + \phi_{2k} \mathrm{VE}_{ij} + \varepsilon_{ijk}$$

$$\varepsilon_{ijk} \sim N(0, \sigma_k^2 \exp(\mathrm{PC}_{2i}))$$

where μ_0 is the global macula-wide intercept; μ_1 is the global macula-wide slope; α_{0k} is the population average intercept in superpixel k; α_{1k} is the population average slope in superpixel k; β_{0ik} is the patient-superpixel intercept for the ith patient; β_{1ik} is the patient-superpixel interaction slope for the ith patient; ϕ_{0k} is the spatial effects eigenvector component for the patient-specific intercepts; ϕ_{1k} is the spatial effects eigenvector component for the patient-specific slopes; ϕ_{2k} is the spatial visit effects eigenvector component; PC_{0i} is the macula-wide patient-specific intercept random effect; PC_{1i} is the macula-wide patient-specific slope random effect; PC_{2i} is the macula-wide patient-specific log residual variance random effect; VE_{ij} is the macula-wide visit random effect; the population component of the patient-superpixel log residual variance is $\log \sigma_k^2$. Our model employs novel, semi-parametric spatial effects for the intercept, $\phi_{0k}\text{PC}_{0i}$, slope $\phi_{1k}\text{PC}_{1i}$, residual variance $\sigma_k^2 \exp(\text{PC}_{2i})$, and visit effects $\phi_{2k}\text{VE}_{ij}$. In each component, terms with a k subscript ϕ_{0k}, ϕ_{1k}, σ_k^2, and ϕ_{2k} define the spatial pattern of variation across the macula while the random effects PC_{0i}, PC_{1i}, $\exp(\text{PC}_{2i})$, and VE_{ij} are random subject effects, or visit effects, that indicate how much subject i's observations deviate from the population averages. Each of these 4 terms are factor analytic models with 1 factor each where the ϕ_{0k}, ϕ_{1k}, σ_k^2, and ϕ_{2k} are called *loadings* and the random effects are also called *factors* (Kline, 2014). Superpixel-specific population intercepts and slopes are the sum of the macula-wide global intercept and slope plus the superpixel-specific intercept and slope deviations. Similarly, a superpixel-patient slope (intercept) is the sum of (a), (b), (c), and (d): the sum of the macula-wide global slope (intercept), superpixel-specific slope (intercept), the macula-wide patient-specific intercept slope (intercept), and the patient-superpixel specific slope (intercept) random effect. Patient-superpixel log residual variances are modeled as the sum of a superpixel component and a patient-specific component. The outlier removal algorithm and full model and priors are given in the supplementary material in Sections 2.5.1 and 2.5.2, respectively.

In Bayesian models, inference is made by summarizing the posterior distributions of the parameters of interest. The posterior distribution is obtained by combining the likelihood (information in the data) with the prior (prior knowledge about the unknown parameters) and quantifies the uncertainty in the parameters after observing the data. The posterior mean, i.e., mean of the posterior

distribution of a parameter of interest, is often taken as the Bayesian point estimate; similarly, the posterior standard deviation (SD), that is, the SD of the posterior distribution, summarizes the uncertainty in the posterior. Posterior means and SDs of superpixel-patient specific intercepts and slopes were calculated from the HSL model and compared to superpixel-patient specific intercepts and slopes obtained from a Bayesian SLR model with a flat prior analyzing longitudinal data from each superpixel-patient separately and a Bayesian CAR model with intrinsic CAR priors analyzing data from each patient separately.

The Bayesian SLR model produces posterior means and posterior Credible Intervals (CrI) that are identical to the classical point estimate and confidence intervals (CI) from classical least squares regression on the same data. The Bayesian SLR posterior SDs are slightly larger than classical SLR standard errors (SEs) by a factor of $\left(\frac{n_i-4}{n_i-2}\right)^{1/2}$ where n_i is the number of observations for patient i because SEs are not standard deviations and SEs do not account for the classical estimates being t distributed with n_i-2 degrees of freedom (df). Using the posterior SD from this Bayesian SLR model puts SLR, CAR and HSL on an equal footing.

For the CAR model inspired by Betz-Stablein et al. (2013), we assumed GCC thicknesses in neighboring superpixels would be more similar than in more distant superpixels. We defined the neighborhood structure, or adjacency matrix, such that superpixels sharing an edge are weighted as 1. For each patient i, we model spatial dependence in intercepts and slopes

$$y_{jk} = \alpha_0 + \alpha_1 t_j + \beta_{0k} + \beta_{1k} t_j + \varepsilon_{jk}$$
$$\varepsilon_{jk} \sim N(0, \sigma^2)$$

where α_0 is the overall eye intercept; α_1 is the overall eye slope; β_{0k} is the superpixel-specific spatial deviation from the overall eye intercept; β_{1k} is the superpixel-specific spatial deviation from the overall eye slope; and σ^2 is the residual variance specific to patient i. We fit the CAR models using the R package NIMBLE v0.13.1 (de Valpine et al., 2017, 2022). The full CAR model and priors are given in the supplementary material in Section 2.5.3 .

We identified superpixel-patient specific rates of change as worsening or improving (significantly negative or positive) when the upper or lower limit of the symmetric 95% CrI was less than or greater than 0, respectively. We compared the proportion of significant negative or significant positive slopes at each superpixel from the HSL and SLR models using an $\alpha = 0.05$ level classical McNemar's test with the null hypothesis being that the 2 proportions were equal.

We summarize differences (SLR minus HSL and CAR minus HSL) of posterior means and of SDs for individual patient-superpixel intercepts and slopes. We compared differences of posterior means averaged within each of 49 superpixels and also averaged across all 5,419 patient-superpixel profiles. We similarly compared the SLR/HSL and CAR/HSL ratios of posterior SDs for superpixel averaged intercepts and slopes, and patient-superpixel intercepts and slopes. Ratios of posterior variances (i.e., ratios of squared SDs) are measures of the improved estimation efficiency of the better model. Individual patient-superpixel intercept and slope posterior means, posterior SDs, and 95% CrI lengths from HSL, CAR, and SLR were plotted against each other to compare the inferences from the three approaches. A *patient-superpixel profile* is the set of measurements over time for a single patient in a single superpixel. We omitted 18 patient-superpixel profiles in the data cleaning step because there were two or more outliers in a single patient-superpixel profile and omitted 2 more patient-superpixel profiles with constant GCC at all time points, as SLR, unlike the HSL model or CAR, is unable to provide an appropriate inference. Thus 5,419 is slightly smaller than $111 \times 49 = 5,439$ patient-superpixel profiles.

2.2.3 Simulation study

We ran a simulation study to provide comparisons between the accuracy of HSL, CAR, and SLR estimates for rates of change (slopes); this comparison is only possible in a simulation study where true individual eye slopes are known. In the initial step, a true slope and intercept for each superpixel-patient combination (49 superpixels in each of 111 patient-eyes) was set by sampling each from a normal distribution with mean equal to the sum of the superpixel-specific population and macula-wide patient-specific parameters and variance equal to the superpixel intercept and slope random

effects variance. All superpixel-specific and patient-specific parameters were set as the posterior means from our model fit to the actual data; log residual variance was set as the posterior mean of the patient plus superpixel log components; visit times were copied from the data set and visit effect parameters were set to the posterior mean of those parameters. In the next step, we generated 100 data sets with random residual errors and visit effects.

We recorded bias, estimator variance, and root mean squared error (RMSE) from the simulation for all patient-superpixel intercepts and slopes. We took the posterior means from the HSL and CAR models as the estimates, and the least squares estimate was the SLR estimate. Bias was calculated as the *average estimate over simulated datasets minus truth*, which is a measure of whether the model estimates the true value on average. Estimator variance is the variance over simulations of the estimate around its average estimate over simulated data sets, a measure of precision. Estimator variance is estimator SD squared. Root mean squared error (RMSE) is the square root of the sum of squared bias and estimator variance

$$\text{RMSE} = \sqrt{\frac{\sum_{m=1}^{100} \text{estimate}_m - \text{truth}_m}{100}} = \sqrt{\text{bias}^2 + \text{estimator variance}}$$

where m indexes data sets. The RMSE is the gold standard for assessing the accuracy of a model for estimating the true parameter value. We calculated average over simulations of 95% Credible Interval Length (CrIL) which is the length of the interval with 2.5% posterior probability to the left and to the right of the interval for HSL and CAR and the usual 95% CI for SLR. Finally, 95% CrI Coverage is the proportion of the time where the 95% CrI contains the truth. For bias, estimator SD, RMSE, and 95% CrIL, smaller is better. For 95% CrI Coverage, values close to 95% are preferred. For the estimator SD, RMSE, and 95% CrIL, we report the SLR/HSL and CAR/HSL ratios; ratios > 1 represent better performance of HSL. We report superpixel summaries by averaging reported measures over all patients at each superpixel location and global summaries by averaging each component over all patients and superpixel locations.

Table 2.1: Averages over all patient-superpixel intercepts and slopes of root mean squared error (RMSE), 95% credible interval length (CrIL), and 95% credible interval (CrI) coverage probability of intercepts and slopes for simple linear regression (SLR), conditional autoregressive (CAR), and hierarchical spatial longitudinal (HSL) models and the averages of the CAR/HSL and SLR/HSL ratios from the simulation study. Lower RMSE and 95% CrIL indicates better model performance and ratios greater than one favor HSL over CAR or SLR. The HSL model outperforms SLR for RMSE and 95% CrIL for both intercepts and slopes. CAR does substantially worse on 95% CrI coverage probability compared to HSL and SLR.

Model	RMSE	95% CrIL	95% CrI Coverage Probability
Intercept			
HSL	1.07	4.21	0.949
CAR	1.31	4.24	0.882
SLR	1.44	7.17	0.951
Ratio (CAR/HSL)	1.25	1.03	
Ratio (SLR/HSL)	1.38	1.70	
Slope			
HSL	0.38	1.55	0.947
CAR	0.58	1.59	0.810
SLR	0.69	3.44	0.951
Ratio (CAR/HSL)	1.65	1.08	
Ratio (SLR/HSL)	2.05	2.25	

2.3 Results

A total of 39,625 GCC superpixel measurements in 49 superpixels from 815 visits were included in the analysis. Supplementary Table 2.3 presents clinical and demographic characteristics of the study sample. Mean (SD) baseline 10-2 visual field mean deviation was -8.9 (5.9) dB. Mean (SD) follow-up time was 3.6 (0.4) years with an average (SD) of 7.3 (1.1) OCT scans per eye.

2.3.1 Simulation results

Table 2.1 presents the average RMSE, 95% CrIL, and 95% CrI coverage probability for SLR, CAR, and HSL averaged over all 5,439 patient-superpixel for intercepts and for slopes. Compared to SLR, the HSL model had substantially better RMSE (intercept SLR/HSL mean ratio: 1.38; slope SLR/HSL mean ratio: 2.05) and 95% CrIL (intercept SLR/HSL mean ratio: 1.70; slope SLR/HSL

mean ratio: 2.25), and similar 95% CrI coverage probability. Compared to CAR, the HSL model had better RMSE (intercept CAR/HSL mean ratio: 1.25; slope CAR/HSL mean ratio: 1.65), similar 95% CrIL, and much better 95% CrI coverage probability (HSL vs. CAR coverage probability for intercepts: 0.949 vs. 0.882; for slopes: 0.947 vs. 0.810).

Patient-superpixel intercepts and slopes. Supplementary Table 2.4 presents summaries of the grand mean and 10th to 90th percentile range of posterior means, posterior SDs, estimator SDs, 95% CrILs, RMSEs, and 95% CrI coverage probabilities for intercepts and slopes across all 5,439 patient-superpixel combinations for HSL, CAR, SLR and the differences between the means (SLR − HSL and CAR − HSL) and ratios (SLR/HSL and CAR/HSL) of the SDs, CrILs, and RMSEs.

Intercepts- For the 5,439 individual patient-superpixel intercepts, SLR, CAR, and HSL models demonstrated similar average, 10th-, and 90th-percentile posterior means. The SLR model had larger (worse) estimator SDs (ratio range: 1.12 to 2.31) and larger (worse) 95% CrILs (ratio range: 1.23 to 3.09) than HSL for all patient-superpixels. While CAR had similar posterior SDs and 95% CrILs to HSL, HSL had higher (better) 95% CrI coverage probabilities for 80.2% (4,364 out of 5,439) of patient-superpixel intercepts. Compared to SLR and CAR, the HSL model had lower (better) RMSE values for 95.1% (5,174/5,439) and 92.7% (5,028/5,439) of patient-superpixel intercepts, respectively.

Slopes- Averaged across all 5,439 estimated patient-superpixel slopes, SLR, CAR, and HSL had similar average slope estimates. However, the 10th to 90th percentile range of SLR slope posterior means (-1.36, 0.48 μm/year) was much wider than HSL (-1.17, 0.19 μm/year) and CAR (-1.01, 0.14 μm/year). The SLR model had larger (worse) estimator SDs than HSL (SLR/HSL ratio range: 1.17 to 11.84) and 95% CrILs (SLR/HSL ratio range: 1.32 to 6.44) for all patient-superpixel slopes. The CAR model had larger estimator SDs than HSL (CAR/HSL ratio range: 0.71 to 7.57) with comparable 95% CrILs (CAR/HSL ratio range: 0.55 to 2.75); however, HSL had better 95% CrI coverage probabilities for 4,965/5,439 (92.1%) patient-superpixel slopes. The average 95% CrI coverage probabilities for patient-superpixel slopes were 0.951, 0.810, and 0.949 for SLR, CAR, and HSL, respectively. Compared to SLR and CAR, the HSL model had lower RMSE values for

5,211/5,439 (95.8%) and 5,097/5,439 (93.7%) patient-superpixel slopes, respectively. Figure 2.1 plots histograms of mean squared error (MSE), squared bias, and estimator variance for patient-superpixel slopes for all patients and superpixels for the SLR (blue), CAR (green), and HSL (red) models. The HSL model had noticeably smallest MSE due to having much smaller estimator variance compared to CAR and SLR. Equivalent histograms are plotted for the intercepts in Supplementary Figure 2.6, where the HSL model also demonstrates noticeably smaller MSE compared to CAR and SLR.

Superpixel-averages of intercepts and of slopes. For the intercepts, there was significant bias averaged across patients for HSL (range: -0.054 to 0.082 μm) and CAR (range: -0.486 to 0.290 μm) for 28/49 and 45/49 superpixels, respectively, while SLR had no bias in any superpixels. While the magnitude of the observed bias for HSL was not clinically relevant both in comparison to the size of the intercepts (posterior mean range: 52.9 to 101.6 μm) and compared to the posterior SDs, the magnitude of bias for CAR was more comparable to the posterior SDs. In contrast, SLR had substantially larger (worse) estimator SD (ratio range: 1.29 to 1.74), RMSE (ratio range: 1.19 to 1.58), and 95% CrILs (ratio range: 1.47 to 1.94) than HSL for all 49 superpixels. The CAR model had larger estimator SD (ratio range: 0.97 to 1.40) and RMSE (ratio range: 1.15 to 1.36) than HSL across superpixels, but similar 95% CrILs (ratio range: 0.76 to 1.33). There were no important differences (fraction coverage $-$ 0.95) in 95% CrI coverage probability for HSL and SLR, but noticeable differences in parafoveal and nasal superpixels for CAR (HSL range: -0.014 to 0.008; CAR range: -0.169 to 0.012; SLR range: -0.004 to 0.006). Superpixel averages of estimator SD, RMSE, 95% CrIL, and 95% CrI coverage probability for intercepts and slopes are shown in Supplementary Figures 2.7, 2.8, 2.9, and 2.10, respectively.

For the slopes, there was significant bias for HSL (range: -0.040 to 0.038 μm/year) and CAR (range: -0.155 to 0.252 μm/year) for 27/49 and 45/49 superpixels, respectively. The magnitude of the bias for HSL was generally modest in comparison to the slopes (range: -0.982 to 0.025 μm/year) and was small in comparison to posterior SDs; however, the magnitude of the bias for CAR was on par with the slope estimates. For the 49 superpixel slopes, SLR had larger (worse) estimator SD

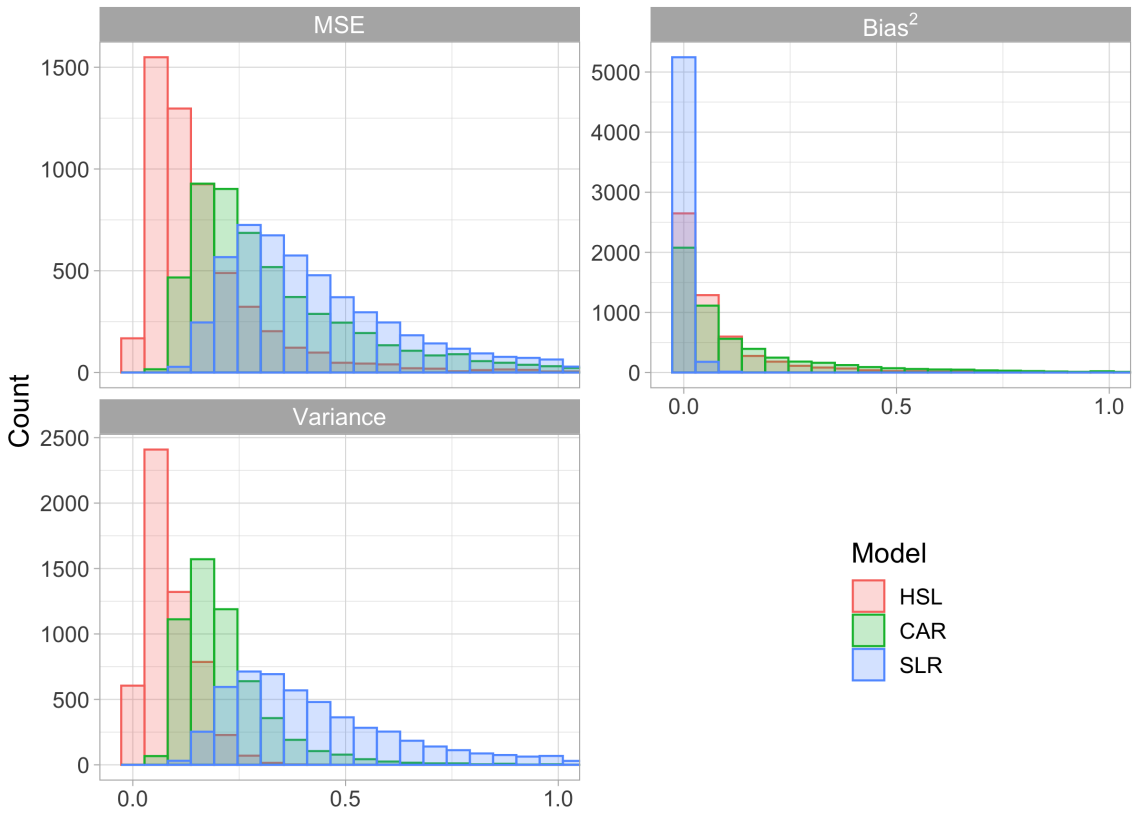

Figure 2.1: Histograms of mean squared error (MSE) (upper left), squared bias (upper right), and estimator variance (lower left) for patient-superpixel slopes comparing simple linear regression (SLR), conditional autoregressive (CAR), and hierarchical spatial longitudinal (HSL) models pooled across all superpixels from the simulation study. Mean squared error is squared bias plus estimator variance; the x-axis values are directly comparable in all three plots. Lower MSE indicates better model performance. Counts of large values omitted due to truncating the x-axis at 1: MSE (HSL: 43; CAR: 241; SLR: 500), squared bias (HSL: 31; CAR: 120; SLR: 0), and estimator variance (HSL: 0; CAR: 8; SLR: 492).

(ratio range: 1.55 to 4.56), RMSE (ratio range: 1.33 to 3.08), and 95% CrILs (ratio range: 1.63 to 3.14) than HSL across superpixels. CAR had larger estimator SD (ratio range: 0.94 to 3.05) and RMSE (ratio range: 1.27 to 2.28) than HSL across superpixels, but similar 95% CrILs (ratio range: 0.73 to 1.69). There were no substantive differences (coverage $-$ 0.95) in 95% CrI coverage probability for HSL and SLR, but noticeable differences in parafoveal and nasal superpixels for CAR (HSL range: -0.033 to 0.024; CAR range: -0.273 to -0.006; SLR range: -0.004 to 0.005). Figure 2.2 displays heat maps of the intercept and slope average RMSE ratios (CAR/HSL and SLR/HSL) across superpixels. In all superpixels, HSL outperforms CAR and SLR with all ratios > 1. Figure 2.3 gives the proportion of significant negative and positive slopes detected by HSL, CAR, and SLR when the true slope is negative. The HSL model detected a higher proportion of significant negative slopes than SLR in 38/49 superpixels, with notably larger differences in the central superpixels. It also detected a lower proportion of significant positive slopes than SLR in all superpixels. In contrast, CAR detected a higher proportion of significant negative slopes in 36/49 superpixels, but also a higher proportion of significant positive slopes in all 49 superpixels than HSL when the true slopes were negative. The proportions of significant negative and positive slopes detected by HSL, CAR, and SLR when the true slope is positive are shown in Supplementary Figure 2.11. The HSL model detected a lower proportion of significant negative slopes in 42/49 superpixels than SLR when the true slopes were positive and a lower proportion of significant positive slopes in 43/49 superpixels than SLR. In addition, HSL detected a lower proportion of significant negative slopes in all 49 superpixels and a lower proportion of significant positive in 39/49 superpixels than CAR. A detailed breakdown of the proportion of significant slopes detected by superpixel is shown in Supplementary Tables 2.5, 2.6, and 2.7.

2.3.2 Analysis of AGPS data

Table 2.2 presents summaries of the grand mean and 10th to 90th percentile range of posterior means and SDs for intercepts and slopes across all 5,419 patient-superpixel combinations for HSL, CAR, and SLR and the differences between the means (CAR $-$ HSL and SLR $-$ HSL) and ratios

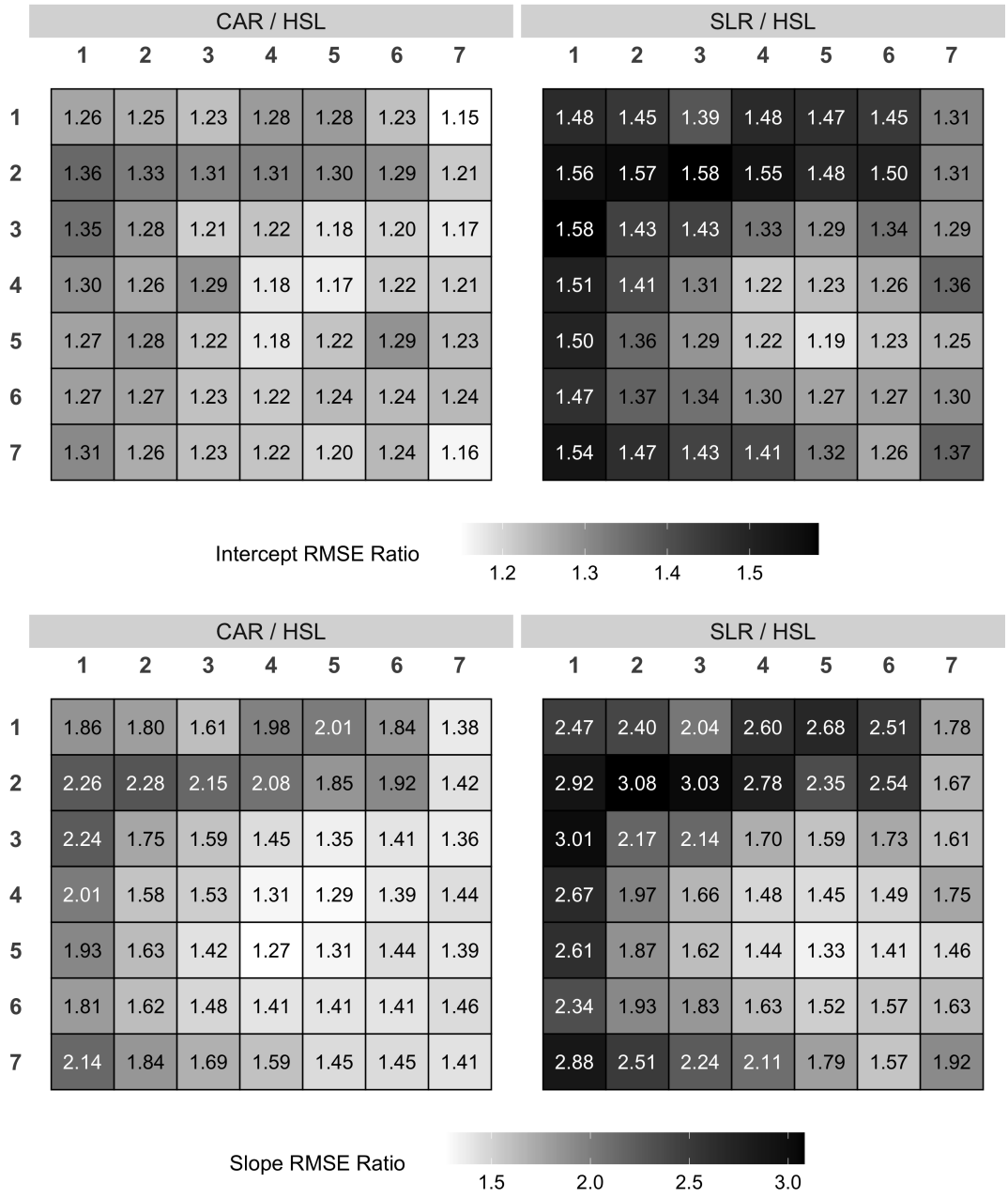

Figure 2.2: Heat map of the average root mean squared error (RMSE) ratio for simple linear regression (SLR) over hierarchical spatial longitudinal (HSL) or conditional autoregressive (CAR) over HSL for intercepts (top) and slopes (bottom) by superpixel location from the simulation study. HSL outperforms SLR and CAR in all superpixels as ratios greater than one favor HSL over the alternative model.

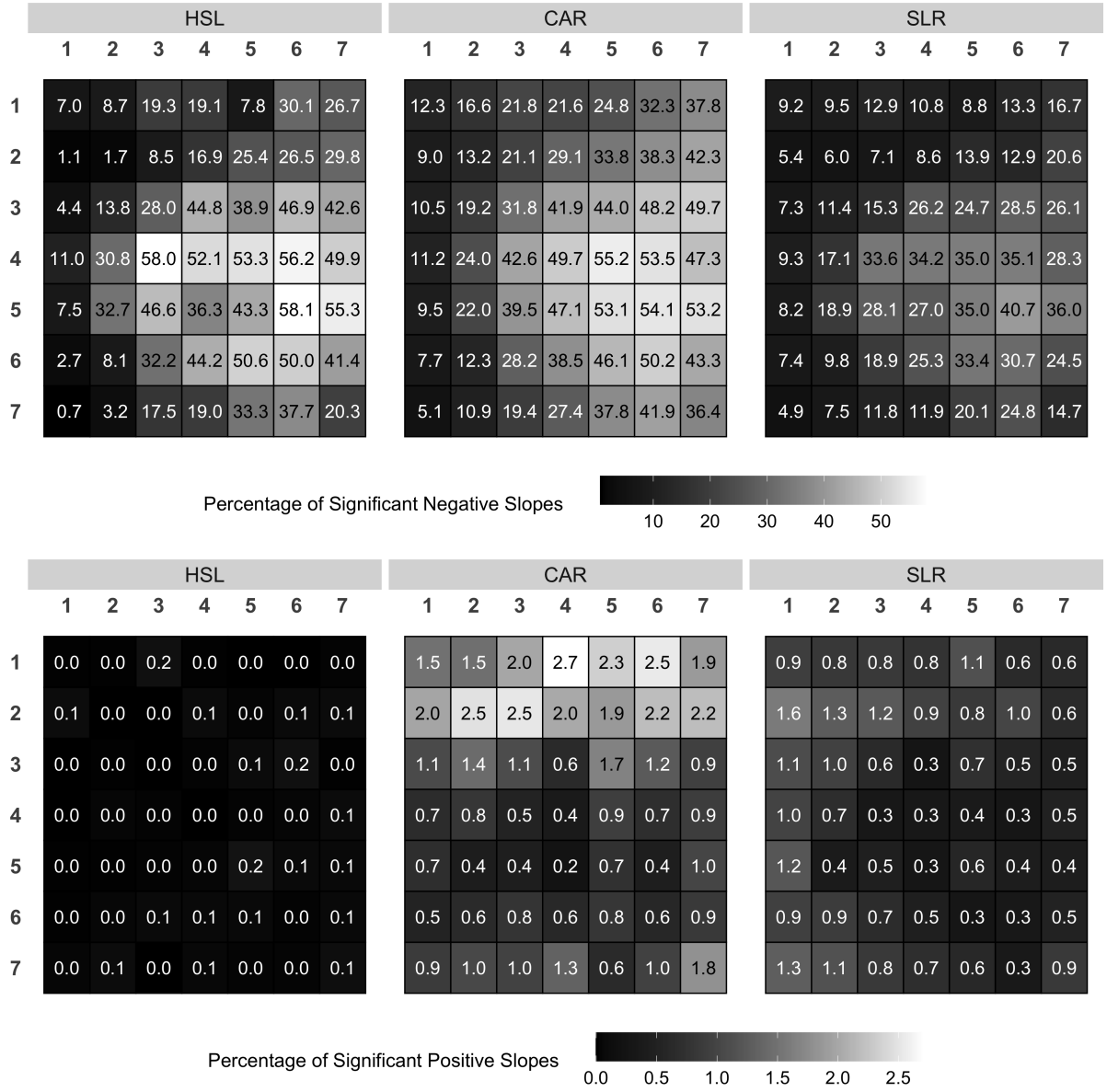

Figure 2.3: Heat map of the percentage of significant negative slopes (top) and significant positive slopes (bottom) detected by simple linear regression (SLR), conditional autoregressive (CAR), and hierarchical spatial longitudinal (HSL) models when the true slope is negative in the simulation study.

of the SDs. Globally, HSL, CAR, and SLR posterior means were similar on average for both the intercepts (CAR − HSL mean difference [10th, 90th percentile]: 0.01 [−1.36, 1.39 μm]; SLR − HSL difference: 0.01 [−1.53, 1.55 μm]) and slopes (CAR − HSL mean difference [10th, 90th percentile]: −0.02 [−0.76, 0.71 μm/year]; SLR − HSL difference: −0.02 [−0.85, 0.82 μm/year]); however, substantial differences existed across patients and superpixels. The striking difference between HSL and SLR was the systematically higher posterior SDs for SLR compared to HSL particularly for the slopes (median 1.69, 10th, 90th percentile = 0.93, 3.28). The mean and 10th and 90th percentiles of posterior means and SDs of slopes by superpixel for HSL, CAR, and SLR are shown in Supplementary Tables 2.8 and 2.9. While the posterior means of the slopes are similar, the ranges are noticeably smaller and the posterior SDs are also uniformly smaller across all superpixels for HSL and CAR. Across 5,419 patient-superpixel curves, the HSL posterior SDs were smaller than those of SLR for 75.2% of intercepts and 87.1% of slopes but larger than those of CAR for 63.4% of intercepts and 54.6% of slopes. Figure 2.4 displays the median and 10th and 90th percentile of posterior SD ratios of CAR/HSL and SLR/HSL. Across all 49 superpixels, HSL has smaller posterior SDs than SLR; HSL has smaller posterior SDs than CAR in temporal superpixels and larger posterior SDs in nasal superpixels. Figure 2.5 shows scatter plots of the posterior means from SLR against HSL. There is noticeable shrinkage towards the population mean in the HSL estimates in the peripheral superior and temporal superpixels. The scatter plots of the posterior means from CAR against HSL are shown in Supplementary Figure 2.12, where there is noticeable shrinkage in the HSL estimates in the temporal regions as well. We provide the scatter plots of posterior means from SLR against HSL and from CAR against HSL on the same axes for all 49 superpixels in Supplementary Figures 2.13 and 2.14, respectively.

The HSL model identified a higher proportion of significant negative slopes compared to SLR and a lower proportion compared to CAR (HSL: 17.6%; CAR: 26.6%; SLR: 15.6%); it detected a lower proportion of significant positive slopes compared to both CAR and SLR (HSL: 1.2%; CAR: 6.9%; SLR: 4.6%). Supplementary Figure 2.15 presents the McNemar's test results comparing the proportion of significant negative slopes between HSL and SLR or HSL and CAR, where a higher

Table 2.2: Summary of the grand mean (10th, 90th percentile) of posterior means and posterior standard deviations (SD) for intercepts and for slopes for all 5,419 patient-superpixel combinations for simple linear regression (SLR), conditional autoregressive (CAR), and hierarchical spatial longitudinal (HSL) models for the data from the patient cohort. The difference in posterior means for intercepts and slopes is defined as SLR minus HSL or CAR minus HSL. The ratio in posterior SDs is defined as SLR over HSL or CAR over HSL and we report the median (10th, 90th percentile) ratio of posterior SDs.

	Mean	10th Percentile	90th Percentile
Intercept Posterior Mean			
HSL	72.95	52.73	102.97
CAR	72.96	52.49	102.92
SLR	72.96	52.29	103.32
Difference (CAR − HSL)	0.01	–1.36	1.39
Difference (SLR − HSL)	0.01	–1.53	1.55
Intercept Posterior SD			
HSL	1.09	0.82	1.40
CAR	1.03	0.72	1.49
SLR	1.61	0.78	2.76
Ratio (CAR/HSL), median	0.92	0.71	1.28
Ratio (SLR/HSL), median	1.35	0.76	2.43
Slope Posterior Mean			
HSL	–0.39	–1.19	0.26
CAR	–0.41	–1.40	0.60
SLR	–0.41	–1.80	0.87
Difference (CAR – HSL)	–0.02	–0.76	0.71
Difference (SLR – HSL)	–0.02	–0.85	0.82
Slope Posterior SD			
HSL	0.40	0.27	0.56
CAR	0.41	0.23	0.59
SLR	0.77	0.35	1.33
Ratio (CAR/HSL), median	0.96	0.63	1.55
Ratio (SLR/HSL), median	1.69	0.93	3.28

proportion of significant negative slopes was identified in 13/49 superpixels for HSL compared to SLR, 4/49 superpixels for SLR compared to HSL, and 31/49 superpixels for CAR compared to HSL. Supplementary Figure 2.16 shows the McNemar's test results for comparing the proportion of significant positive slopes, where a lower proportion of significant positive slopes was identified in 21/49 superpixels for HSL compared to SLR and 36/49 for HSL compared to CAR.

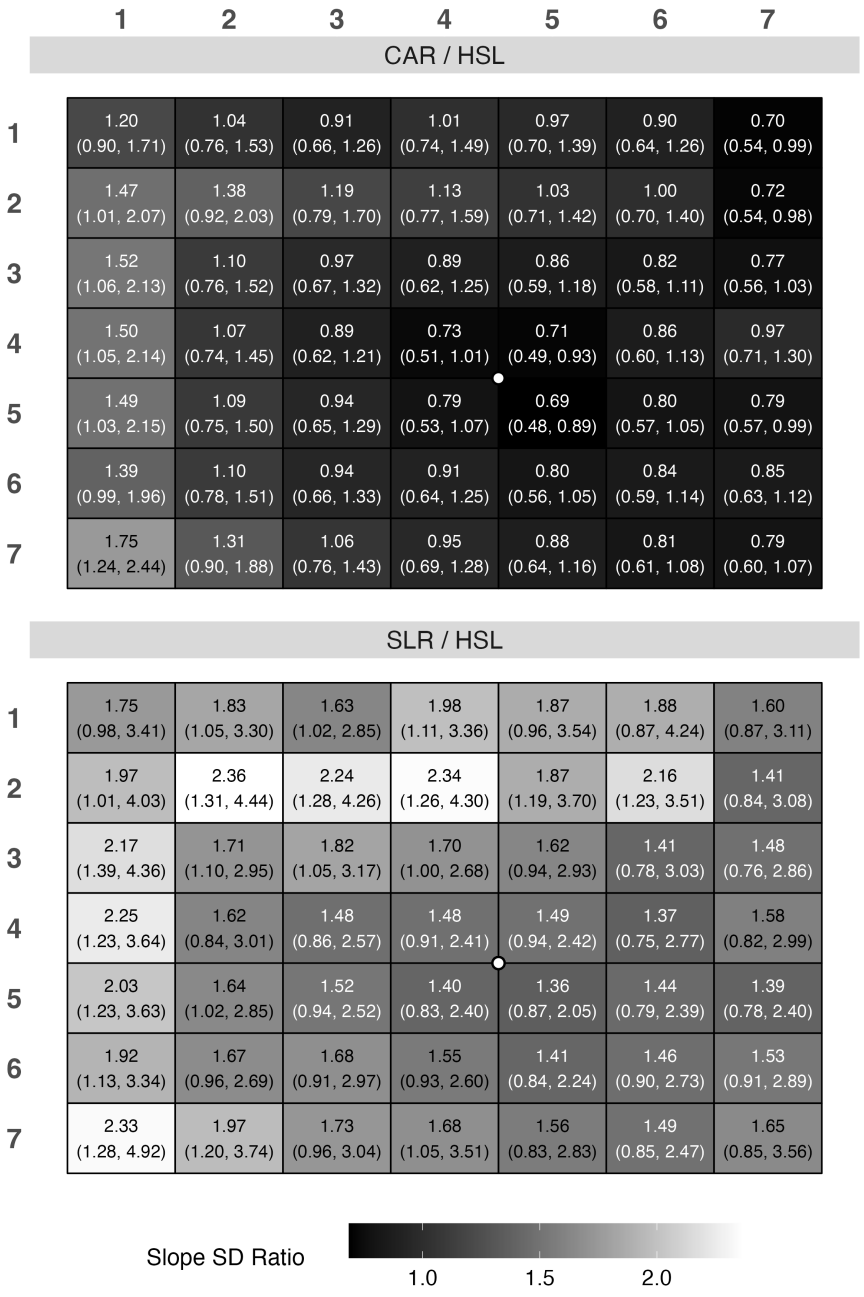

Figure 2.4: Heat map of the median (10th, 90th percentile) slope SD ratio as a function of superpixel location for the patient cohort data. The ratio was defined as either conditional autoregressive (CAR) over hierarchical spatial longitudinal (HSL) or simple linear regression (SLR) over HSL. Ratios greater than 1 indicate better performance for HSL. On average, HSL outperforms SLR at all superpixel locations.

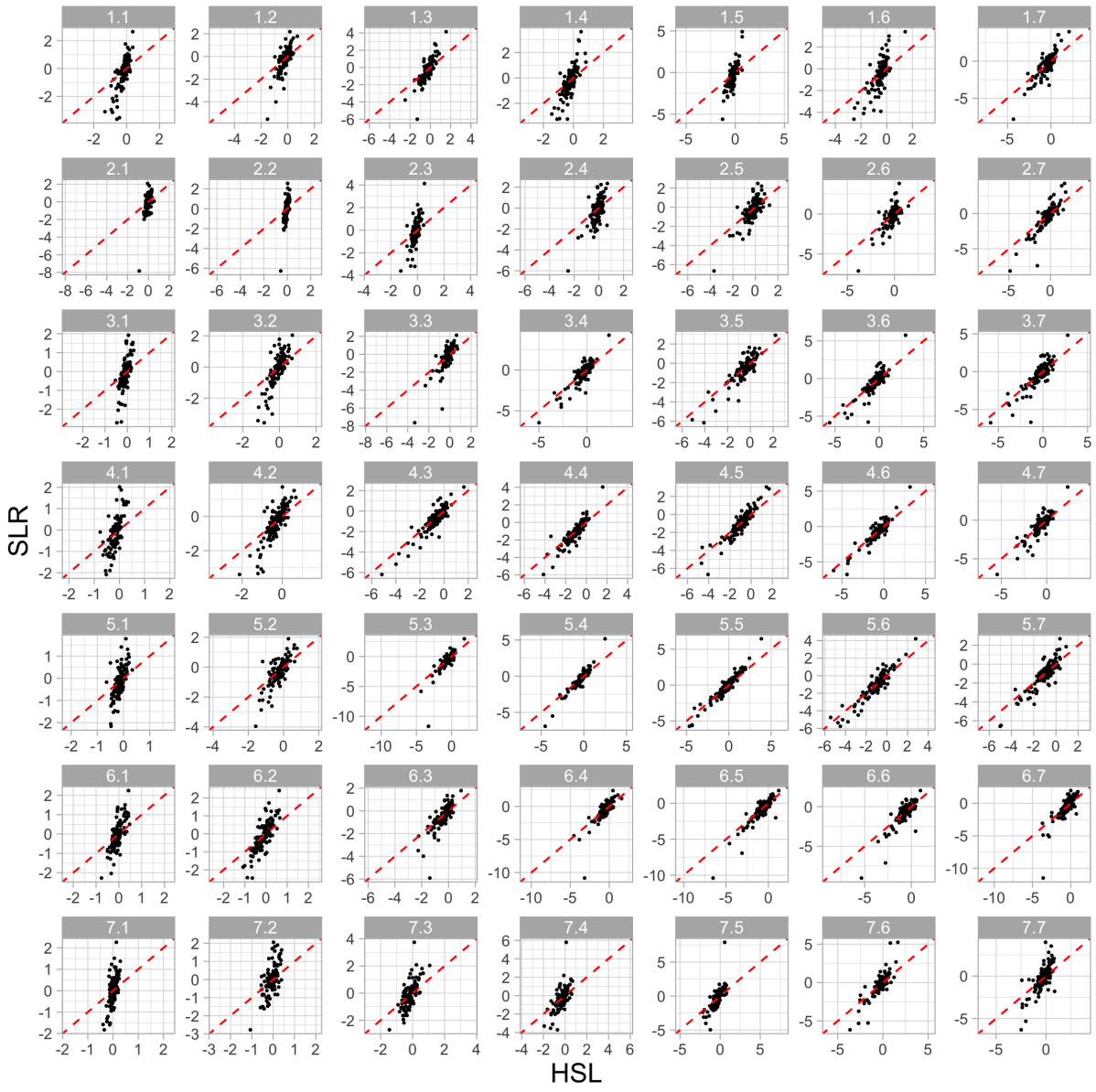

Figure 2.5: Scatter plots of slope posterior means from simple linear regression model (SLR, y-axis) against those from the hierarchical spatial longitudinal model (HSL, x-axis) in each superpixel for the patient cohort data (right eye format). The SLR posterior means are much more variable than the HSL means. There is noticeable shrinkage towards the population mean in the HSL estimates in the peripheral superior and temporal regions. Each plot is square with its own axes with the x- and y- axes having the same range. The red dashed line represents the $x = y$ diagonal.

2.4 Discussion

We examined the ability of SLR, CAR, and HSL to accurately estimate rates of GCC thinning within macular superpixels in a cohort of eyes with central or moderate to advanced glaucoma damage at baseline. Our novel HSL model resulted in lower posterior SD for both intercepts (SLR/HSL median ratio: 1.35) and slopes (SLR/HSL median ratio: 1.69, Table 1) indicating a marked improvement in the certainty in estimated intercepts and rates of change for HSL over SLR. The simulation study showed a significantly higher performance by HSL compared to SLR in terms of detecting actual change; on simulated models, HSL detected 21% of slopes as significantly negative while SLR detected only 13% when the true slopes were negative. A smaller and still significantly better performance was observed with the cohort data. While CAR also reduced the posterior SD, CAR estimates were more biased by superpixel location and offered substantially reduced 95% CrI coverage probability (mean intercept probability: 0.882; slope probability 0.810). Based on the simulation study findings, CAR underreported posterior SDs or over-estimated coverage, implying that in the AGPS data analysis, the more frequent declarations of significance are overly optimistic. Therefore, with the advantage of the reduction in noise and appropriate coverage, HSL could detect changes in GCC more efficiently and earlier with relevant clinical and research implications for earlier detection of glaucoma progression.

Simple linear regression is still frequently used to estimate global or local rates of change of structural or functional measures in the field of glaucoma. We show that SLR has numerous weaknesses in this context. Data from the patient cohort are not used to help draw inferences about individual eyes, sectors or superpixels. Spatial correlations are ignored, and correlation of baseline thickness and slopes are not accounted for. Visit effects cannot be accommodated in SLR. These limitations in SLR lead to substantially larger uncertainty in estimating individual patient-superpixel rates of change. In contrast, our HSL model accommodates all these features of the data and thus reduces uncertainty in estimating rates of change.

Our Bayesian hierarchical spatial longitudinal model addresses these shortcomings in SLR by

modeling the structure of the data with patients nested in a cohort, spatial correlations across the macula, and visit effects. The HSL model provides intercept and slope estimates with substantially higher accuracy by incorporating information from the cohort and other superpixels from the same person and reduces the posterior variances for both intercepts and slopes. We implemented this updated version of the HSL model developed in our research laboratory in this study. The current version allows random variation across superpixels of individual intercepts and slopes from a global patient estimate. Similarly, intercepts and slopes within individual superpixels are allowed to randomly vary from the population intercept and slope. Population level intercepts and slopes vary across the macula and log residual variances are modeled with superpixel and individual patient components. We recently identified spatial distributions of patient-intercepts, patient-slopes, patient-residual variances, and residuals across superpixels through residual analysis of a prior model that did not include spatial effects except for superpixel level parameters (Mohammadzadeh et al., 2022b). The current model in this study improves on the model in Mohammadzadeh et al. (2022b) by including such spatial effects for patient-superpixel intercepts, slopes, and residual variances. Moreover, we now include visit effects, which model correlations in residuals across superpixels for a single patient-visit. These spatial effects are modeled as factors in a factor analysis with one factor per dimension.

Visit effects can be estimated with our HSL model but are impossible to implement in SLR. Visit effects model the correlation in regression residuals across superpixels in a single visit. Presence of visit effects means that all GCC thickness measurements from a single patient-visit tend to be randomly above or below the patient's actual GCC thickness. Estimating visit effects requires multivariate modeling of all superpixel thickness measures in one model. We believe this is the first study to identify visit effects in structural OCT data. Visit effects in visual field data are well known (Phu and Kalloniatis, 2021; Chauhan et al., 2008; Rui et al., 2021), though modeling of visual field visit effects has been rarely done (Bryan et al., 2015).

Our findings indicate a more than twofold reduction in the variance of slope estimates with the HSL model compared to SLR. This significant reduction in posterior variance shows that the HSL

model is much more efficient than SLR in using the available data and it allows HSL to detect significant rates of change earlier as compared to SLR. For example, HSL identified a higher proportion of significant negative slopes (17.6% vs. 15.6%) and lower proportion of significant positive slopes (1.2% vs. 4.6%) as compared to SLR. While these numbers may not seem clinically impressive, the overall superior performance of the HSL is clinically relevant as it not only increases negative hit rates, an indication of higher sensitivity for identifying actual decreasing thicknesses, but it also results in marked reduction of significant positive slopes, an indication of potentially higher specificity as GCC thickness is not expected to increase over time. These findings were confirmed in the simulation study.

There are previous studies that used hierarchical linear models for detection of longitudinal change in visual fields; our study is unique as it addresses detection of longitudinal changes in macular structural measurements (Montesano et al., 2021; Zhu et al., 2014b; Betz-Stablein et al., 2013; Bryan et al., 2015; Henson et al., 2000). Montesano et al. (2021) developed a Bayesian model for visual field data accounting for the within-eye hierarchical structure, data censoring, and the heteroskedastic variance as a function of the mean threshold sensitivity; they found that time to detect progression was shorter for Bayesian models compared to SLR with permutation analysis of point-wise linear regression.

In the paper by Betz-Stablein et al. (2013), a conditional autoregressive prior was used to model spatial correlation across the visual field intercepts and slopes within a single person but data from multiple patients were not included in a single model. Intercepts and slopes were not modeled as correlated, and conditional on neighboring superpixels, distant superpixels were considered independent. Our CAR model, while inspired by Betz-Stablein et al. (2013), is novel; like in Betz-Stablein et al. (2013), our CAR model was fit to data from each patient separately. While the variances in slope estimates were on par with HSL, there was substantially more bias and reduced 95% CrI coverage probability, translating to a higher proportion of significant slopes identified even when the true slopes were of the opposite sign; thus the CAR model had higher error rates. In contrast, our model also has random patient-superpixel intercepts and slopes, but then allows for finding global

patterns of associations in both intercepts and slopes, as we demonstrated in our previous work (Mohammadzadeh et al., 2022b). It also accommodates residual variance varying by both superpixel and patient and allows for visit effects as well.

Based on the simulation study, HSL has substantially lower RMSE than SLR and CAR for both individual patient intercepts and slopes and for superpixel aggregated effects compared to SLR. The SLR model does not have population or superpixel parameters that summarize the population of patients globally or in a superpixel, so any superpixel level inference from SLR is ad hoc; by taking a Bayesian approach with our SLR models, we were able to create an appropriate, if simplistic, inference for SLR results aggregated across patients within a superpixel. For patient-superpixel intercepts and slopes, HSL had lower RMSE in an overwhelming proportion of intercepts (95%) and slopes (96%). Estimates from HSL had thus lower noise as compared to SLR.

Our simulation results demonstrated a small bias in some estimated HSL intercepts and slopes. This phenomenon of shrinkage towards the prior mean is a well-known feature of Bayesian estimators. By borrowing information from the population of patients, the Bayesian models shrink all estimates towards population averages. Shrinkage depends on the play between uncertainty in the estimate without population information and the variability in the population. Unusually high or low estimates without shrinkage are typically due to noise in addition to possibly higher or lower underlying true values; these values are shrunken by greater absolute amounts in hierarchical Bayesian models than estimates nearer to the center of the distribution. Posterior means from a hierarchical model are more stable than classical non-Bayesian estimates such as from SLR. This shrinkage mitigates against erroneous high/low slope estimates and, hence, helps prevent making aggressive therapeutic decisions when relying on uncertain and possibly erroneous estimates, pending additional data. The magnitude of the superpixel-averaged slope bias was at most 10% of the average slope across all patients and superpixels (-0.040 μm/year vs. -0.39 μm/year). Despite this bias, the HSL model was still able to identify a higher proportion of worsening slopes while at the same time minimizing the significant positive slopes, which is desirable within both clinical and research frameworks.

Our model, as with any model, assumes data generated by the model. As we have normality of residuals built into the model, it does not necessarily model well a distribution with long tails. The obvious outliers are quite extreme, and we tried to remove all the obvious outliers. We believe these are due to measure error. Undoubtedly there are still outliers in the data set. A future extension of our methods may allow for diagnosing outliers via the model itself, or we may develop a novel and improved outlier removal algorithm.

Optical coherence tomography data are somewhat noisy. If clinicians use raw data with outliers, they will not be using the best data possible, and the conclusions could be flawed. Undoubtedly OCT image quality will improve with time and outliers will lessen or disappear. In the meantime, our outlier removal algorithm does not require model fitting; clinicians could use it in clinical practice before trying to fit any model including SLR. The implications of our findings go beyond macular structural measures. The proposed framework can be applied to other structural measures such as retinal nerve fiber layer or neuroretinal rim measurements with modifications for differing geometry and scale of measurements given the fact that all the limitations of SLR apply to those structural measures as well. We are also developing a similar hierarchical model for visual field measurements. The proposed framework is a useful starting point for analysis of visual fields; however, modifications will be required given properties of visual fields such as increased variability with worsening thresholds (heteroskedasticity), censoring, and possible loss of information once threshold sensitivity at individual test locations drops to 15-19 dB or below (Montesano et al., 2021; Russell et al., 2012a).

In conclusion, we present a novel Bayesian HSL model that improves estimation accuracy of local GCC rates of change. In a simulation study, SLR and CAR have median MSE ratios over the proposed model of 3.3 and 2.4, respectively, for estimating superpixel-patient slopes; in both the simulation study and in the patient cohort data, HSL identifies a higher proportion of deteriorating superpixels when compared to SLR while minimizing positive detection rates. This efficiency is found by more fully utilizing already available information from measurements on a cohort of glaucoma patients and jointly analyzing measurements on all superpixels. Our findings have important

implications for improved detection of glaucoma deterioration both clinically and in the research setting.

2.5 Supplementary material

2.5.1 Outlier removal algorithm

As previously described (Mohammadzadeh et al., 2021, 2022b), we developed a semi-automated algorithm to identify and subsequently remove outliers which we also explain here. Let y_{ijk} denote a single observation of GCC thickness measured on patient i's jth visit at time t_{ij} in superpixel k. The follow-up time t_{ij} is years since baseline, with the first visit at $t_{ij} = 0$ years. The total number of observations for patient i at superpixel k is n_{ik}. For each patient's data at each superpixel, we calculated consecutive visit absolute differences $|y_{ijk} - y_{i(j-1)k}|$ and consecutive visit absolute centered slopes $|y_{ijk} - y_{i(j-1)k}/(t_{ij} - t_{i(j-1)}) - 0.41|$ for $j = 2, \ldots, n_i$. Slopes were centered around -0.41 μm/year, the mean of all slopes from all pairs of consecutive visits across all patients and superpixels. We flagged absolute centered slopes greater than 24 μm/year that also had consecutive visit absolute differences greater than 5 μm. These values were chosen to identify approximately 0.5% of the observations as possible outliers. The flagged slope identified two consecutive points as candidates for removal. We calculated the sum of the absolute visit differences $\sum_{j=2}^{n_{ik}} |y_{ijk} - y_{i(j-1)k}|$ for each superpixel in each patient and further considered the candidate that caused the largest decrease in the sum of the absolute visit differences. If removing the candidate resulted in a replacement absolute slope that was not one-half or less of the original flagged slope, we did not remove the observation; otherwise, we removed the candidate. If an observation was removed, we applied the same algorithm to the reduced data set to see if another observation from the same superpixel should be removed as well. For each curve, if 2 or more points were identified as outliers, we removed the entire curve. All together this resulted in removing 0.7% of the observations as outliers.

2.5.2 Bayesian hierarchical spatial longitudinal model specification

Let $N(a,b)$ be a normal distribution with mean a and variance b, $IG(a,b)$ be an inverse gamma distribution with mean $b/(a-1)$ (for $a>1$) and variance $b^2/[(a-1)^2(a-2)]$ (for $a>2$), $U(a,b)$ be a uniform distribution over the interval (a,b), $a<b$, with mean $(a+b)/2$ and variance $(b-a)^2/12$, and let $y \sim HN(a)$ be distributed as a half-normal with scale a where the half-normal is the absolute value of a mean zero normal density with variance a.

We fit data from all individuals and superpixels together in a single model. The current version of our Bayesian Hierarchical Spatial Longitudinal (HSL) model includes (a) a macula-wide global intercept and slope, (b) superpixel-specific intercept and slope deviations from the global intercept and slope (superpixel-level random effects), (c) macula-wide patient-specific intercept and slope effects (global patient random effects), (d) patient-superpixel interaction intercept and slope random effects, and (e) macula-wide visit effects. Patient-superpixel log residual variances are modeled as the sum of a superpixel component and a patient-specific component. We mean center the outcome y_{ijk} by the grand mean of all GCC measurements $\bar{y}$ and mean center the time t_{ij} by the mean follow-up time across patients $\bar{t}$.

For superpixel k, α_{0k} is the population average deviation from the grand mean $\bar{y}$ at the mean follow-up time $\bar{t}$; α_{1k} is the population average slope; β_{0ik} is the patient-superpixel interaction deviation at mean follow-up time $\bar{t}$ for the ith patient; β_{1ik} is the patient-superpixel interaction slope for the ith patient; ϕ_{0k} is the spatial effects eigenvector component for the patient-specific deviation at mean follow-up time $\bar{t}$; ϕ_{1k} is the spatial effects eigenvector component for the patient-specific slopes; ϕ_{2k} is the spatial effects eigenvector component for the visit effects; PC_{0i} is the macula-wide patient-specific deviation at mean follow-up time $\bar{t}$; PC_{1i} is the macula-wide patient-specific slope; PC_{2i} is the macula-wide patient-specific log residual variance component; VE_{ij} is the macula-wide visit effect; and the population component of the patient-superpixel log residual variance is $\log \sigma_k^2$.

The full model is

$$
\begin{aligned}
y_{ijk}|\boldsymbol{\theta} &\sim N(\alpha_{0k}+\alpha_{1k}t_{ij}+\beta_{0ik}+\beta_{1ik}t_{ij}+\phi_{0k}\mathrm{PC}_{0i}+\phi_{1k}\mathrm{PC}_{1i}t_{ij}+\phi_{2k}\mathrm{VE}_{ij},\sigma_k^2\exp(\mathrm{PC}_{2i})) \\
\boldsymbol{\theta} &= \{\alpha_{0k},\alpha_{1k},\beta_{0ik},\beta_{1ik},\phi_{0k},\phi_{1k},\phi_{2k},\mathrm{PC}_{0i},\mathrm{PC}_{1i},\mathrm{PC}_{2i},\mathrm{VE}_{ij},\sigma_k^2\} \\
\beta_{0ik}|D_{00k} &\sim N(0,D_{00k}) \\
\beta_{1ik}|\gamma_k,\beta_{0ik},D_{11.0k} &\sim N(\frac{\gamma_k\beta_{0ik}}{100},D_{11.0k})
\end{aligned}
$$

We rescale the regression coefficient γ_k of the random slopes given the random intercept by a factor of 100 for ease of computation and prior specification.

$$
\begin{aligned}
\alpha_{0k}|\sigma_1^2 &\sim N(\mu_1,\sigma_1^2) \\
\log D_{00k}|\mu_2,\eta_{12},\alpha_{0k},\sigma_2^2 &\sim N(\mu_2+\eta_{12}\alpha_{0k},\sigma_2^2) \\
\log D_{11.0k}|\mu_3,\eta_{13},\alpha_{0k},\eta_{23},\log D_{00k},\mu_2,\eta_{12},\sigma_3^2 &\sim N(\mu_3+\eta_{13}\alpha_{0k}+\eta_{23}(\log D_{00k}-\mu_2-\eta_{12}\alpha_{0k}),\sigma_3^2) \\
\alpha_{1k}|\mu_4,\sigma_4^2 &\sim N(\mu_4,\sigma_4^2) \\
\gamma_k|\mu_5,\sigma_5^2 &\sim N(\mu_5,\sigma_5^2) \\
\log\sigma_k^2|\mu_6,\sigma_6^2 &\sim N(\mu_6,\sigma_6^2) \\
\mathrm{PC}_{0i}|\sigma_7^2 &\sim N(0,\sigma_7^2) \\
\mathrm{PC}_{1i}|\lambda,\mathrm{PC}_{0i},\sigma_8^2 &\sim N(\lambda\mathrm{PC}_{0i},\sigma_8^2) \\
\mathrm{PC}_{2i}|\sigma_9^2 &\sim N(0,\sigma_9^2) \\
\mathrm{VE}_{ij}|\sigma_{10}^2 &\sim N(0,\sigma_{10}^2) \\
\phi_{0k}|\sigma_{11}^2 &\sim N(1/7,\sigma_{11}^2) \\
\phi_{1k}|\sigma_{12}^2 &\sim N(1/7,\sigma_{12}^2) \\
\phi_{2k}|\sigma_{13}^2 &\sim N(1/7,\sigma_{13}^2)
\end{aligned}
$$

In our previous work, we conducted principal component (PC) decompositions of the covariance matrices for random intercepts, slopes, and residuals (Mohammadzadeh et al., 2022b). Here, ϕ_{0k}, ϕ_{1k},

and ϕ_{2k} are the spatial effects eigenvectors from the largest PC of each decomposition for random intercepts, slopes, and residuals, respectively, and are all centered at 1/7.

The matrix D_k is a 2×2 variance-covariance matrix of the random intercepts and slopes with elements

$$D_k = \begin{pmatrix} D_{00k} & D_{01k} \\ D_{10k} & D_{11k} \end{pmatrix}$$

and $D_{11.0k} = D_{11k} - D_{10k}D_{00k}^{-1}D_{01k}$ is the variance of the conditional distribution of $\beta_{1ik}|\beta_{0ik}$. We are particularly interested in the correlation

$$\rho_k = \frac{D_{01k}}{(D_{00k}D_{11k})^{1/2}} = \gamma_k\sqrt{\frac{D_{00k}}{D_{11k}}}$$

between the random intercepts and slopes. The grand mean intercept μ_1 is forced equal to zero

$$\mu_1 \equiv 0,$$

first because the data is centered and second because it greatly improves convergence of the Markov chain Monte Carlo (MCMC) algorithms. We add $\bar{y}$ back into our inferences when we report intercept estimates. The centering in time changes the meaning of the intercepts (the average deviation from the grand mean $\bar{y}$ at the mean follow-up time $\bar{t}$. We transform parameters back to time centered at $t_{i1} = 0$ for inferences in the paper. For example, random intercepts β_{0ik}^* are transformed using the following equation: $\beta_{0ik}^* = \beta_{0ik} - \beta_{1ik}\bar{t}$. The priors are

$$\begin{aligned}
\mu_2 &\sim N(5.4161, 0.804719) \\
\mu_3 &\sim N(-0.4462871, 0.804719) \\
\mu_4 &\sim N(-0.8, 0.36) \\
\mu_5 &\sim N(0, 9) \\
\mu_6 &\sim N(1.1, 0.49)
\end{aligned}$$

$$
\begin{aligned}
\sigma_1 &\sim HN(18.7997121) \\
\sigma_2 &\sim HN(0.6266571) \\
\sigma_3 &\sim HN(1.0026513) \\
\sigma_4^2 &\sim IG(2.5, 0.1666667) \\
\sigma_5^2 &\sim IG(2.5, 13.5) \\
\sigma_6 &\sim HN(0.8773199) \\
\sigma_7 &\sim HN(95.07715) \\
\sigma_8 &\sim HN(5.42492) \\
\sigma_9 &\sim HN(1.253314) \\
\sigma_{10} &= u_1 + u_2 \\
u_1 &\sim U(0, 10) \\
u_2 &\sim U(0, 10) \\
\sigma_{11} &\sim HN(0.06266571) \\
\sigma_{12} &\sim HN(0.06266571) \\
\sigma_{13} &\sim HN(0.06266571) \\
\eta_{12} &\sim N(0, 0.25) \\
\eta_{13} &\sim N(0, 0.25) \\
\eta_{23} &\sim N(0, 0.25) \\
\lambda &\sim N(0, 0.25).
\end{aligned}
$$

2.5.3 Bayesian conditional autoregressive model specification

For the conditional autoregressive (CAR) model inspired by Betz-Stablein et al. (2013), we fit data for each individual separately. We assumed GCC thicknesses in neighboring superpixels would be more similar than in more distant superpixels. We defined the neighborhood structure, or adjacency

matrix, such that superpixels sharing an edge are weighted as 1 and nonadjacent superpixels are weighted as 0. For each patient i, we model spatial dependence in intercepts and slopes

$$y_{jk} = \alpha_0 + \alpha_1 t_j + \beta_{0k} + \beta_{1k} t_j + \varepsilon_{jk}$$
$$\varepsilon_{jk} \sim N(0, \sigma^2)$$

where α_0 is the overall eye intercept; α_1 is the overall eye slope; β_{0k} is the superpixel-specific spatial deviation from the overall eye intercept; β_{1k} is the superpixel-specific spatial deviation from the overall eye slope; and σ^2 is the residual variance specific to patient i. The priors are

$$\sigma^2 \sim IG(3,8)$$
$$\alpha_0 \sim N(77,400)$$
$$\alpha_1 \sim N(-0.45, 0.5625)$$
$$\beta_{0k} | \beta_{0-k}, \tau_0 \sim N(\mu_{0k}, \sigma^2_{0k})$$

where

$$\mu_{0k} = \left(\frac{1}{\sum_l w_{kl}} \right) \sum_l w_{kl} \beta_{0l}$$
$$\sigma^2_{0k} = \frac{1}{\tau_0 \sum_l w_{kl}}$$

and β_{0-k} represents all elements of β_0 except β_{0k}. When superpixel k and l are adjacent, that is, they share an edge, $w_{kl} = 1$; otherwise, $w_{kl} = 0$. The parameter τ_0 is the precision for the spatial intercept process. For the slopes we have

$$\beta_{1k} | \beta_{1-k}, \tau_1 \sim N(\mu_{1k}, \sigma^2_{1k})$$

where

$$\mu_{1k} = \left(\frac{1}{\sum_l w_{kl}}\right) \sum_l w_{kl}\beta_{1l}$$

$$\sigma^2_{1k} = \frac{1}{\tau_1 \sum_l w_{kl}}$$

$$\tau_0 \sim IG(0.001, 0.001)$$

$$\tau_1 \sim IG(0.001, 0.001).$$

2.5.4 Supplementary tables

Table 2.3: Clinical and demographic characteristics of study sample. SD = standard deviation; MD = mean deviation; IQR = interquartile range; GCC = ganglion cell complex.

Age (years)	
Mean (SD)	66.9 (8.5)
Range	39.7 to 81.2
Gender (%)	
Female	70 (63.1%)
Male	40 (36.0%)
Not reported	1 (0.9%)
Ethnicity (%)	
Caucasian	59 (53.2%)
Asian	24 (21.6%)
African American	15 (13.5%)
Hispanic	13 (11.7%)
Baseline 10-2 MD (dB)	
Median (IQR)	-7.6 (-12.0 to -4.1)
Mean (SD)	-8.9 (5.9)
Range	-25.1 to -0.4
Baseline 24-2 MD (dB)	
Median (IQR)	-6.7 (-12.3 to -4.3)
Mean (SD)	-8.7 (6.1)
Range	-26.4 to -0.3
Follow up (years)	
Mean (SD)	3.59 (0.44)
Range	1.94 to 4.20
Visits per Subject	
Mean (SD)	7.3 (1.1)
Range	4 to 10
Baseline GCC (μm)	
Mean (SD)	73.1 (20.1)
Range	37 to 154

Table 2.4: Summary of the grand mean (10th, 90th percentile) of posterior means, posterior standard deviations (SD)s, estimator SDs, 95% Credible Interval Lengths (CrIL)s, root mean squared error (RMSE), and 95% Credible Interval (CrI) coverage probability for intercepts and for slopes for all 5,439 patient-superpixel combinations for simple linear regression (SLR), conditional autoregressive (CAR), and hierarchical spatial longitudinal (HSL) models for the simulation study. The difference in posterior means for intercepts and slopes is defined as CAR or SLR minus HSL. All ratios are defined as CAR or SLR over HSL and we report the median (10th, 90th percentile) ratios. One patient had 4 repeated measures and infinite posterior variance and was omitted from the posterior SD calculations, which represents a slight advantage on the scale for SLR over HSL.

	Intercepts			Slopes		
	Mean	10th Percentile	90th Percentile	Mean	10th Percentile	90th Percentile
			Posterior Mean			
HSL	72.91	51.36	100.57	–0.39	–1.17	0.19
CAR	72.89	51.40	100.34	–0.38	–1.01	0.14
SLR	72.89	51.23	100.78	–0.38	–1.36	0.48
Difference (CAR – HSL)	–0.02	–0.56	0.48	0.01	–0.26	0.30
Difference (SLR – HSL)	–0.02	–0.64	0.61	0.01	–0.33	0.36
			Posterior SD			
HSL	1.07	0.82	1.37	0.39	0.27	0.54
CAR	1.08	0.91	1.27	0.40	0.31	0.49
SLR	1.80	1.30	2.43	0.86	0.56	1.27
Ratio (CAR/HSL), median	1.01	0.83	1.26	1.02	0.75	1.47
Ratio (SLR/HSL), median	1.63	1.41	1.99	2.04	1.60	3.02
			Estimator SD			
HSL	0.95	0.71	1.24	0.28	0.16	0.41
CAR	1.13	0.88	1.42	0.45	0.34	0.59
SLR	1.44	1.08	1.88	0.68	0.47	0.98
Ratio (CAR/HSL), median	1.19	1.02	1.39	1.59	1.08	2.76
Ratio (SLR/HSL), median	1.50	1.30	1.75	2.32	1.60	4.15
			95% CrIL			
HSL	4.21	3.23	5.38	1.55	1.08	2.13
CAR	4.24	3.59	4.97	1.59	1.23	1.94
SLR	7.17	5.21	9.66	3.44	2.24	5.10

Table 2.4: (continued)

	Intercepts			Slopes		
	Mean	10th Percentile	90th Percentile	Mean	10th Percentile	90th Percentile
Ratio (CAR/HSL), median	1.01	0.84	1.26	1.03	0.76	1.48
Ratio (SLR/HSL), median	1.66	1.44	2.00	2.09	1.63	3.06
			RMSE			
HSL	1.07	0.76	1.43	0.38	0.20	0.58
CAR	1.31	0.95	1.76	0.58	0.37	0.84
SLR	1.44	1.08	1.89	0.69	0.47	0.98
Ratio (CAR/HSL), median	1.23	1.02	1.48	1.54	1.08	2.36
Ratio (SLR/HSL), median	1.38	1.11	1.66	1.80	1.18	3.18
			95% CrI Coverage Probability			
HSL	0.949	0.90	0.99	0.947	0.87	1.00
CAR	0.882	0.77	0.97	0.810	0.60	0.95
SLR	0.951	0.92	0.98	0.951	0.92	0.98
Difference (CAR – HSL)	–0.067	–0.16	0.01	–0.137	–0.31	–0.01
Difference (SLR – HSL)	0.002	–0.04	0.05	0.004	–0.06	0.08

Table 2.5: The proportion of significant negative and positive slopes detected by simple linear regression (SLR), conditional autoregressive (CAR), and hierarchical spatial longitudinal (HSL) models by superpixel (SP) location from the simulation study. The proportions are given first for all patient-superpixel slopes, second when the true slope is negative, and third when the true slope is positive. The first row is the global summary across all superpixels.

	All Patient-Superpixel Slopes					
	Significantly Negative			Significantly Positive		
SP	HSL	CAR	SLR	HSL	CAR	SLR
All	20.7%	24.4%	13.8%	1.8%	3.5%	3.3%
1.1	4.9%	8.9%	6.5%	0.2%	3.1%	3.1%
1.2	5.5%	12.1%	6.4%	0.4%	3.9%	3.4%
1.3	13.8%	17.0%	9.4%	1.5%	4.5%	3.8%
1.4	15.4%	19.0%	8.7%	0.0%	3.4%	1.8%
1.5	5.0%	18.6%	6.0%	0.2%	4.3%	3.0%
1.6	24.5%	28.5%	10.9%	0.1%	3.7%	1.5%
1.7	19.6%	29.0%	12.4%	1.3%	6.1%	2.9%
2.1	0.6%	6.5%	3.2%	0.7%	3.3%	4.0%
2.2	1.0%	10.3%	3.8%	0.0%	3.9%	2.7%
2.3	5.9%	16.8%	5.2%	0.1%	4.7%	2.9%
2.4	10.3%	20.2%	5.4%	1.8%	5.3%	3.3%
2.5	15.5%	23.5%	8.8%	3.6%	6.2%	3.5%
2.6	15.7%	27.2%	8.1%	2.3%	6.0%	3.0%
2.7	18.4%	28.4%	12.9%	4.0%	7.2%	4.8%
3.1	2.7%	8.0%	4.9%	0.1%	2.1%	2.3%
3.2	8.2%	13.5%	7.0%	1.0%	3.3%	3.9%
3.3	18.3%	23.2%	10.2%	0.4%	3.5%	2.9%
3.4	31.7%	30.9%	18.7%	4.9%	4.7%	4.4%
3.5	27.8%	32.7%	17.7%	3.9%	5.2%	4.1%
3.6	35.1%	37.4%	21.6%	4.7%	6.2%	4.2%
3.7	27.7%	34.1%	17.1%	5.5%	7.2%	5.2%
4.1	8.6%	9.5%	7.4%	0.0%	1.3%	2.0%
4.2	22.3%	18.9%	12.7%	0.5%	1.9%	2.6%
4.3	46.6%	35.0%	27.2%	1.9%	2.6%	2.4%

Table 2.5: (continued)

	All Patient-Superpixel Slopes					
	Significantly Negative			Significantly Positive		
SP	HSL	CAR	SLR	HSL	CAR	SLR
4.4	42.8%	41.7%	28.1%	0.5%	2.1%	2.1%
4.5	43.0%	45.8%	28.3%	0.8%	2.3%	1.7%
4.6	44.3%	43.3%	27.6%	3.0%	3.7%	3.2%
4.7	39.7%	38.5%	22.7%	1.7%	3.4%	2.3%
5.1	5.5%	7.9%	6.2%	0.1%	1.0%	2.3%
5.2	21.6%	16.5%	12.9%	0.9%	1.4%	2.5%
5.3	34.2%	30.1%	20.7%	3.3%	2.1%	3.7%
5.4	24.6%	33.5%	18.4%	3.8%	2.4%	5.0%
5.5	28.3%	37.9%	23.0%	5.0%	3.4%	5.8%
5.6	46.7%	44.5%	32.8%	2.7%	2.9%	3.0%
5.7	42.5%	42.7%	27.7%	2.1%	3.6%	2.8%
6.1	1.5%	5.3%	4.5%	0.5%	1.6%	3.6%
6.2	4.9%	8.9%	6.2%	1.7%	1.8%	4.7%
6.3	22.2%	21.0%	13.2%	0.8%	1.6%	2.6%
6.4	32.4%	29.4%	18.7%	2.7%	2.2%	3.5%
6.5	40.7%	38.7%	27.0%	2.6%	2.9%	3.0%
6.6	37.5%	40.0%	23.3%	2.4%	2.8%	2.7%
6.7	30.3%	33.7%	18.1%	3.4%	4.1%	4.0%
7.1	0.4%	3.5%	3.0%	0.4%	1.8%	3.5%
7.2	1.6%	6.9%	4.2%	0.6%	2.2%	3.5%
7.3	10.8%	13.7%	7.6%	1.3%	2.5%	3.5%
7.4	12.6%	20.2%	8.1%	2.2%	3.7%	4.1%
7.5	22.6%	27.6%	13.8%	1.8%	3.7%	4.1%
7.6	24.3%	29.2%	16.1%	4.1%	4.0%	4.8%
7.7	12.8%	25.9%	9.5%	2.1%	5.8%	4.3%

Table 2.6: The proportion of significant negative and positive slopes detected by simple linear regression (SLR), conditional autoregressive (CAR), and hierarchical spatial longitudinal (HSL) models by superpixel (SP) location from the simulation study. The proportions are given first for all patient-superpixel slopes, second when the true slope is negative, and third when the true slope is positive. The first row is the global summary across all superpixels.

	When True Slope is Negative					
	Significantly Negative			Significantly Positive		
SP	HSL	CAR	SLR	HSL	CAR	SLR
All	30.3%	33.2%	19.8%	0.0%	1.2%	0.7%
1.1	7.0%	12.3%	9.2%	0.0%	1.5%	0.9%
1.2	8.7%	16.6%	9.5%	0.0%	1.5%	0.8%
1.3	19.3%	21.8%	12.9%	0.2%	2.0%	0.8%
1.4	19.1%	21.6%	10.8%	0.0%	2.7%	0.8%
1.5	7.8%	24.8%	8.8%	0.0%	2.3%	1.1%
1.6	30.1%	32.3%	13.3%	0.0%	2.5%	0.6%
1.7	26.7%	37.8%	16.7%	0.0%	1.9%	0.6%
2.1	1.1%	9.0%	5.4%	0.1%	2.0%	1.6%
2.2	1.7%	13.2%	6.0%	0.0%	2.5%	1.3%
2.3	8.5%	21.1%	7.1%	0.0%	2.5%	1.2%
2.4	16.9%	29.1%	8.6%	0.1%	2.0%	0.9%
2.5	25.4%	33.8%	13.9%	0.0%	1.9%	0.8%
2.6	26.5%	38.3%	12.9%	0.1%	2.2%	1.0%
2.7	29.8%	42.3%	20.6%	0.1%	2.2%	0.6%
3.1	4.4%	10.5%	7.3%	0.0%	1.1%	1.1%
3.2	13.8%	19.2%	11.4%	0.0%	1.4%	1.0%
3.3	28.0%	31.8%	15.3%	0.0%	1.1%	0.6%
3.4	44.8%	41.9%	26.2%	0.0%	0.6%	0.3%
3.5	38.9%	44.0%	24.7%	0.1%	1.7%	0.7%
3.6	46.9%	48.2%	28.5%	0.2%	1.2%	0.5%
3.7	42.6%	49.7%	26.1%	0.0%	0.9%	0.5%
4.1	11.0%	11.2%	9.3%	0.0%	0.7%	1.0%
4.2	30.8%	24.0%	17.1%	0.0%	0.8%	0.7%
4.3	58.0%	42.6%	33.6%	0.0%	0.5%	0.3%

Table 2.6: (continued)

	When True Slope is Negative					
	Significantly Negative			Significantly Positive		
SP	HSL	CAR	SLR	HSL	CAR	SLR
4.4	52.1%	49.7%	34.2%	0.0%	0.4%	0.3%
4.5	53.3%	55.2%	35.0%	0.0%	0.9%	0.4%
4.6	56.2%	53.5%	35.1%	0.0%	0.7%	0.3%
4.7	49.9%	47.3%	28.3%	0.1%	0.9%	0.5%
5.1	7.5%	9.5%	8.2%	0.0%	0.7%	1.2%
5.2	32.7%	22.0%	18.9%	0.0%	0.4%	0.4%
5.3	46.6%	39.5%	28.1%	0.0%	0.4%	0.5%
5.4	36.3%	47.1%	27.0%	0.0%	0.2%	0.3%
5.5	43.3%	53.1%	35.0%	0.2%	0.7%	0.6%
5.6	58.1%	54.1%	40.7%	0.1%	0.4%	0.4%
5.7	55.3%	53.2%	36.0%	0.1%	1.0%	0.4%
6.1	2.7%	7.7%	7.4%	0.0%	0.5%	0.9%
6.2	8.1%	12.3%	9.8%	0.0%	0.6%	0.9%
6.3	32.2%	28.2%	18.9%	0.1%	0.8%	0.7%
6.4	44.2%	38.5%	25.3%	0.1%	0.6%	0.5%
6.5	50.6%	46.1%	33.4%	0.1%	0.8%	0.3%
6.6	50.0%	50.2%	30.7%	0.0%	0.6%	0.3%
6.7	41.4%	43.3%	24.5%	0.1%	0.9%	0.5%
7.1	0.7%	5.1%	4.9%	0.0%	0.9%	1.3%
7.2	3.2%	10.9%	7.5%	0.1%	1.0%	1.1%
7.3	17.5%	19.4%	11.8%	0.0%	1.0%	0.8%
7.4	19.0%	27.4%	11.9%	0.1%	1.3%	0.7%
7.5	33.3%	37.8%	20.1%	0.0%	0.6%	0.6%
7.6	37.7%	41.9%	24.8%	0.0%	1.0%	0.3%
7.7	20.3%	36.4%	14.7%	0.1%	1.8%	0.9%

Table 2.7: The proportion of significant negative and positive slopes detected by simple linear regression (SLR), conditional autoregressive (CAR), and hierarchical spatial longitudinal (HSL) models by superpixel (SP) location from the simulation study. The proportions are given first for all patient-superpixel slopes, second when the true slope is negative, and third when the true slope is positive. The first row is the global summary across all superpixels.

	When True Slope is Positive					
	Significantly Negative			Significantly Positive		
SP	HSL	CAR	SLR	HSL	CAR	SLR
All	0.4%	5.9%	1.0%	5.6%	8.5%	9.0%
1.1	0.3%	2.1%	0.9%	0.7%	6.4%	7.8%
1.2	0.2%	4.6%	1.1%	1.0%	8.0%	7.8%
1.3	0.3%	5.2%	0.9%	4.9%	10.9%	11.1%
1.4	2.0%	9.6%	1.1%	0.2%	6.0%	5.5%
1.5	0.1%	8.2%	1.3%	0.5%	7.7%	6.3%
1.6	1.5%	12.8%	1.5%	0.4%	8.3%	5.0%
1.7	0.4%	5.2%	0.8%	4.6%	17.4%	9.3%
2.1	0.0%	4.0%	1.1%	1.3%	4.6%	6.3%
2.2	0.1%	6.8%	1.3%	0.0%	5.6%	4.4%
2.3	0.1%	7.5%	1.0%	0.4%	9.3%	6.6%
2.4	0.5%	7.1%	0.8%	4.3%	10.1%	6.8%
2.5	0.4%	7.8%	1.0%	9.1%	12.7%	7.6%
2.6	0.5%	11.4%	1.3%	5.4%	11.3%	5.8%
2.7	0.3%	6.5%	0.8%	10.1%	15.2%	11.3%
3.1	0.3%	4.4%	1.5%	0.2%	3.5%	4.2%
3.2	0.3%	5.5%	0.7%	2.3%	5.9%	8.0%
3.3	1.2%	7.9%	1.2%	1.2%	7.9%	6.8%
3.4	0.7%	5.0%	0.9%	16.3%	14.4%	14.1%
3.5	0.3%	5.0%	0.6%	13.3%	14.1%	12.4%
3.6	0.4%	5.5%	1.0%	18.2%	20.7%	15.0%
3.7	0.3%	5.1%	0.3%	15.6%	18.7%	14.0%
4.1	0.7%	3.6%	1.1%	0.2%	3.6%	5.5%
4.2	0.5%	5.7%	1.4%	1.5%	4.7%	7.7%
4.3	0.5%	4.3%	1.1%	9.3%	11.0%	10.8%

Table 2.7: (continued)

	When True Slope is Positive					
	Significantly Negative			Significantly Positive		
SP	HSL	CAR	SLR	HSL	CAR	SLR
4.4	0.4%	5.5%	0.6%	2.8%	10.1%	10.3%
4.5	1.2%	7.6%	1.0%	4.2%	8.0%	7.2%
4.6	1.2%	6.2%	0.5%	14.0%	14.8%	13.6%
4.7	0.7%	4.6%	1.1%	8.1%	12.9%	9.2%
5.1	0.5%	3.8%	1.3%	0.2%	1.9%	4.9%
5.2	0.4%	5.8%	1.4%	2.5%	3.3%	6.7%
5.3	0.5%	4.6%	0.7%	12.3%	6.8%	12.4%
5.4	0.1%	5.3%	0.5%	11.8%	6.8%	14.7%
5.5	0.5%	9.7%	0.7%	13.9%	8.2%	15.5%
5.6	1.0%	5.7%	0.9%	13.3%	13.0%	13.6%
5.7	0.8%	8.5%	0.7%	8.7%	11.9%	10.5%
6.1	0.0%	2.4%	0.9%	1.0%	2.9%	6.8%
6.2	0.3%	4.0%	0.9%	4.1%	3.4%	10.2%
6.3	0.6%	5.3%	0.9%	2.5%	3.3%	6.8%
6.4	0.5%	5.0%	0.8%	9.9%	6.7%	11.5%
6.5	0.7%	8.8%	0.9%	13.0%	11.4%	14.0%
6.6	0.5%	9.8%	1.2%	9.6%	9.4%	9.9%
6.7	0.4%	7.6%	0.7%	12.6%	12.8%	13.5%
7.1	0.0%	2.1%	1.3%	0.8%	2.7%	5.6%
7.2	0.1%	2.9%	1.1%	1.1%	3.4%	5.8%
7.3	0.4%	4.6%	1.1%	3.4%	5.0%	7.6%
7.4	0.2%	6.3%	0.9%	6.3%	8.2%	10.7%
7.5	0.3%	6.4%	0.7%	5.4%	10.1%	11.4%
7.6	0.3%	6.6%	0.5%	11.2%	9.5%	12.7%
7.7	0.4%	8.6%	1.0%	5.3%	12.5%	9.9%

Table 2.8: The mean and percentile range (10th, 90th) of posterior means and standard deviations (SD) of slopes by superpixel (SP) location from the analysis of data from the patient cohort for simple linear regression (SLR), conditional autoregressive (CAR), and hierarchical spatial longitudinal (HSL) models. The first row is the global summary across all superpixels.

	Posterior Mean					
	HSL		CAR		SLR	
SP	Mean	Range	Mean	Range	Mean	Range
All	–0.39	(–1.19, 0.26)	–0.41	(–1.40, 0.60)	–0.41	(–1.80, 0.87)
1.1	–0.19	(–0.59, 0.14)	–0.23	(–1.34, 0.66)	–0.26	(–1.45, 0.79)
1.2	–0.18	(–0.66, 0.20)	–0.26	(–1.38, 0.56)	–0.21	(–1.52, 0.83)
1.3	–0.28	(–0.91, 0.29)	–0.29	(–1.39, 0.85)	–0.31	(–1.58, 1.10)
1.4	–0.25	(–0.69, 0.17)	–0.32	(–1.32, 0.67)	–0.29	(–1.54, 0.91)
1.5	–0.20	(–0.68, 0.23)	–0.35	(–1.49, 0.71)	–0.28	(–1.71, 0.96)
1.6	–0.52	(–1.26, –0.04)	–0.47	(–1.97, 0.60)	–0.55	(–2.22, 0.99)
1.7	–0.48	(–1.56, 0.45)	–0.41	(–1.77, 0.98)	–0.45	(–2.31, 1.05)
2.1	–0.05	(–0.27, 0.23)	–0.16	(–1.03, 0.65)	–0.11	(–1.21, 0.81)
2.2	–0.06	(–0.20, 0.12)	–0.17	(–1.00, 0.72)	–0.08	(–1.13, 1.03)
2.3	–0.15	(–0.40, 0.12)	–0.23	(–1.11, 0.68)	–0.17	(–1.30, 0.97)
2.4	–0.12	(–0.54, 0.29)	–0.28	(–1.29, 0.75)	–0.19	(–1.61, 1.08)
2.5	–0.24	(–0.90, 0.42)	–0.32	(–1.22, 0.66)	–0.27	(–1.57, 0.95)
2.6	–0.22	(–1.06, 0.37)	–0.35	(–1.32, 0.60)	–0.22	(–1.71, 0.87)
2.7	–0.38	(–1.57, 0.63)	–0.41	(–1.75, 0.67)	–0.42	(–2.54, 1.17)
3.1	–0.10	(–0.31, 0.13)	–0.18	(–0.88, 0.61)	–0.13	(–0.81, 0.89)
3.2	–0.12	(–0.51, 0.27)	–0.20	(–1.12, 0.63)	–0.16	(–1.44, 0.96)
3.3	–0.28	(–0.74, 0.24)	–0.35	(–1.27, 0.52)	–0.35	(–1.48, 0.86)
3.4	–0.47	(–1.14, 0.33)	–0.43	(–1.48, 0.58)	–0.47	(–2.16, 0.91)
3.5	–0.53	(–1.60, 0.34)	–0.48	(–1.41, 0.65)	–0.56	(–1.84, 0.91)
3.6	–0.53	(–1.49, 0.41)	–0.50	(–1.48, 0.64)	–0.54	(–1.99, 0.78)
3.7	–0.41	(–1.83, 0.64)	–0.50	(–1.98, 0.67)	–0.41	(–2.20, 1.09)
4.1	–0.16	(–0.43, 0.06)	–0.20	(–0.94, 0.49)	–0.16	(–1.07, 0.65)
4.2	–0.32	(–0.92, 0.15)	–0.32	(–1.23, 0.46)	–0.36	(–1.49, 0.76)
4.3	–0.75	(–1.75, 0.11)	–0.58	(–1.52, 0.42)	–0.79	(–2.27, 0.48)
4.4	–1.02	(–2.11, –0.17)	–0.77	(–1.98, 0.17)	–1.06	(–2.57, 0.28)

Table 2.8: (continued)

	Posterior Mean					
	HSL		CAR		SLR	
SP	Mean	Range	Mean	Range	Mean	Range
4.5	–0.94	(–1.94, –0.06)	–0.74	(–1.79, 0.43)	–0.94	(–2.67, 0.42)
4.6	–0.80	(–1.77, 0.24)	–0.67	(–1.98, 0.56)	–0.80	(–2.09, 0.43)
4.7	–0.75	(–1.80, 0.08)	–0.69	(–2.24, 0.41)	–0.77	(–2.28, 0.32)
5.1	–0.14	(–0.38, 0.11)	–0.17	(–0.86, 0.51)	–0.13	(–0.91, 0.65)
5.2	–0.27	(–0.79, 0.21)	–0.31	(–1.02, 0.46)	–0.29	(–1.25, 0.55)
5.3	–0.59	(–1.63, 0.17)	–0.55	(–1.30, 0.41)	–0.67	(–1.93, 0.43)
5.4	–0.34	(–1.20, 0.55)	–0.56	(–1.41, 0.36)	–0.33	(–1.45, 0.75)
5.5	–0.36	(–1.70, 0.99)	–0.58	(–1.80, 0.61)	–0.30	(–1.94, 1.66)
5.6	–1.08	(–2.76, 0.07)	–0.81	(–2.59, 0.53)	–1.07	(–3.30, 0.67)
5.7	–1.02	(–2.39, 0.01)	–0.90	(–2.51, 0.23)	–1.09	(–2.97, 0.40)
6.1	–0.05	(–0.28, 0.26)	–0.09	(–0.82, 0.70)	–0.04	(–0.89, 1.03)
6.2	–0.12	(–0.60, 0.27)	–0.17	(–0.96, 0.48)	–0.12	(–1.05, 0.86)
6.3	–0.39	(–1.07, 0.18)	–0.39	(–1.30, 0.46)	–0.43	(–1.55, 0.61)
6.4	–0.60	(–1.36, 0.28)	–0.54	(–1.37, 0.46)	–0.66	(–1.71, 0.81)
6.5	–0.79	(–1.97, 0.20)	–0.69	(–1.62, 0.41)	–0.89	(–2.40, 0.39)
6.6	–0.74	(–1.65, 0.23)	–0.74	(–1.87, 0.28)	–0.84	(–2.46, 0.47)
6.7	–0.52	(–1.51, 0.40)	–0.68	(–1.56, 0.45)	–0.59	(–2.09, 1.03)
7.1	0.01	(–0.13, 0.16)	–0.05	(–0.81, 0.57)	0.03	(–0.80, 0.84)
7.2	–0.07	(–0.42, 0.21)	–0.09	(–0.97, 0.67)	0.01	(–0.91, 1.16)
7.3	–0.16	(–0.60, 0.24)	–0.20	(–1.09, 0.66)	–0.12	(–1.17, 1.13)
7.4	–0.33	(–0.93, 0.22)	–0.34	(–1.32, 0.71)	–0.32	(–1.68, 0.85)
7.5	–0.33	(–1.08, 0.41)	–0.41	(–1.59, 0.69)	–0.34	(–1.87, 1.01)
7.6	–0.33	(–1.53, 0.54)	–0.44	(–1.41, 0.59)	–0.30	(–1.79, 1.05)
7.7	–0.34	(–1.12, 0.41)	–0.38	(–1.39, 1.20)	–0.27	(–1.80, 1.58)

Table 2.9: The mean and percentile range (10th, 90th) of posterior means and standard deviations (SD) of slopes by superpixel (SP) location from the analysis of data from the patient cohort for simple linear regression (SLR), conditional autoregressive (CAR), and hierarchical spatial longitudinal (HSL) models. The first row is the global summary across all superpixels.

	Posterior SD					
	HSL		CAR		SLR	
SP	Mean	Range	Mean	Range	Mean	Range
All	0.40	(0.27, 0.56)	0.40	(0.23, 0.59)	0.77	(0.35, 1.29)
1.1	0.36	(0.31, 0.42)	0.46	(0.28, 0.68)	0.70	(0.30, 1.14)
1.2	0.37	(0.32, 0.43)	0.42	(0.25, 0.60)	0.79	(0.36, 1.31)
1.3	0.43	(0.35, 0.51)	0.42	(0.24, 0.61)	0.78	(0.38, 1.22)
1.4	0.38	(0.33, 0.43)	0.41	(0.24, 0.61)	0.81	(0.39, 1.19)
1.5	0.40	(0.35, 0.46)	0.42	(0.24, 0.61)	0.86	(0.36, 1.44)
1.6	0.44	(0.38, 0.51)	0.42	(0.25, 0.61)	0.99	(0.36, 2.04)
1.7	0.61	(0.51, 0.75)	0.46	(0.27, 0.68)	1.14	(0.52, 2.13)
2.1	0.27	(0.24, 0.29)	0.42	(0.25, 0.60)	0.63	(0.30, 1.05)
2.2	0.27	(0.24, 0.29)	0.39	(0.22, 0.55)	0.73	(0.35, 1.23)
2.3	0.30	(0.27, 0.34)	0.38	(0.21, 0.55)	0.80	(0.38, 1.38)
2.4	0.32	(0.28, 0.36)	0.38	(0.22, 0.54)	0.84	(0.37, 1.38)
2.5	0.35	(0.30, 0.41)	0.39	(0.22, 0.54)	0.78	(0.37, 1.32)
2.6	0.36	(0.31, 0.43)	0.39	(0.22, 0.56)	0.87	(0.39, 1.35)
2.7	0.56	(0.44, 0.69)	0.43	(0.24, 0.63)	0.98	(0.42, 1.62)
3.1	0.26	(0.23, 0.28)	0.41	(0.24, 0.60)	0.68	(0.33, 1.12)
3.2	0.33	(0.28, 0.39)	0.38	(0.22, 0.56)	0.65	(0.34, 1.09)
3.3	0.38	(0.32, 0.43)	0.38	(0.22, 0.56)	0.76	(0.39, 1.27)
3.4	0.41	(0.33, 0.48)	0.38	(0.21, 0.56)	0.74	(0.38, 1.17)
3.5	0.43	(0.35, 0.52)	0.38	(0.21, 0.56)	0.74	(0.35, 1.22)
3.6	0.45	(0.37, 0.54)	0.38	(0.22, 0.56)	0.76	(0.35, 1.38)
3.7	0.52	(0.42, 0.62)	0.41	(0.24, 0.59)	0.87	(0.38, 1.41)
4.1	0.26	(0.23, 0.29)	0.41	(0.24, 0.60)	0.62	(0.31, 0.98)
4.2	0.34	(0.28, 0.40)	0.38	(0.22, 0.56)	0.64	(0.28, 0.95)
4.3	0.41	(0.33, 0.50)	0.38	(0.22, 0.56)	0.70	(0.32, 1.23)
4.4	0.51	(0.41, 0.62)	0.39	(0.22, 0.56)	0.84	(0.46, 1.43)

Table 2.9: (continued)

	Posterior SD					
	HSL		CAR		SLR	
SP	Mean	Range	Mean	Range	Mean	Range
4.5	0.54	(0.44, 0.65)	0.39	(0.23, 0.57)	0.91	(0.43, 1.44)
4.6	0.44	(0.34, 0.54)	0.39	(0.22, 0.56)	0.70	(0.30, 1.19)
4.7	0.41	(0.33, 0.49)	0.41	(0.25, 0.59)	0.71	(0.30, 1.25)
5.1	0.26	(0.24, 0.29)	0.41	(0.24, 0.61)	0.61	(0.31, 0.99)
5.2	0.33	(0.28, 0.40)	0.38	(0.22, 0.56)	0.60	(0.31, 0.99)
5.3	0.39	(0.32, 0.47)	0.38	(0.21, 0.55)	0.66	(0.34, 1.02)
5.4	0.47	(0.37, 0.58)	0.38	(0.22, 0.55)	0.74	(0.35, 1.30)
5.5	0.56	(0.43, 0.71)	0.39	(0.22, 0.56)	0.80	(0.42, 1.32)
5.6	0.48	(0.36, 0.60)	0.39	(0.22, 0.55)	0.72	(0.36, 1.14)
5.7	0.53	(0.41, 0.66)	0.42	(0.25, 0.59)	0.83	(0.39, 1.30)
6.1	0.28	(0.24, 0.33)	0.42	(0.25, 0.60)	0.60	(0.32, 1.00)
6.2	0.33	(0.28, 0.40)	0.39	(0.22, 0.56)	0.58	(0.32, 0.99)
6.3	0.38	(0.32, 0.46)	0.38	(0.22, 0.56)	0.68	(0.31, 1.20)
6.4	0.41	(0.32, 0.49)	0.39	(0.22, 0.54)	0.71	(0.36, 1.09)
6.5	0.47	(0.38, 0.58)	0.39	(0.23, 0.54)	0.74	(0.37, 1.16)
6.6	0.44	(0.36, 0.54)	0.39	(0.22, 0.56)	0.75	(0.37, 1.24)
6.7	0.47	(0.38, 0.59)	0.42	(0.25, 0.60)	0.86	(0.41, 1.39)
7.1	0.24	(0.22, 0.27)	0.45	(0.28, 0.64)	0.69	(0.32, 1.25)
7.2	0.30	(0.27, 0.34)	0.42	(0.24, 0.60)	0.72	(0.36, 1.24)
7.3	0.37	(0.31, 0.42)	0.41	(0.24, 0.59)	0.77	(0.34, 1.14)
7.4	0.41	(0.35, 0.48)	0.41	(0.24, 0.59)	0.86	(0.39, 1.34)
7.5	0.44	(0.37, 0.53)	0.41	(0.25, 0.59)	0.83	(0.35, 1.40)
7.6	0.50	(0.39, 0.61)	0.42	(0.25, 0.60)	0.81	(0.37, 1.37)
7.7	0.55	(0.46, 0.66)	0.46	(0.28, 0.67)	1.11	(0.45, 2.01)

2.5.5 Supplementary figures

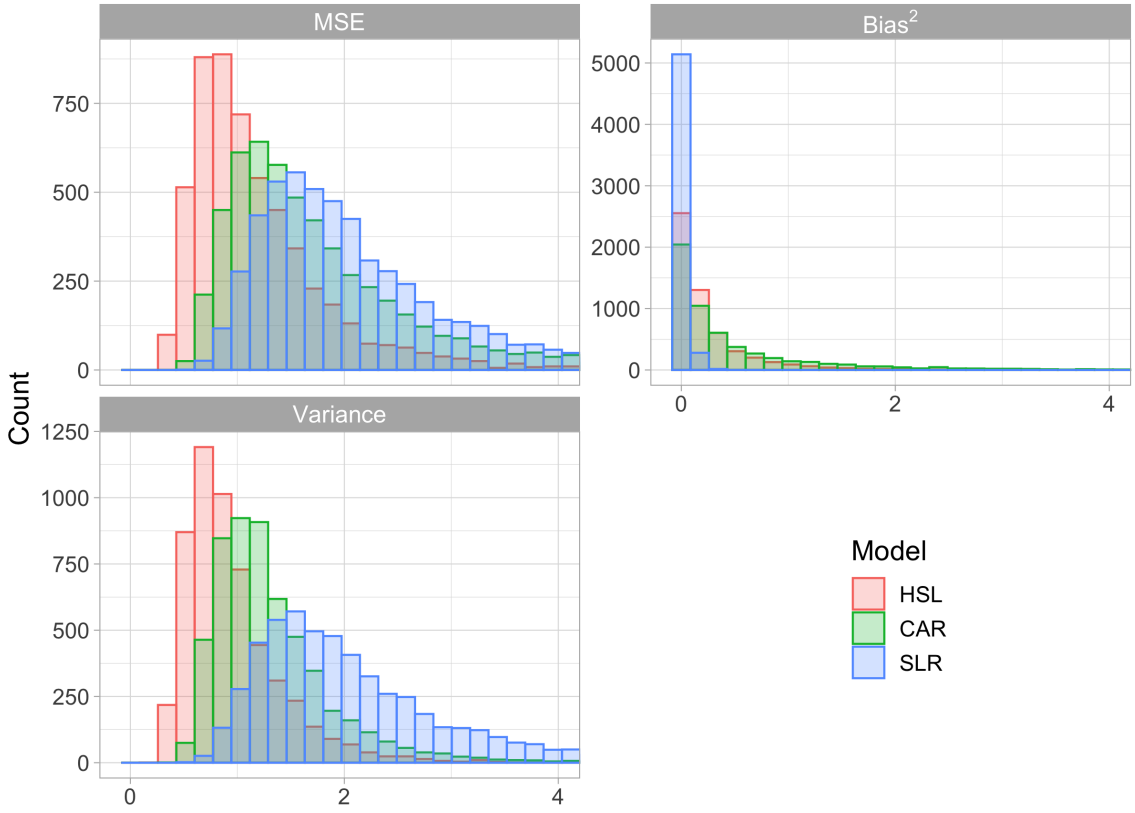

Figure 2.6: Histograms of mean squared error (MSE) (upper left), squared bias (upper right), and estimator variance (lower left) for patient-superpixel intercepts comparing simple linear regression (SLR), conditional autoregressive (CAR), and hierarchical spatial longitudinal (HSL) models pooled across all superpixels from the simulation study. Mean squared error is squared bias plus estimator variance; the x-axis values are directly comparable in all three plots. Lower MSE indicates better model performance. Counts of large values omitted due to truncating the x-axis at 4: MSE (HSL: 75; CAR: 271; SLR: 377), squared bias (HSL: 12; CAR: 68; SLR: 0), and estimator variance (HSL: 4; CAR: 25; SLR: 368).

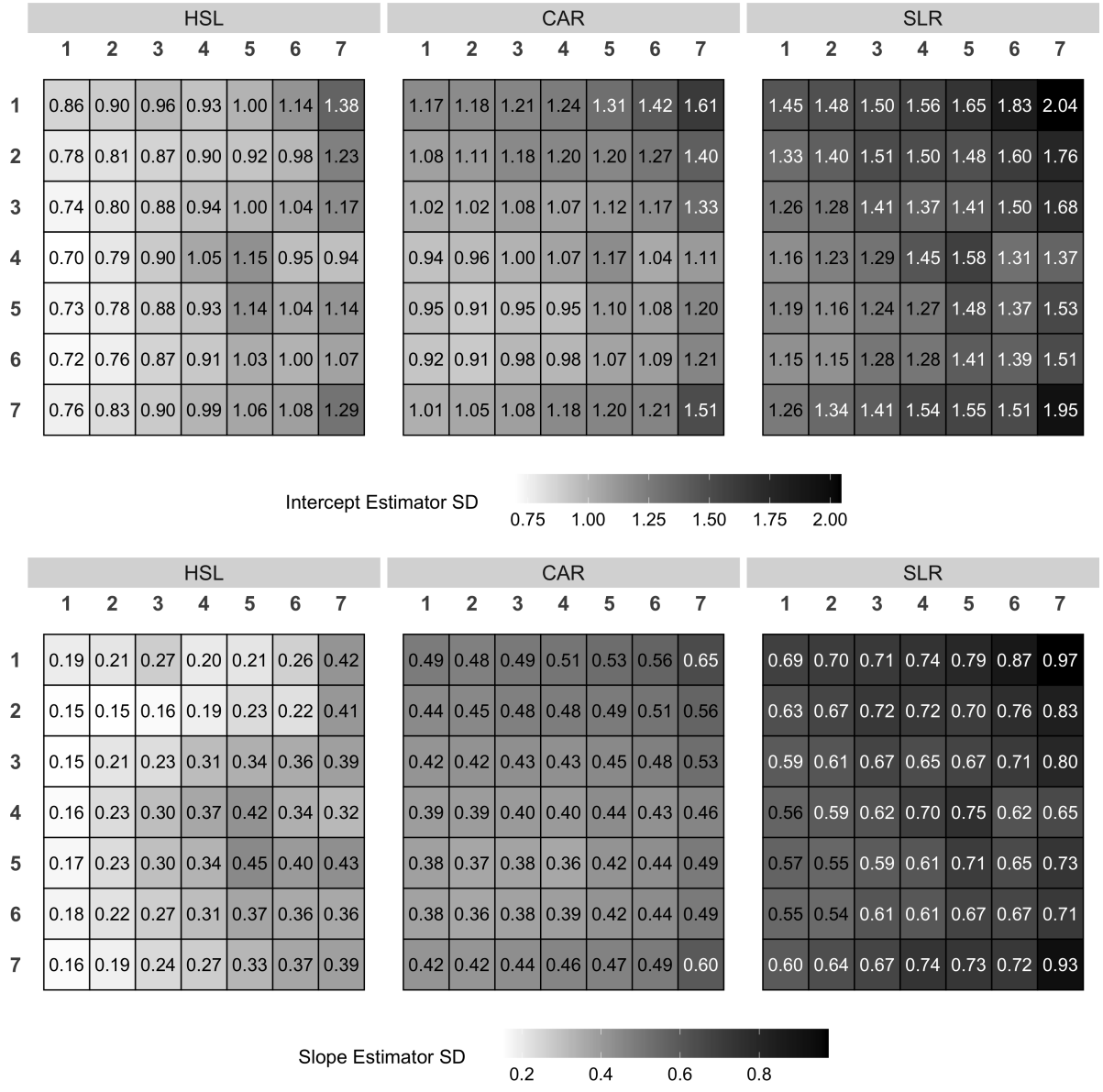

Figure 2.7: Heat map of the average estimator standard deviations (SD) of intercepts (top) and slopes (bottom) by superpixel location for hierarchical spatial longitudinal (HSL) (left), conditional autoregressive (CAR) (middle), and simple linear regression (SLR) (right) models from the simulation study.

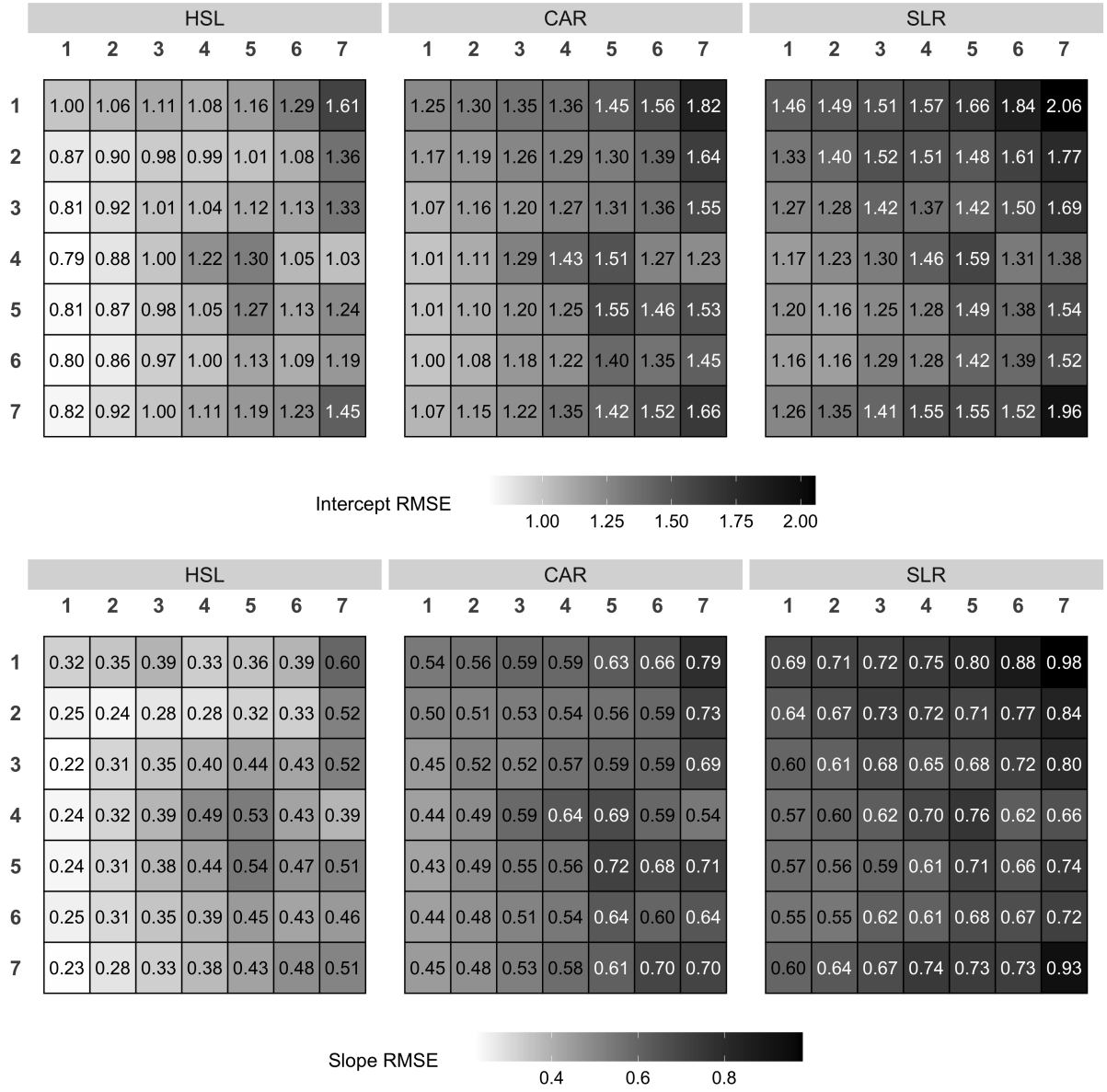

Figure 2.8: Heat map of the average root mean squared error (RMSE) of intercepts (top) and slopes (bottom) by superpixel location for hierarchical spatial longitudinal (HSL) (left), conditional autoregressive (CAR) (middle), and simple linear regression (SLR) (right) models from the simulation study.

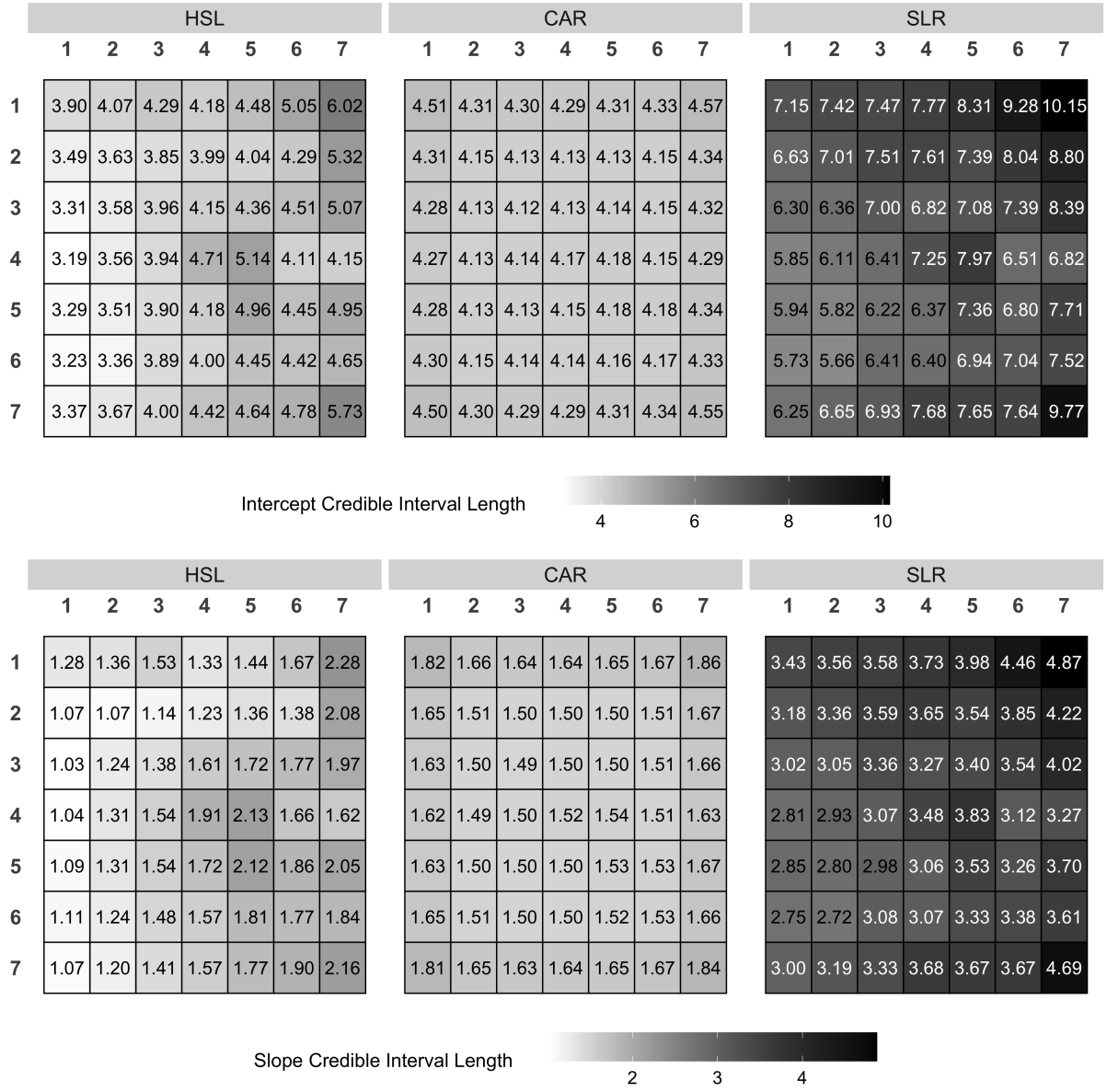

Figure 2.9: Heat map of the average 95% Credible Interval Length of intercepts (top) and slopes (bottom) by superpixel location for hierarchical spatial longitudinal (HSL) (left), conditional autoregressive (CAR) (middle), and simple linear regression (SLR) (right) models from the simulation study.

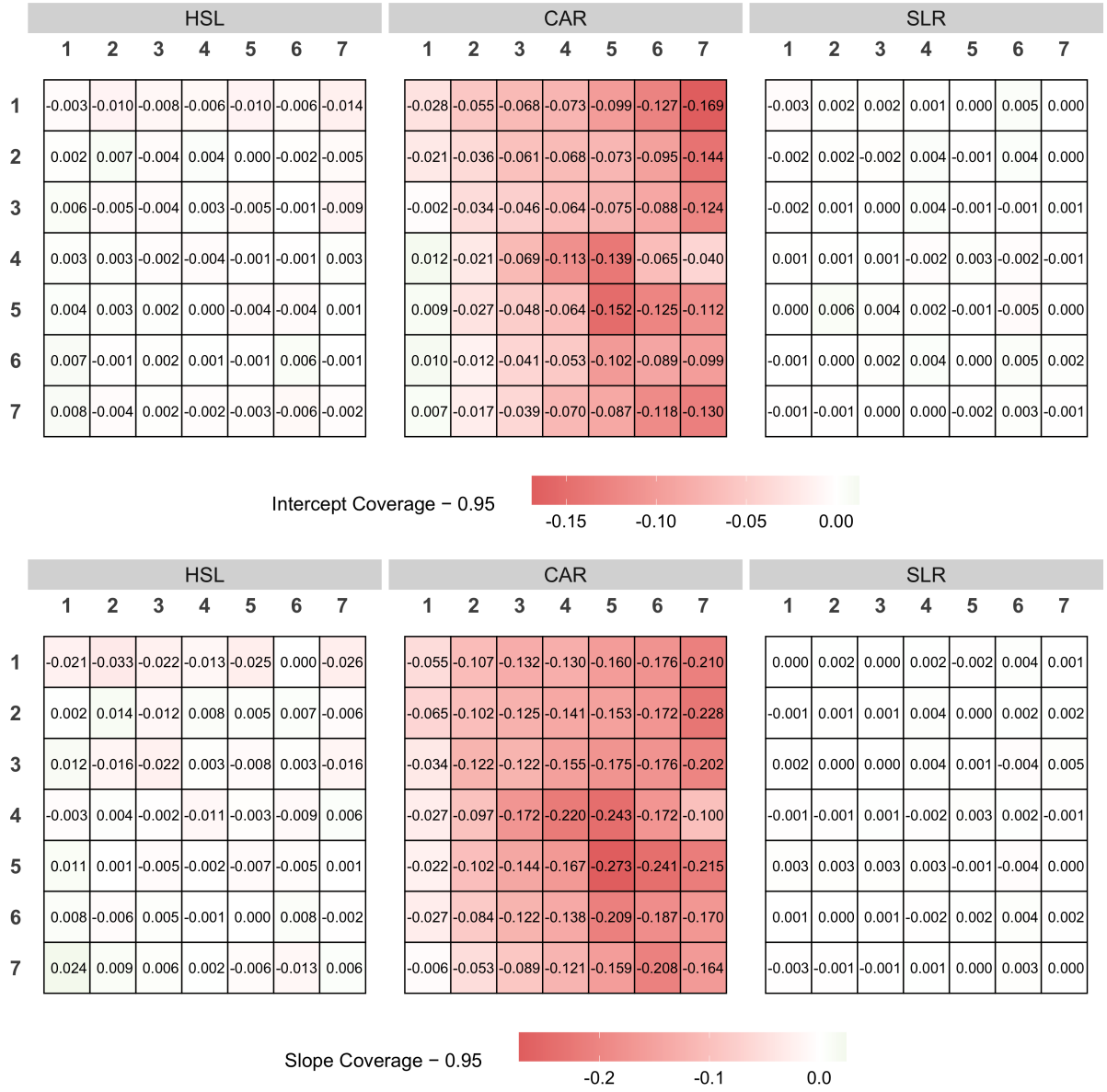

Figure 2.10: Heat map of the 95% coverage probability minus 0.95 of intercepts (top) and slopes (bottom) by superpixel location for hierarchical spatial longitudinal (HSL) (left), conditional autoregressive (CAR) (middle), and simple linear regression (SLR) (right) models from the simulation study.

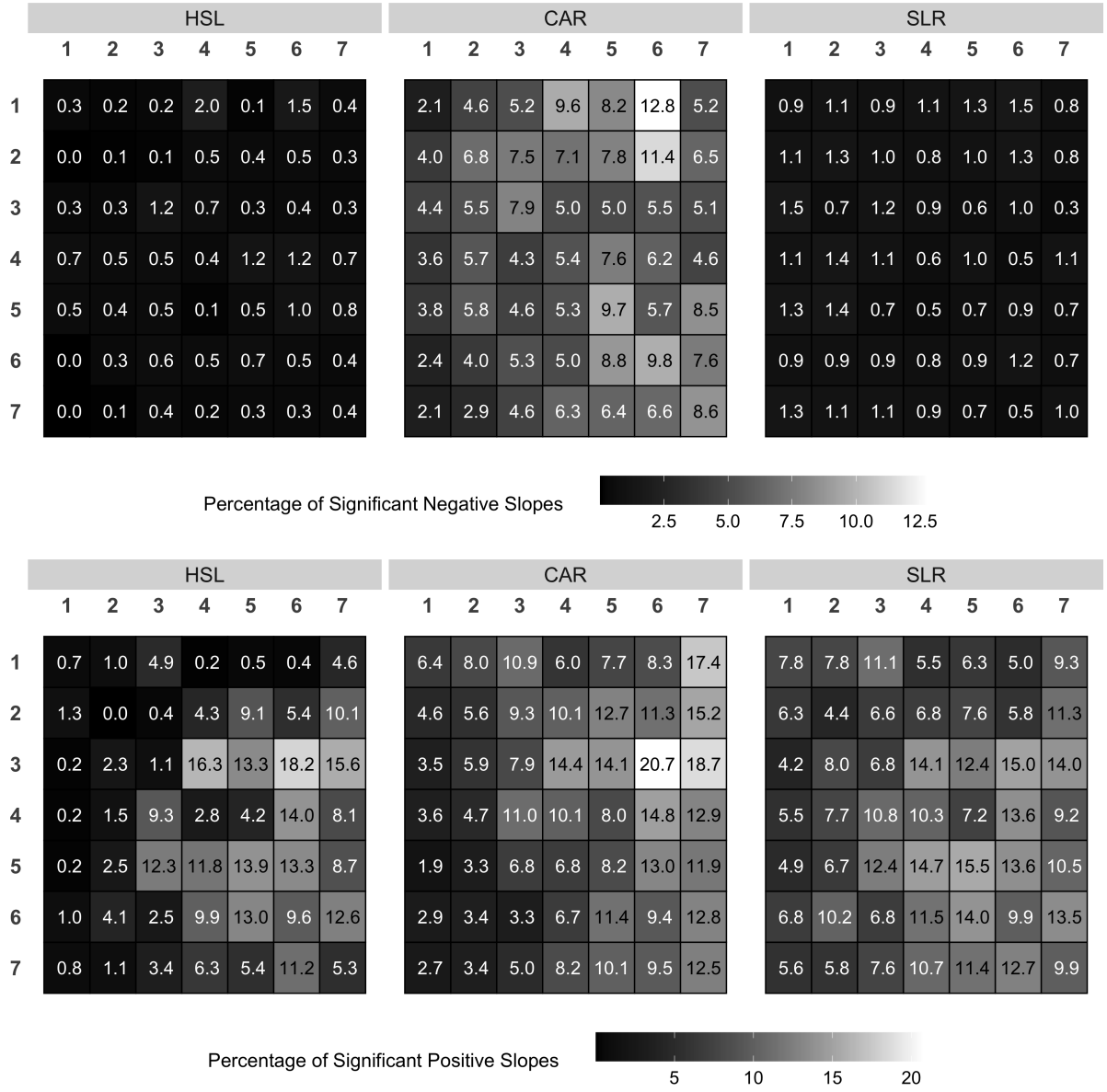

Figure 2.11: Heat map of the percentage of significant negative slopes (top) and significant positive slopes (bottom) detected by hierarchical spatial longitudinal (HSL) (left), conditional autoregressive (CAR) (middle), and simple linear regression (SLR) (right) models when the true slope is positive from the simulation study.

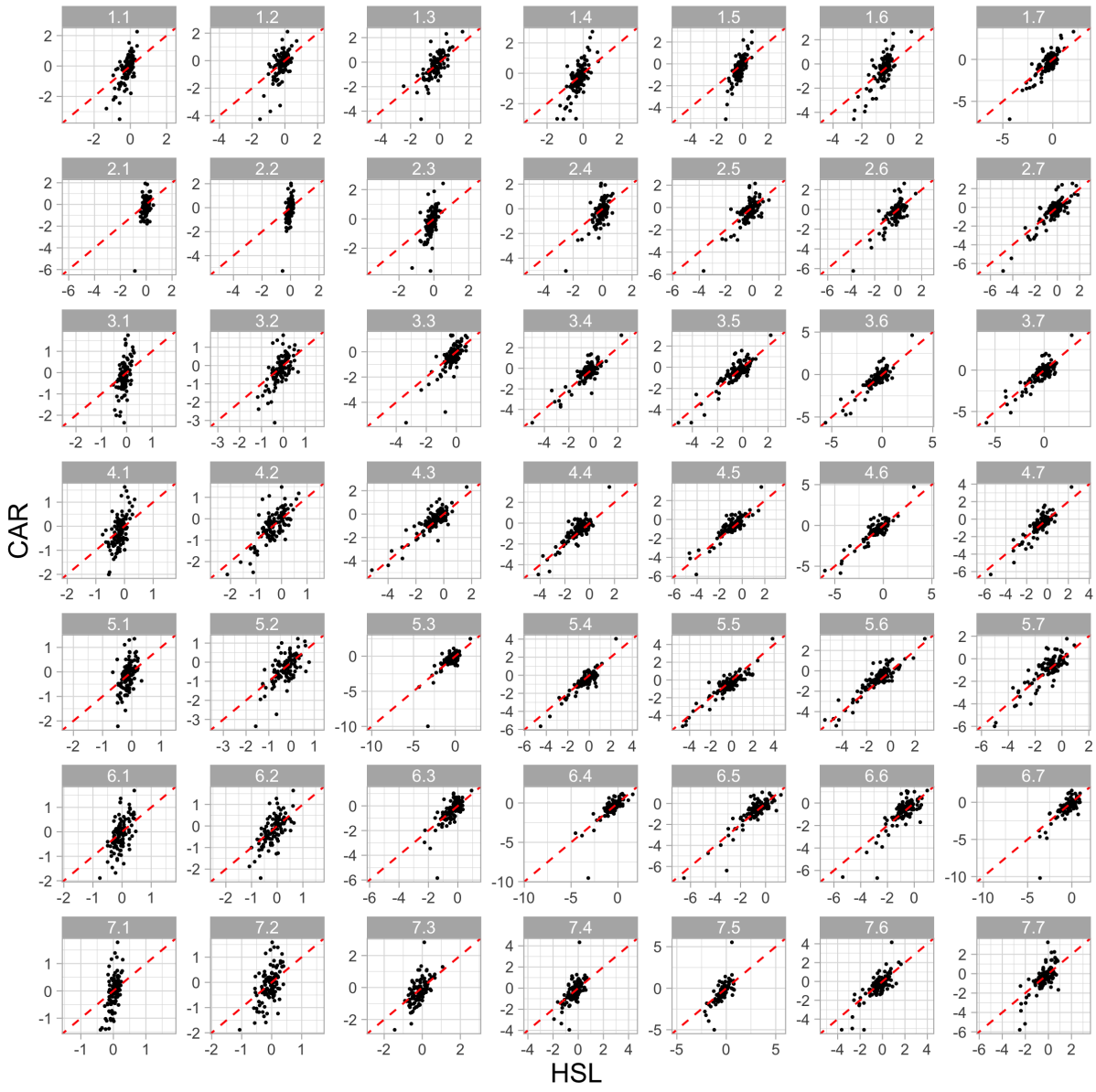

Figure 2.12: Scatter plots of slope posterior means from conditional autoregressive model (CAR, y-axis) against those from the hierarchical spatial longitudinal model (HSL, x-axis) in each superpixel for the patient cohort data. The CAR posterior means are more variable than the HSL means in temporal superpixels. There is noticeable shrinkage towards the population mean in the HSL estimates in the temporal regions. Each plot is square with its own axes with the x- and y- axes having the same range. The red dashed line represents the $x = y$ diagonal.

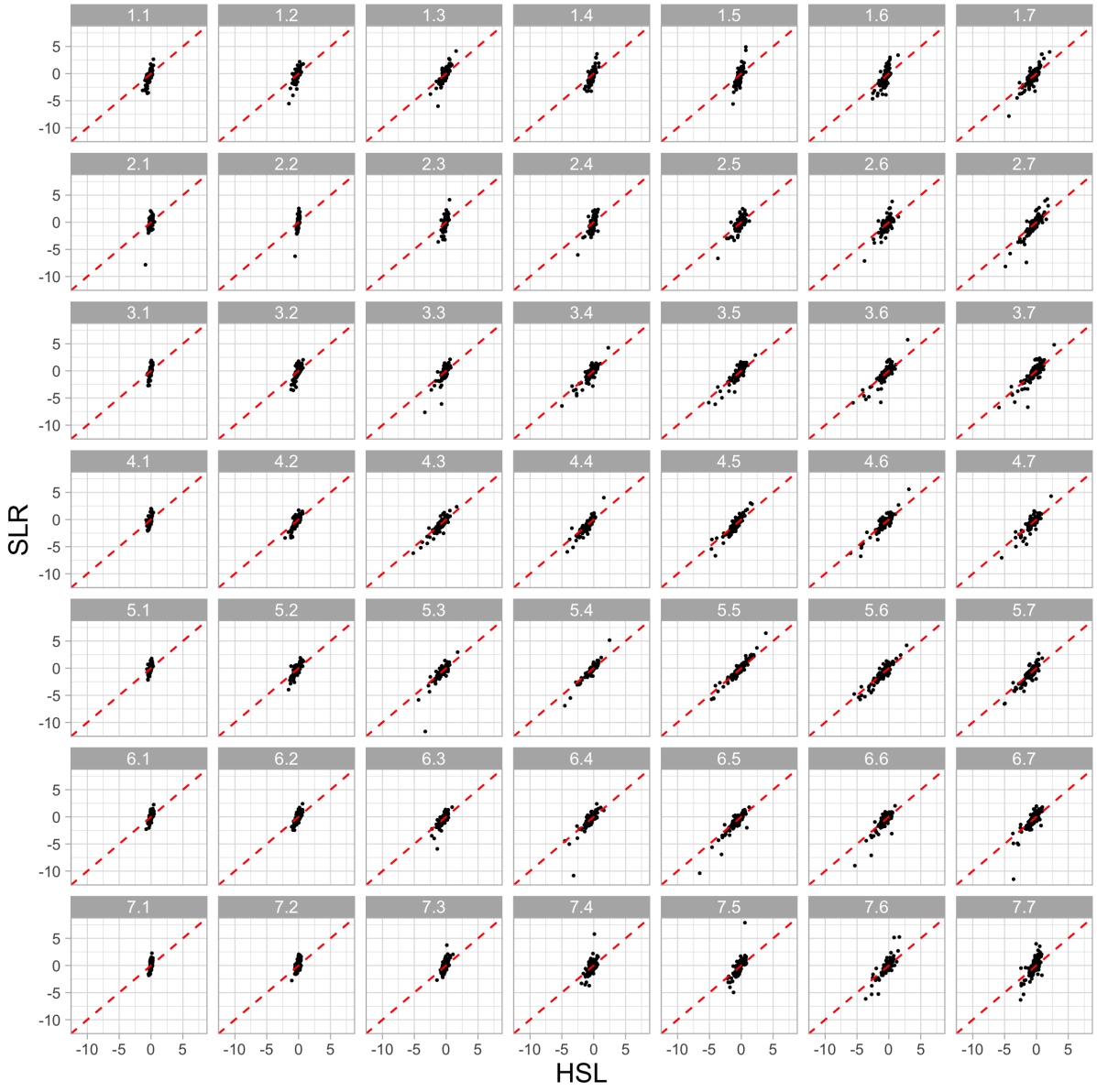

Figure 2.13: Scatter plots of slope posterior means from simple linear regression model (SLR, y-axis) against those from the hierarchical spatial longitudinal model (HSL, x-axis) in each superpixel for the patient cohort data. The SLR posterior means are much more variable than the HSL means. There is noticeable shrinkage towards the population mean in the HSL estimates in the peripheral superior and temporal regions. Each plot is square with the x- and y- axes having the same range across all 49 superpixels. The red dashed line represents the $x = y$ diagonal.

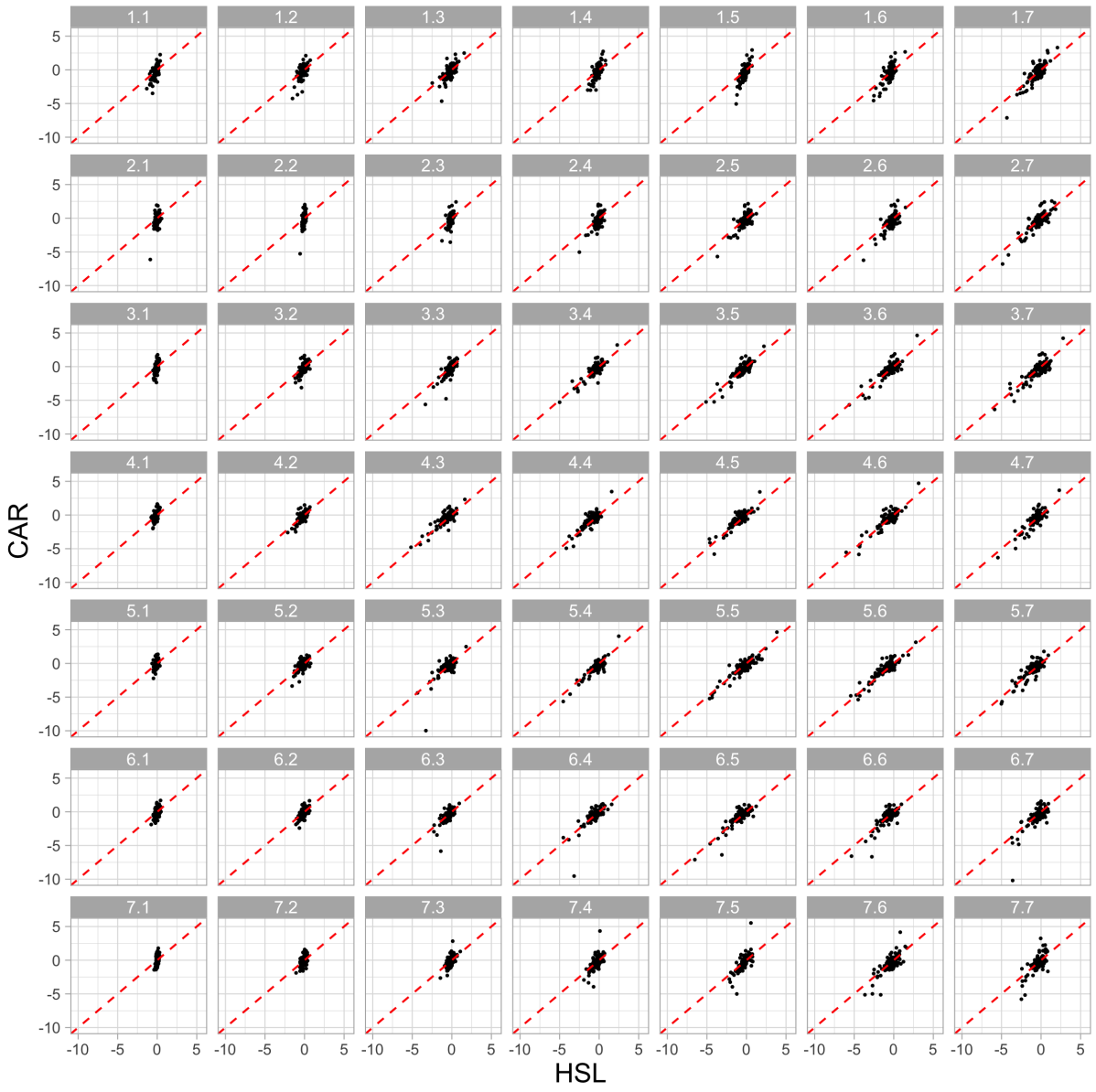

Figure 2.14: Scatter plots of slope posterior means from conditional autoregressive model (CAR, y-axis) against those from the hierarchical spatial longitudinal model (HSL, x-axis) in each superpixel for the patient cohort data. The CAR posterior means are more variable than the HSL means in temporal superpixels. There is noticeable shrinkage towards the population mean in the HSL estimates in the temporal regions. Each plot is square with the x- and y- axes having the same range across all 49 superpixels. The red dashed line represents the $x = y$ diagonal.

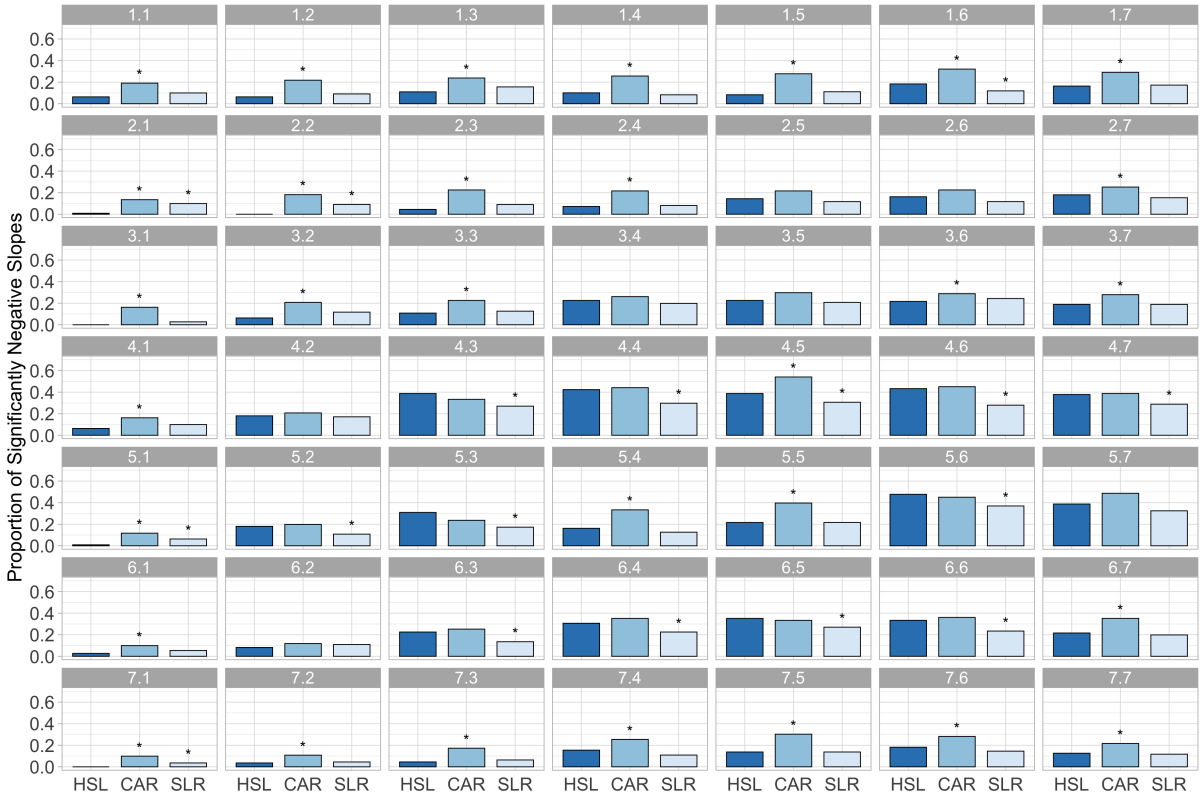

Figure 2.15: The proportion of significant negative slopes for each superpixel detected by hierarchical spatial longitudinal (HSL), conditional autoregressive (CAR), and simple linear regression (SLR) for the data from the patient cohort. The asterisk on the CAR bar indicates a significant p-value from McNemar's tests comparing proportions between HSL and CAR and similarly for the asterisk on the SLR bar for the comparison between HSL and SLR.

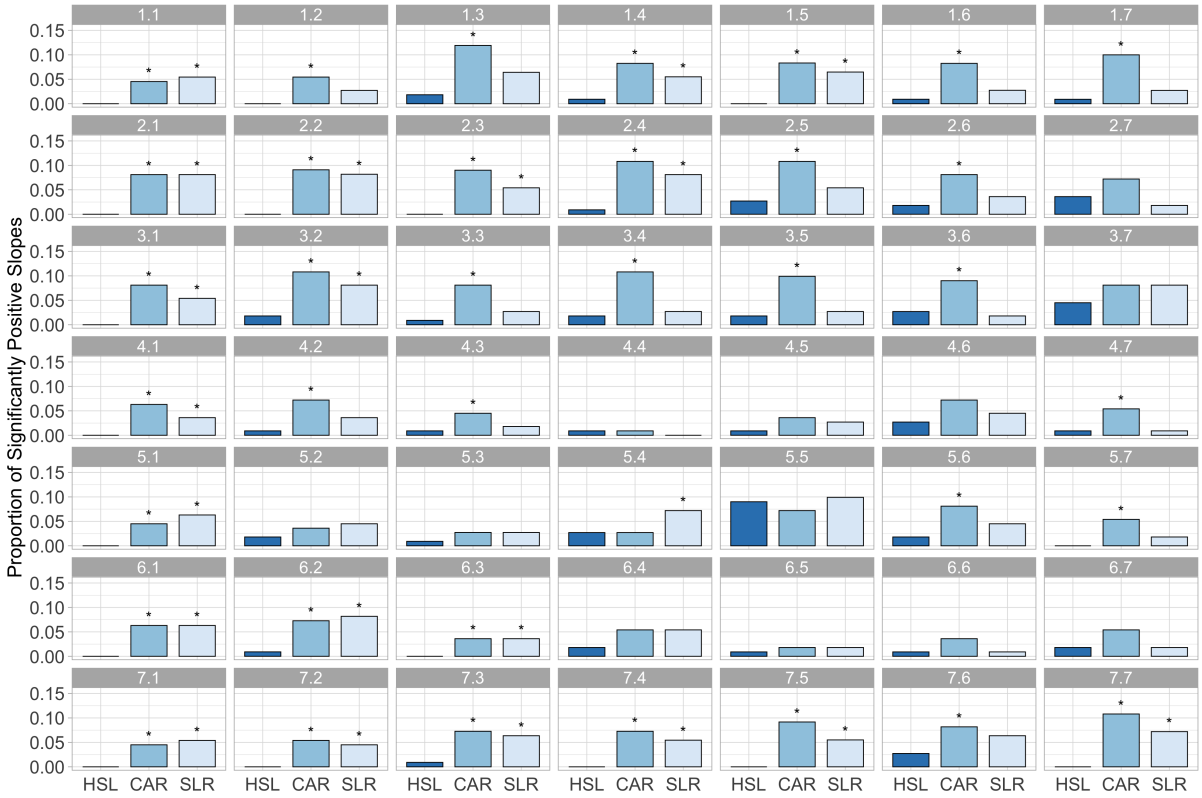

Figure 2.16: The proportion of significant positive slopes for each superpixel detected by hierarchical spatial longitudinal (HSL), conditional autoregressive (CAR), and simple linear regression (SLR) for the data from the patient cohort. The asterisk on the CAR bar indicates a significant p-value from McNemar's tests comparing proportions between HSL and CAR and similarly for the asterisk on the SLR bar for the comparison between HSL and SLR.

CHAPTER 3

A spatially varying hierarchical random effects model for longitudinal macular structural data in glaucoma patients

3.1 Introduction

Glaucoma damages the optic nerve and is the second leading cause of blindness worldwide (Kingman, 2004). As there is no cure, timely detection of disease progression is imperative to identify eyes at high risk of or demonstrating early progression so that timely treatment can be provided and further visual loss prevented. Ophthalmologists assess glaucomatous progression by monitoring functional changes in visual fields or structural changes in the retina over time. Visual field (VF) measurements assess functional changes by measuring how well eyes are able to detect light. Repeatedly measuring the thickness of retinal layers, such as macular ganglion cell complex (GCC), with optical coherence tomography (OCT) allows ophthalmologists to evaluate central retinal (macular) structural change over time. Both VF and OCT obtain data from multiple locations across the retina. In current practice, clinicians detect progression by modeling functional or structural changes over time using simple linear regression (SLR) for each subject-location combination (Gardiner and Crabb, 2002; Nouri-Mahdavi et al., 2007; Tatham and Medeiros, 2017; Thompson et al., 2020a). SLR does not accommodate the hierarchical structure that patients are members of a population and ignores the spatial arrangement of the data. For analyzing VF data at individual locations, Montesano et al. (2021) introduce a hierarchical model accounting for location and cluster levels fit to data from a single eye, Betz-Stablein et al. (2013) and Berchuck et al. (2019) present models accounting for spatial correlation fit to data from a single eye, and Bryan et al. (2017) describe a two-stage

approach to fit a hierarchical model taking subject, eye, hemifield (one half of the VF), and location into account. While these methods exist for VF data, they cannot be directly applied to structural macular data as the measurement processes are markedly different. Key features of VF data that differ from structural data include censoring, heteroskedasticity, and a different underlying spatial structure.

We analyze data from the Advanced Glaucoma Progression Study (AGPS), a cohort of eyes with moderate to severe glaucoma. To monitor glaucoma progression, we model longitudinal macular GCC thickness measurements over a square 6×6 grid of 36 superpixels (roughly a $20° \times 20°$ area) for all subjects. For a single subject, the intercepts, slopes, and residual standard deviations (SD) vary spatially across superpixel locations. Mohammadzadeh et al. (2021) model GCC data from each superpixel separately and compare different Bayesian hierarchical models, preferring a model with random intercepts, random slopes, and random residual SDs. Our desired model needs to account for both the hierarchical structure of the data and the spatial correlations in both the population- and subject-level intercepts, slopes, and residual SDs and in the residuals. The parameters at the population level summarize information from the whole cohort at each superpixel location. Additional difficulties in modeling GCC data arise from the amount and sources of measurement error. Thickness measurements are reliant on automated segmentation algorithms, which may introduce spatially correlated errors unique to each imaging scan. We show that including visit effects to account for visit-specific errors reduces error in predicting future thickness measurements and greatly improves model fit. In this study, we motivate and develop the Spatially varying Hierarchical Random Effects with Visit Effects (SHREVE) model, a novel Bayesian hierarchical model with spatially varying population- and subject-level coefficients and SDs, accounting for spatial and within-subject correlation, between-subject variation, and spatially correlated visit-specific errors.

For the AGPS data, we allow the intercepts, slopes, and residual SDs to vary over space. Varying coefficient models are natural extensions to classical linear regression and extensively used in imaging studies and the analysis of spatial data (Hastie and Tibshirani, 1993; Ge et al., 2014; Zhu et al., 2014a; Liu et al., 2019), where regression coefficients are allowed to vary smoothly as a func-

tion of one or more variables, and in our case, over spatial locations. Regression coefficients may vary over space in a discrete fashion as with areal units or in a continuous manner as with point-referenced data (Gelfand et al., 2010). In the context of imaging studies with grid data, a conditional autogressive (CAR) model (Gössl et al., 2001; Penny et al., 2005; Ge et al., 2014) or a Gaussian process (GP) model (Zhang et al., 2016; Castruccio et al., 2018) may be assumed for discrete or continuous spatial variation, respectively. In a GP model, coefficients from any finite set of locations has a multivariate normal distribution with a mean function and valid covariance function specifying the expected value at each location and covariance between coefficients at any two locations, respectively (Gelfand et al., 2010).

Gelfand et al. (2003) first proposed the use of GPs to model spatially varying regression coefficients and multivariate Gaussian processes (MGP) for multiple spatially varying regression coefficients in a hierarchical Bayesian framework. We can assign GP priors at different levels in the hierarchy, which allows for flexible specification in hierarchical models (Gelfand and Schliep, 2016; Kim and Lee, 2017). In our case with three components, spatially varying intercepts, slopes, and residual SDs, we employ MGPs to model the correlations between components within a location and across locations at both the subject and population level. MGPs are specified with a multivariate mean function and cross-covariance function, defining the covariance between any two coefficients at any two locations (Banerjee et al., 2014). For simplicity and computational convenience, separable cross-covariance functions are often used where components share the same spatial correlation and components within a location share a common covariance matrix, and the resulting covariance matrix is the Kronecker product of a covariance matrix between components and a spatial correlation matrix (Banerjee et al., 2014). Assuming all components share a common spatial correlation structure is likely inadequate in practice, as processes may be very different from each other in nature. Instead, we propose a nonseparable cross-covariance function to allow each process to have its own spatial correlation function.

Constructing valid cross-covariance models is a challenging task for nonseparable MGPs. Genton and Kleiber (2015) review approaches to construct valid cross-covariance functions for MGPs in-

cluding the linear model of coregionalization (Wackernagel, 2013; Schmidt and Gelfand, 2003) and kernel and covariance convolution methods (Ver Hoef and Barry, 1998; Gaspari and Cohn, 1999). For univariate GPs, the Matérn class of covariance models is widely used, featuring a smoothness parameter that defines the level of mean square differentiability and a lengthscale parameter that defines the rate of correlation decay (Guttorp and Gneiting, 2006). Gneiting et al. (2010) and Apanasovich et al. (2012) introduce multivariate Matérn models and provide necessary and sufficient conditions to allow the cross-covariance functions to have any number of components (processes) while allowing for different smoothnesses and rates of correlation decay for each component. We propose such a multivariate Matérn construction to model our spatially varying intercepts, slopes, and residual SDs, so that each component is allowed its own spatial correlation structure.

In Section 3.2, we describe the motivating data. In Section 3.3, we briefly review GPs and develop the SHREVE model. In Section 3.4, we apply the SHREVE model to GCC data and compare its performance to several nested models lacking visit effects or other model components. We give a concluding discussion in Section 3.5.

3.2 Ganglion cell complex data

This section highlights data characteristics that motivate model development. We provide details on the imaging procedure and study subjects.

3.2.1 Macular optical coherence tomography

Macular OCT has emerged as a standard imaging modality to assess changes in retinal ganglion cells (RGCs) (Mohammadzadeh et al., 2020a). As glaucoma is characterized by progressive loss of RGCs, clinicians use macular OCT as a means to monitor changes in retinal thickness over time (Weinreb and Khaw, 2004). Macular GCC thickness, measured in microns (μm), has been shown to be more efficient for detecting structural loss regardless of glaucoma severity compared to measures of other macular layers (Mohammadzadeh et al., 2022a). Glaucomatous damage to the macular area,

reflected in thinning of GCC, has been associated with VF loss (Mohammadzadeh et al., 2020b). Visual field loss occurs when part(s) of the peripheral vision is (are) lost.

3.2.2 Advanced Glaucoma Progression Study

We analyze data from the AGPS (Mohammadzadeh et al., 2021, 2022a,b), an ongoing longitudinal study at the University of California, Los Angeles. The study adhered to the tenets of the Declaration of Helsinki and conformed to Health Insurance Portability and Accountability Act policies. All patients provided written informed consent at the time of enrollment in the study. The data include GCC thickness measurements from 111 eyes with at least 4 OCT scans and a minimum of approximately 2 years of observed follow-up time, up to 4.25 years from baseline. Subjects returned approximately every 6 months for imaging using Spectralis OCT (Heidelberg Engineering, Heidelberg, Germany). This device acquires $30^\circ \times 25^\circ$ volume scans centered on the fovea, the center of the macula represented as a white dot in Figure 3.1 and as a black dot in subsequent figures (Mohammadzadeh et al., 2020a). We used built-in software, the Glaucoma Module Premium Edition, to automatically segment macular layers of interest. GCC thickness is calculated by summing the thicknesses of the retinal nerve fiber layer, inner plexiform layer, and ganglion cell layer. The posterior pole algorithm of the Spectralis reports layer thickness averaged over pixels within a *superpixel* with superpixels forming an 8×8 grid of locations, as shown in Figure 3.1. We display superpixels in right eye orientation with superpixels labeled as row number 1-8, a dot, then column number 1-8. Superpixels in rows 1-4 are located in the *superior hemiretina* and rows 5-8 are located in the *inferior hemiretina*; the temple and nose are to the left and right, respectively. Left eyes are mirror images of right eyes and are flipped left-right for presentation and analysis. Because there is substantial measurement noise in the outer ring of superpixels, rows 1 and 8 and columns 1 and 8 (Miraftabi et al., 2016), we analyze only the central 6×6 superpixels as shown in Figure 3.1.

1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8
2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8
3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8
4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8
5.1	5.2	5.3	5.4	5.5	5.6	5.7	5.8
6.1	6.2	6.3	6.4	6.5	6.6	6.7	6.8
7.1	7.2	7.3	7.4	7.5	7.6	7.7	7.8
8.1	8.2	8.3	8.4	8.5	8.6	8.7	8.8

Figure 3.1: Visualization of the 8×8 grid of superpixels and labels from the Spectralis posterior pole algorithm. The inner 36 superpixels included in the analysis are shaded in gray and delineated with thicker lines. Superpixels are shown in right eye orientation where rows 1-4 are located in the *superior hemiretina* and rows 5-8 are located in the *inferior hemiretina*; the temple and nose are to the left and right, respectively. Superpixels labels are row number 1-8, a dot, then column number 1-8. The black dot indicates the foveal center for visual orientation.

3.2.3 Data exploration

Let observation y_{ijk} be the GCC thickness measure in μm of subject $i = 1,\ldots,n$ at visit $j = 1,\ldots,J_i$, where J_i is the number of visits for subject i, in superpixel $k = 1,\ldots,K$ observed at time t_{ij}, with $t_{i1} = 0$ for all subjects. Location $\boldsymbol{s}_k = (\text{row}_k, \text{column}_k)$ denotes the spatial coordinates of superpixel k in two-dimensional space. Initially, we remove any zero thickness values $y_{ijk} = 0$, which indicate errors of measurement. We define a profile for subject i in superpixel k as the sequence of observations (t_{ij}, y_{ijk}) from visits $j = 1,\ldots,J_i$ and plot profiles of GCC thickness against time by connecting consecutive observations with line segments. For all subjects and superpixels, we plotted data in profile plots, which identified a number of outliers. We applied a semi-automated algorithm to identify pairs of consecutive points that have large differences in GCC thicknesses between the consecutive visits. For each subject and superpixel, we calculated the consecutive-visit absolute differences $|y_{ijk} - y_{i(j-1)k}|$ and the consecutive-visit centered-slopes $|y_{ijk} - y_{i(j-1)k}/(t_{ij} - t_{i(j-1)}) + 0.5|$, which were centered around -0.5 μm/year, the mean of slopes across all pairs of consecutive

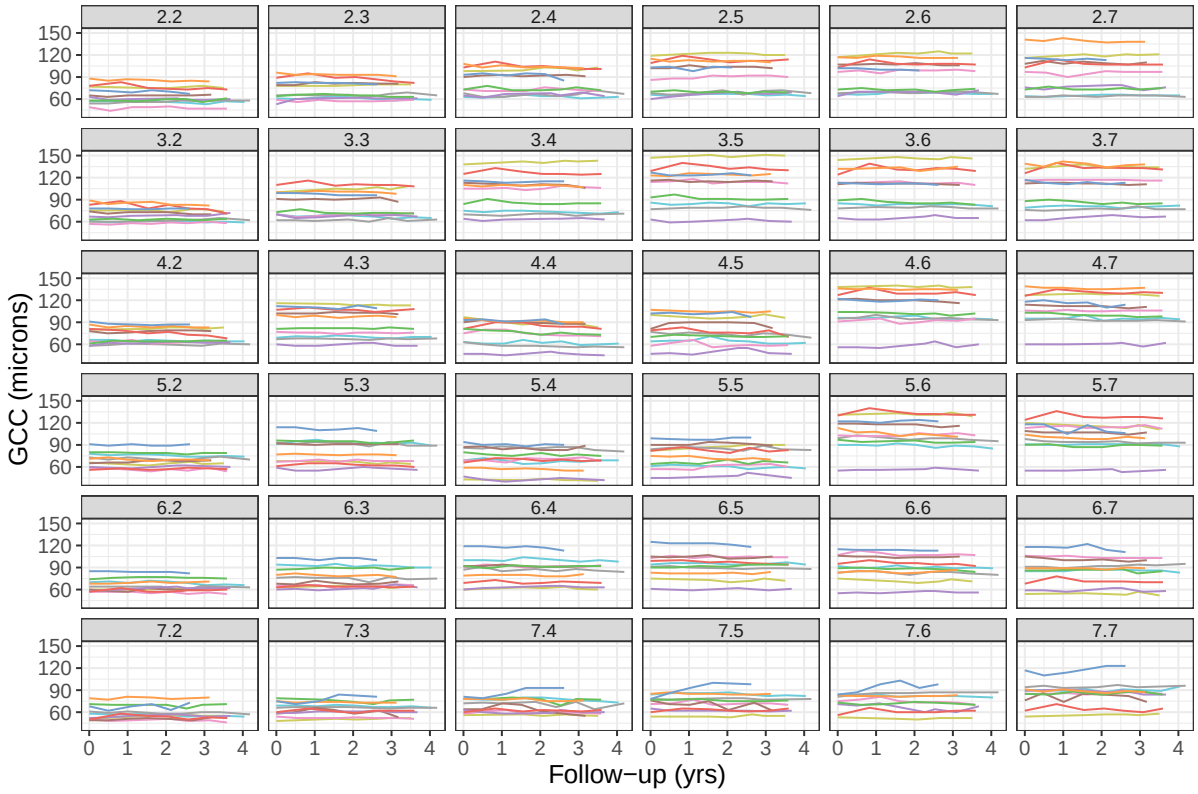

Figure 3.2: Profile plots of ganglion cell complex (GCC) thickness measurements for 10 subjects across 36 superpixels against follow-up time in years since baseline visit. Each color represents a different subject. These profiles illustrate the variability in baseline GCC thickness across the 10 subjects within superpixels, with a range within a superpixel of up to 84 μm. The average baseline thicknesses over subjects vary across superpixels, generally increasing from the temporal to nasal regions (left to right).

visits for all subjects and superpixels. We flagged pairs of observations $(y_{ijk}, y_{i(j-1)k})$ with absolute centered-slopes greater than 24 μm/year with absolute differences greater than 5 μm as candidates for removal. We calculated the sum of absolute visit differences for each profile $\sum_{j=2}^{J_i} |y_{ij} - y_{i(j-1)k}|$ and removed the point that resulted in the largest reduction in the sum of absolute visit differences. For each profile, if two or more observations were identified as outliers, we removed all remaining observations as well.

Eyes enrolled in the AGPS had moderate to severe glaucoma, thus exhibit a range of glaucomatous damage. Figure 3.2 shows profile plots after outlier removal of GCC thickness in μm against time in years since baseline visit for 10 subjects at all 36 superpixels. Baseline GCC varies across subjects within superpixels, with maximum differences in thicknesses between any of the AGPS subjects ranging from 40 to 100 μm across superpixels. From Figure 3.2, we note that intercepts are spatially correlated and repeated thickness measurements for each subject at each superpixel are highly correlated. The leftmost, temporal superpixels tend to have lower baseline thicknesses and smaller spread than rightmost, nasal superpixels and nasal superpixels show more variability both within and between subjects.

Figure 3.3 shows heat maps of GCC measurements over time for four subjects. Each row represents a different subject and each block of 6×6 superpixels displays the GCC thicknesses observed in rows 2-7 and columns 2-7 at the labeled follow-up time above the block. The range of baseline thicknesses across superpixels varies across subjects, with the first subject's baseline values ranging between 53 and 82 μm, while the third subject's baseline values range between 59 and 115 μm. Changes in GCC thickness over time also differ between Subject 1 and Subject 3. Subject 3 has noticeable decrease in thickness, thinning over time in many superpixels (e.g., 2.7, 3.3, and 4.3), while Subject 1 is more stable over time. Within subjects, there is a range of baseline thicknesses and changes over time across superpixels. These data characteristics motivate the need to model spatially varying random intercepts and slopes. Analyzing longitudinal GCC data separately in each superpixel, Mohammadzadeh et al. (2021) show that models with subject-specific residual SDs perform better than models with fixed residual SDs. Figure 3.4 shows heat maps of estimated slopes

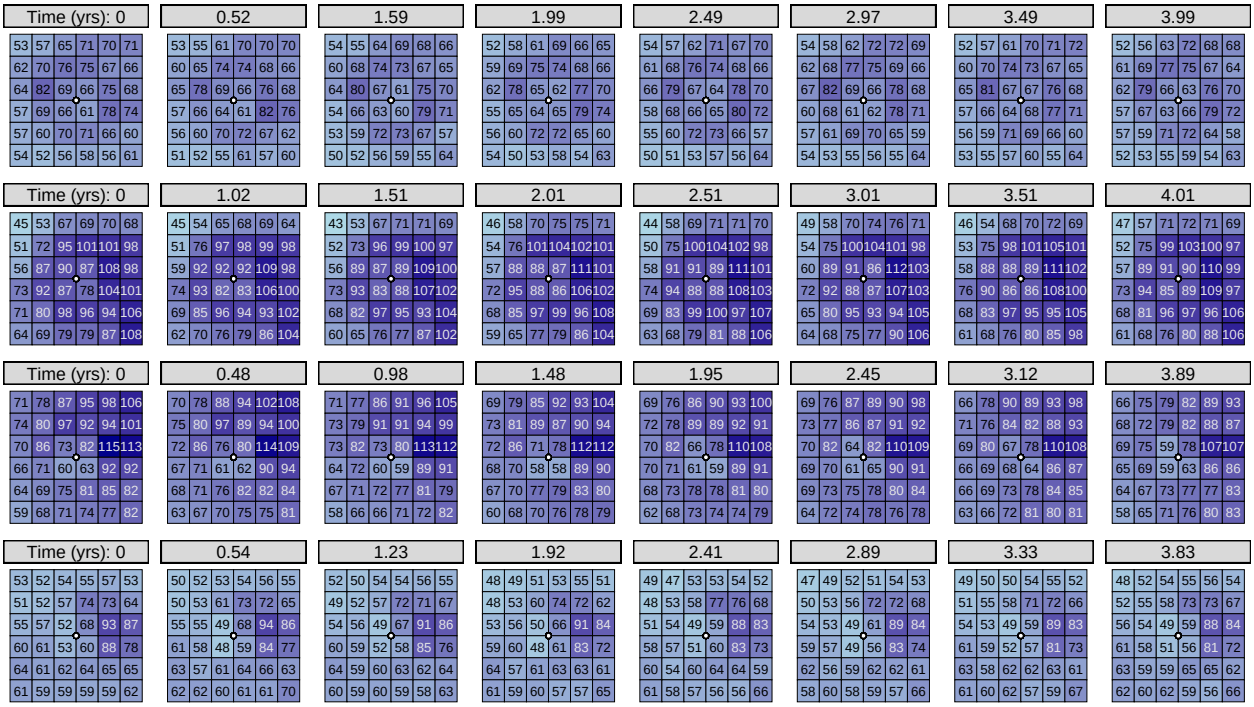

Figure 3.3: Heat maps of ganglion cell complex (GCC) thickness measurements (μm) across 8 visits for 4 subjects for all 36 superpixels (top left 2.2 to bottom right 7.7). Each row is a different subject. The follow-up time of each visit is labeled at the top of each block. All maps share a common color scale for comparison. GCC measurements are highly correlated within subjects over time, illustrated by similar color patterns over time. The color patterns also highlight the spatial correlation between locations. GCC measurements are highly variable across subjects, as seen by the difference in color shades. Over time, the third row subject has noticeable thinning in many superpixels while the other subjects are more stable in comparison.

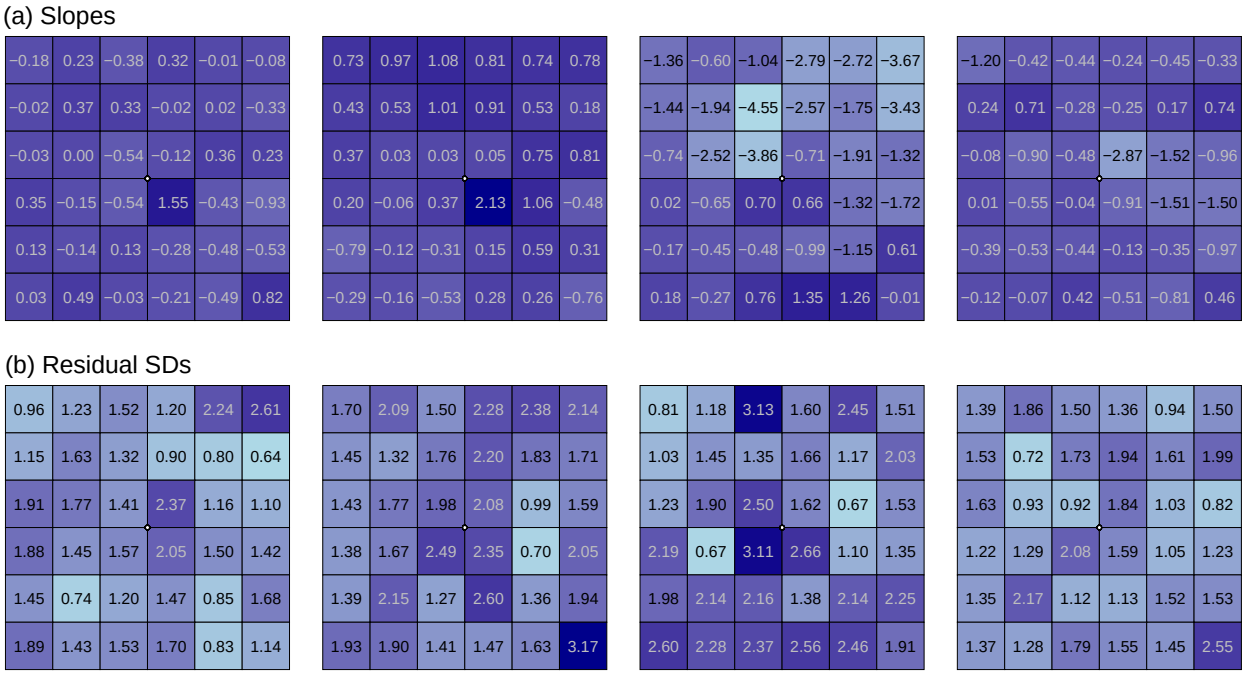

Figure 3.4: Heat maps of (a) estimated slopes (μm/year) and (b) residual standard deviations (SD) (μm) for the same 4 subjects as in Figure 3.3 using simple linear regressions of ganglion cell complex (GCC) thicknesses on time since baseline in each superpixel. Each column is a different subject. Estimated slopes appear spatially correlated within subjects. Subject 3 has particularly steep negative slopes in the upper half of the eye, while Subjects 1 and 2 have more stable slopes across superpixels. The estimated residual SDs vary within subject by superpixel location. Subjects 1 and 4 have more uniform residual SDs across locations while Subjects 2 and 3 have some superpixels with much higher residual SDs.

(top) and residual SDs (bottom) from SLR of GCC thickness on time since baseline in each superpixel for the same four subjects as in Figure 3.3, where each column is a different subject. Estimated slopes and residual SDs appear spatially correlated.

Bryan et al. (2015) model errors that affect all locations at a visit in glaucomatous VFs as global visit effects. Similar to VF data, we suspect there are spatially correlated errors in GCC measurements. We speculate these effects arise from the imaging process and segmentation errors that affect multiple locations. To better visualize these effects, we plot empirical residuals $y_{ijk} - \bar{y}_{ik}$, where $\bar{y}_{ik} = \sum y_{ijk}/J_i$. Empirical residual profile plots allow us to better see time trends within and across superpixels. Figure 3.5 provides an example of correlated errors across superpixels, where there is a noticeable increase at four years of follow-up. It is unlikely that such an increase is due to thickening of GCC, but rather due to errors in the imaging process or layer segmentation. Figure 3.5 shows spatially correlated slopes noticeable in the region from superpixels 3.4 to 3.7 down to 6.4 to 6.7.

3.2.4 Modeling goals

We are interested in estimating individual rates of change at the superpixel level and predicting future GCC observations. To this end, we explicitly model the correlations between intercepts, slopes, and residual SDs at both the population and subject level. The intercepts are correlated with the magnitude of the slopes; as the baseline thickness increases, rates of change are faster (Rabiolo et al., 2020). Healthier eyes tend to have more thickness at baseline, with more potential for progression but also more opportunities for clinicians to intervene and prevent vision loss. Accounting for the relationships between measurement variability and either baseline thickness or slopes may help to better estimate the rates of progression and elucidate whether increased noise is associated with worsening disease. As glaucoma progresses, the ganglion cell and inner plexiform layers, two sublayers of GCC, show increased measurement variability especially as measures tend towards their floor (Miraftabi et al., 2016).

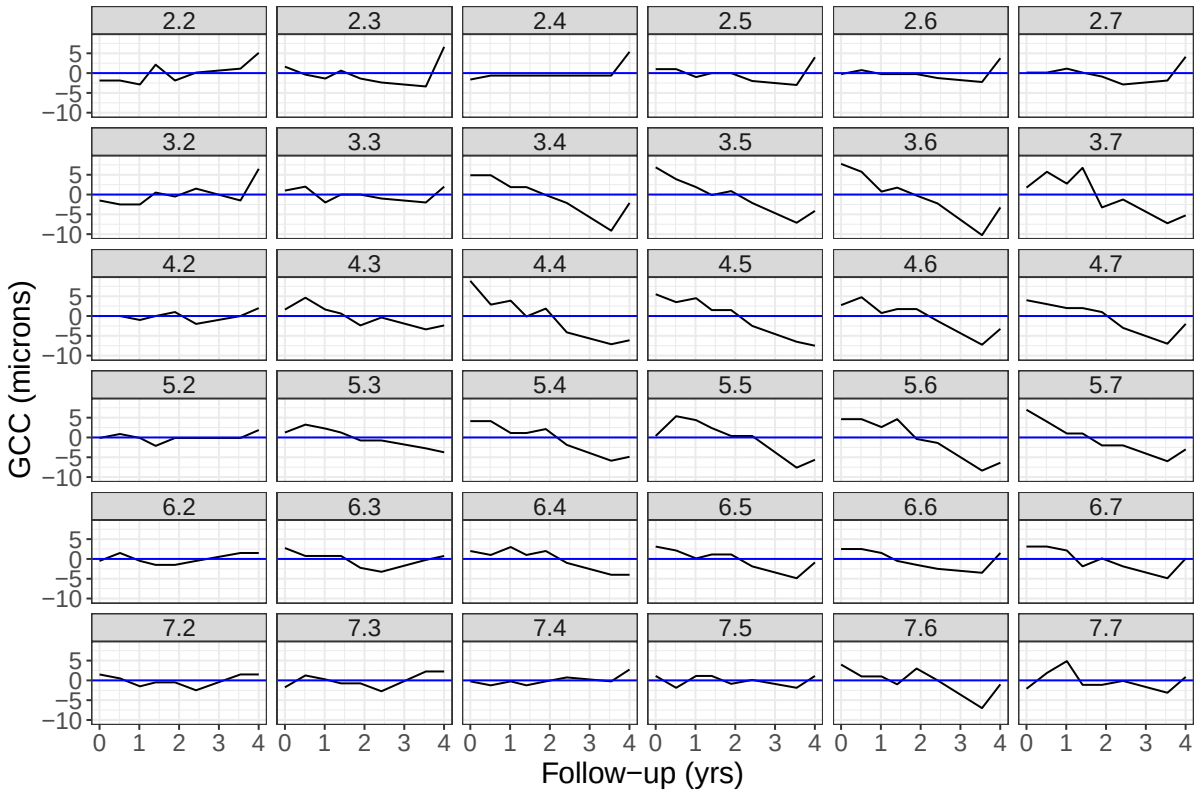

Figure 3.5: Empirical residual profile plots (superpixel mean subtracted from ganglion cell complex (GCC) thickness) for a single subject across 36 superpixels. There is an increase at four years for many superpixel locations suggesting visit-specific spatially correlated errors.

3.3 Methods

This section reviews the MGP priors we use to model the spatially varying visit effects and coefficients, constructs the SHREVE model, defines the priors, and introduces model comparison metrics.

3.3.1 Gaussian processes

A Gaussian spatial process (Williams and Rasmussen, 2006; Bogachev, 1998; Banerjee et al., 2014) is a stochastic process $\{z(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^d\}$ in which any finite collection of real-valued random variables $\{z(\mathbf{s}_1), \ldots, z(\mathbf{s}_K)\}$ is distributed as multivariate normal for every set of $K \geq 1$ spatial locations $\mathbf{s}_1, \ldots, \mathbf{s}_K \in \mathbb{R}^d$, for dimension $d \geq 1$; we work only with $d = 2$. We denote a GP as

$$z(\mathbf{s}) \sim \text{GP}(m(\mathbf{s}), C(\mathbf{s}, \mathbf{s}')),$$

with mean function $m(\mathbf{s}) = \mathbb{E}[z(\mathbf{s})]$ and covariance function $C(\mathbf{s}, \mathbf{s}') = \text{cov}[z(\mathbf{s}), z(\mathbf{s}')]$ for two locations $\mathbf{s}$ and $\mathbf{s}'$, which may be the same or distinct. The covariance function $C(\mathbf{s}, \mathbf{s}')$ models how similar outcomes $z(\mathbf{s})$ and $z(\mathbf{s}')$ are. We assume stationary and isotropic covariance functions $C(\mathbf{s}, \mathbf{s}')$. Stationarity means $C(\mathbf{s}, \mathbf{s}')$ depends only on the spatial separation vector $\mathbf{s} - \mathbf{s}'$ between points, and isotropy means $C(\mathbf{s}, \mathbf{s}')$ depends only on the distance between locations $h = \|\mathbf{s} - \mathbf{s}'\|$, where $\|\cdot\|$ is the Euclidean norm, i.e., $C(\mathbf{s}, \mathbf{s}') \equiv C(h)$.

We use Matérn covariance functions of the form $\sigma^2 M(h|\nu, \ell)$, where $\sigma^2 > 0$ is the variance and $M(h|\nu, \ell)$ is the Matérn correlation function (Matern, 1986)

$$M(h|\nu, \ell) = \frac{2^{1-\nu}}{\Gamma(\nu)} (\sqrt{2\nu} h/\ell)^\nu K_\nu (\sqrt{2\nu} h/\ell),$$

where $\nu > 0$ is the smoothness parameter, $\ell > 0$ is the lengthscale, and K_ν is the modified Bessel function of the second kind of order ν (Abramowitz and Stegun, 1964). In general, the process is m times mean square differentiable if and only if $\nu > m$ (Williams and Rasmussen, 2006). The lengthscale parameter ℓ controls how quickly the correlation decays as a function of distance with

larger ℓ indicating slower correlation decay.

3.3.2 Multivariate Gaussian processes

Let $\mathbf{z}(\mathbf{s}) = (z_1(\mathbf{s}), \ldots, z_P(\mathbf{s}))^T$ be a $P \times 1$ stochastic process where each component $z_p(\mathbf{s})$ for $p = 1, \ldots, P$ is a scalar random variable at location $\mathbf{s}$. Then $\mathbf{z}(\mathbf{s})$ is an MGP if any random vector $(\mathbf{z}(\mathbf{s}_1)^T, \ldots, \mathbf{z}(\mathbf{s}_K)^T)^T$ from any set of $K \geq 1$ locations $\mathbf{s}_1, \ldots, \mathbf{s}_K$ has a multivariate normal distribution. The MGP is an extension of the univariate GP where the random variables $\mathbf{z}(\mathbf{s})$ are vector-valued. We denote an MGP as

$$\mathbf{z}(\mathbf{s}) \sim \text{MGP}(\mathbf{m}(\mathbf{s}), \mathbf{C}(\mathbf{s}, \mathbf{s}')),$$

with $P \times 1$ mean vector $\mathbf{m}(\mathbf{s})$ and $P \times P$ cross-covariance matrix function $\mathbf{C}(\mathbf{s}, \mathbf{s}') = \text{cov}[\mathbf{z}(\mathbf{s}), \mathbf{z}(\mathbf{s}')] = \{C_{pq}(\mathbf{s}, \mathbf{s}')\}_{p,q=1}^{P}$. Functions $C_{pq}(\mathbf{s}, \mathbf{s}') = \text{cov}[z_p(\mathbf{s}), z_q(\mathbf{s}')]$, for $p, q = 1, \ldots, P$, are called marginal covariance functions when $p = q$ and cross-covariance functions when $p \neq q$.

We want to allow each marginal process to have its own spatial correlation function. Each marginal covariance function C_{pp} is modeled with a Matérn correlation function, $C_{pp}(h) = \sigma_{pp}^2 M(h|\nu_{pp}, \ell_{pp})$, for $p = 1, \ldots, P$, with variance parameter $\sigma_{pp}^2 > 0$, smoothness parameter ν_{pp}, and lengthscale parameter ℓ_{pp}. We model each cross-covariance function C_{pq} with a Matérn correlation function, $C_{pq}(h) = \sigma_{pq} M(h|\nu_{pq}, \ell_{pq})$, for $1 \leq p \neq q \leq P$, with covariance parameter σ_{pq}, smoothness parameter ν_{pq}, and lengthscale parameter ℓ_{pq}. We assume marginal covariance C_{pp} and cross-covariance C_{pq} functions to be Matérn following sufficient conditions on parameters ν_{pp}, ν_{pq}, ℓ_p, ℓ_{pq}, σ_{pp}, and σ_{pq} that result in a nonnegative definite cross-covariance function (Apanasovich et al., 2012). We use the simplest parameterization, where no additional parameters beyond σ_{pp}^2, ν_{pp}, and ℓ_{pp} are required to model the smoothness and lengthscale parameters for the cross-covariances.

The cross-covariance function $\mathbf{C}(\mathbf{s},\mathbf{s}')$ is nonnegative definite when

$$\begin{aligned}\nu_{pq}(\nu_{pp},\nu_{qq}) &= \frac{\nu_{pp}+\nu_{qq}}{2},\\ \ell_{pq}(\ell_p,\ell_q) &= \sqrt{\frac{2}{\ell_p^{-2}+\ell_q^{-2}}},\end{aligned} \tag{3.1}$$

$$\sigma_{pq}(\nu_{pp},\nu_{qq},\ell_p,\ell_q,\sigma_{pp},\sigma_{qq},R_{pq}) = \sigma_{pp}\sigma_{qq}\frac{\ell_{pq}(\ell_p,\ell_q)}{\sqrt{\ell_p\ell_q}}\frac{\Gamma(\nu_{pq}(\nu_{pp},\nu_{qq}))}{\Gamma^{1/2}(\nu_{pp})+\Gamma^{1/2}(\nu_{qq})}R_{pq}, \tag{3.2}$$

where $\mathbf{R}=\{R_{pq}\}$ is a nonnegative definite $P\times P$ correlation matrix with diagonal elements equal to 1 and nondiagonal elements in the closed interval $[-1,\,1]$. The cross-correlation $\rho_{pq}=\sigma_{pq}/\sigma_{pp}\sigma_{qq}=$ $\text{corr}(z_p(\mathbf{s}),z_q(\mathbf{s}))$ is the correlation between $z_p(\mathbf{s})$ and $z_q(\mathbf{s})$.

3.3.3 Model specification for a spatially varying hierarchical random effects with visit effects model

The proposed SHREVE model allows random intercepts, slopes, and log residual SDs to be correlated within and across locations while accounting for within-subject variability and spatially correlated visit-specific errors. For ease of notation, we specify the model assuming no missing data but note that complete data is not a requirement. We model y_{ijk} as

$$\begin{aligned}y_{ijk} &= \alpha_{0k}+\alpha_{1k}t_{ij}+\beta_{0ik}+\beta_{1ik}t_{ij}+\gamma_{ijk}+\varepsilon_{ijk}\\ \varepsilon_{ijk}|\tau_{ik}^2 &\sim \text{N}(0,\tau_{ik}^2),\\ \log\tau_{ik} &= \phi_k+\sigma_{ik},\end{aligned}$$

where α_{0k}, α_{1k}, and ϕ_k are the superpixel k population-level intercept, slope, and log residual SD processes, respectively, β_{0ik}, β_{1ik}, and σ_{ik} are subject-specific intercept, slope, and log residual SD processes, respectively, in superpixel k and γ_{ijk} is the visit effect process at location $\mathbf{s}_k$ for subject i visit j. Figure 3.6 presents the model graphically.

Let $\boldsymbol{\alpha}_k=(\alpha_{0k},\alpha_{1k},\phi_k)^T$ denote the population-level (PL) multivariate spatial process, which

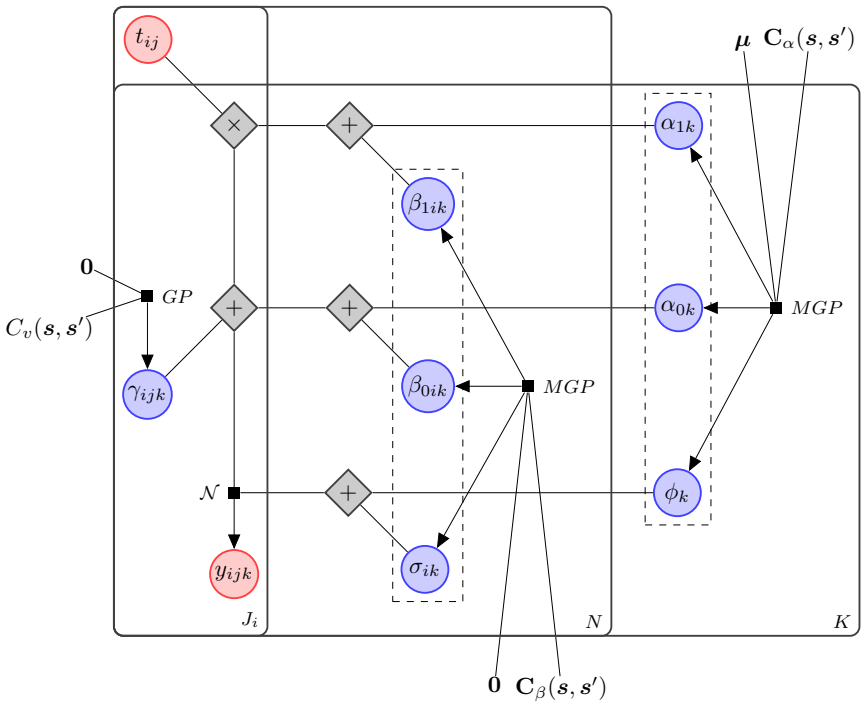

Figure 3.6: Plate diagram of the proposed model. Blue nodes are latent variables, red nodes are observed variables, gray nodes are deterministic nodes, GP stands for *Gaussian process*, and MGP stands for *multivariate Gaussian process*. Plates are used to group variables repeated together over subjects, time, and space, where $i = 1,\ldots,N$ indexes subjects, $j = 1,\ldots,J_i$ indexes subject i's visits, and $k = 1,\ldots,K$ indexes superpixel locations.

we model with MGP $\boldsymbol{\alpha}_k|\boldsymbol{\mu},\boldsymbol{\theta}_\alpha \sim \text{MGP}(\boldsymbol{\mu},\mathbf{C}_\alpha(\mathbf{s}_k,\mathbf{s}_{k'}))$, with mean vector $\boldsymbol{\mu} = (\mu_0,\mu_1,\mu_\phi)^T$ and PL cross-covariance matrix function $\mathbf{C}_\alpha(\mathbf{s}_k,\mathbf{s}_{k'})$ with hyperparameters $\boldsymbol{\theta}_\alpha = \{\sigma_{\alpha,pp},\nu_{\alpha,p},\ell_{\alpha,p},\mathbf{R}_\alpha, p \in \{1,2,3\}\}$. The parameters μ_0, μ_1, and μ_ϕ are the global grand mean intercept, slope, and log residual SD, respectively. PL marginal covariance functions $C_{\alpha,pp}(\mathbf{s}_k,\mathbf{s}_{k'}) = \sigma^2_{\alpha,pp}M(h|\nu_{\alpha,p},\ell_{\alpha,p})$, for $p = 1,...,3$ have PL marginal variances $\sigma^2_{\alpha,pp}$, PL smoothness parameters $\nu_{\alpha,p}$, and PL length-scales $\ell_{\alpha,p}$. PL cross covariance functions $C_{\alpha,pq}(\mathbf{s}_k,\mathbf{s}_{k'}) = \sigma_{\alpha,pq}M(h|\nu_{\alpha,pq},\ell_{\alpha,pq})$ have covariance parameters $\sigma_{\alpha,pq}$ between processes p and q, smoothness parameters $\nu_{\alpha,pq}$, and length-scales $\ell_{\alpha,pq}$. Here $h = \|\mathbf{s}_k - \mathbf{s}_{k'}\|$ is the distance between two superpixel locations, $\sigma_{\alpha,pq} \equiv \sigma_{pq}(\nu_{\alpha,p},\nu_{\alpha,q},\ell_{\alpha,p},\ell_{\alpha,q},\sigma_{\alpha,pp},\sigma_{\alpha,qq},R_{\alpha,pq})$ is a function of $\sigma_{\alpha,pp}$ and $\sigma_{\alpha,qq}$ as defined in (3.2), and $\ell_{\alpha,pq} \equiv \ell_{pq}(\ell_{\alpha,p},\ell_{\alpha,q})$ is a function of $\ell_{\alpha,p}$ and $\ell_{\alpha,q}$ as in (3.1). The 3×3 cross-correlation matrix $\mathbf{R}_\alpha$ is an unknown symmetric matrix with 1's on the diagonal and with (p,q)th element the

correlation parameter $R_{\alpha,pq} = R_{\alpha,qp}$.

Similarly, we model random effects (RE) $\boldsymbol{\beta}_{ik} = (\beta_{0ik}, \beta_{1ik}, \sigma_{ik})^T$ as $\boldsymbol{\beta}_{ik}|\boldsymbol{\theta}_\beta \sim \text{MGP}(\mathbf{0}, \mathbf{C}_\beta(\mathbf{s}_k, \mathbf{s}_{k'}))$, with mean vector $\mathbf{0}$ and cross-covariance matrix function $\mathbf{C}_\beta(\mathbf{s}_k, \mathbf{s}_{k'})$ with hyperparameters $\boldsymbol{\theta}_\beta = \{\sigma_{\beta,pp}, \nu_{\beta,p}, \ell_{\beta,p}, \mathbf{R}_\beta, p \in \{1,2,3\}\}$. RE marginal covariance functions $C_{\beta,pp}(\mathbf{s}_k, \mathbf{s}_{k'}) = \sigma^2_{\beta,pp} M(h|\nu_{\beta,p}, \ell_{\beta,p})$ for $p = 1,...,3$ have RE marginal variances $\sigma^2_{\beta,pp}$, smoothness parameters $\nu_{\beta,p}$, and lengthscales $\ell_{\beta,p}$. RE cross-covariance functions $C_{\beta,pq}(\mathbf{s}_k, \mathbf{s}_{k'}) = \sigma_{\beta,pq} M(\mathbf{h}|\nu_{\beta,pq}, \ell_{\beta,pq})$ have RE covariance parameters $\sigma_{\beta,pq} \equiv \sigma_{pq}(\nu_{\beta,p}, \nu_{\beta,q}, \ell_{\beta,p}, \ell_{\beta,q}, \sigma_{\beta,pp}, \sigma_{\beta,qq}, R_{\beta,pq})$, lengthscales $\ell_{\beta,pq} \equiv \ell_{pq}(\ell_{\beta,p}, \ell_{\beta,q})$, and unknown cross-correlation matrix $\mathbf{R}_\beta$ as defined in (3.1) and (3.2). We model the spatially varying visit effects γ_{ijk} with mean 0 GPs $\gamma_{ijk}|\sigma_v, \nu_v, \ell_v \sim \text{GP}(0, C_v(\mathbf{s}_k, \mathbf{s}_{k'}))$, with visit effects covariance function $C_v(\mathbf{s}_k, \mathbf{s}_{k'}) = \sigma^2_v M(\mathbf{h}|\nu_v, \ell_v)$.

3.3.4 Priors

We use weakly informative priors to keep inferences within a reasonable range and allow computations to proceed satisfactorily. The closest two superpixels can be is 1 unit apart, and the largest separation is $\sqrt{(7-2)^2 + (7-2)^2} \approx 7$ units. We expect lengthscales to plausibly fall in this range. At the same time, we wish to avoid infinitesimal lengthscales. We assign independent and identical inverse gamma priors on all MGP lengthscale parameters $\ell_{\alpha,1}$, $\ell_{\alpha,2}$, $\ell_{\alpha,3}$, $\ell_{\beta,1}$, $\ell_{\beta,2}$, $\ell_{\beta,3}$, $\ell_v \sim IG(2.25, 2.5)$ with mean 2 and SD 4. For the MGP SD parameters, we wish to avoid flat priors that could pull the posterior towards extreme values. We assign truncated-normal priors on all MGP SD parameters $\sigma_{\alpha,11}, \sigma_{\beta,11} \sim N^+(0, 10^2)$, $\sigma_{\alpha,22}, \sigma_{\alpha,33}, \sigma_{\beta,22}, \sigma_{\beta,33}, \sigma_v \sim N^+(0, 2.5^2)$, where $N^+(a,b)$ is a normal distribution restricted to the positive real line with mean a and variance b. We assign independent normal priors on the global effects $\mu_0 \sim N(73, 15^2)$, $\mu_1 \sim N(-0.3, 0.3^2)$, $\mu_\phi \sim N(0.7, 0.3^2)$. For the correlation matrices $\mathbf{R}_\alpha$ and $\mathbf{R}_\beta$, we assign marginally uniform priors on the individual correlations derived from the inverse Wishart distribution with 3×3 identity matrix scale matrix parameter and four degrees of freedom $IW(\mathbf{I}_3, 4)$ (Barnard et al., 2000). When $\boldsymbol{\Sigma}$ has a standard inverse-Wishart distribution, we can decompose $\boldsymbol{\Sigma} = \mathbf{SRS}$ in terms of the diagonal standard deviation matrix $\mathbf{S}$ and

correlation matrix $\mathbf{R}$ to obtain the prior for the correlation matrices. We set all MGP smoothness parameters $\nu_{\alpha,1}$, $\nu_{\alpha,2}$, $\nu_{\alpha,3}$, $\nu_{\beta,1}$, $\nu_{\beta,2}$, $\nu_{\beta,3}$, $\nu_{\nu} = \frac{1}{2}$ since we obtain measurements from a coarse grid of superpixel locations and expect the processes to be rough. When $\nu = \frac{1}{2}$, the Matern correlation function reduces to the popular exponential kernel $M(\mathbf{h}|\frac{1}{2},\ell) = \exp(-\|\mathbf{h}\|/\ell)$.

3.3.5 Computation and inference

For data analysis and visualization, we use the R programming language (R Core Team, 2021) and GGPLOT2 (Wickham, 2016). We use Markov Chain Monte Carlo (MCMC) methods (Metropolis et al., 1953; Robert and Casella, 2005) implemented in NIMBLE v0.13.0 (de Valpine et al., 2017, 2022). We specify the model at the observation level and omit observations removed in the data cleaning step. To sample from the posteriors, we use Gibbs sampling and update specific parameters using the automated factor slice sampler or Metropolis-Hastings sampler within Gibbs. We update the global effects μ_0, μ_1 and μ_ϕ using scalar Metropolis-Hastings random walk samplers; the visit effect GP lengthscale ℓ_ν and subject-level residual SD GP SD parameter $\sigma_{\beta,33}$ together using the automated factor slice sampler (Tibbits et al., 2014); the subject-level random effects β_{0ik}, β_{1ik}, and σ_{ik} and visit effects γ_{ijk} using multivariate Metropolis-Hastings random walk samplers in sub-blocks. We tested various schemes for sampling sub-blocks of the subject-level random effects and visit effects to improve sampling efficiency (Risser and Turek, 2020). We jointly sample subject-level intercepts, slopes, and the first visit effect in sub-blocks of size 3. We separately sample the subject-level residual SDs in sub-blocks of size 6 and the remaining visit effects in sub-blocks of size 3. Each pair of SD and lengthscale parameters from MGPs and GPs were sampled together (e.g., $(\sigma_{\alpha,11},\ell_{\alpha,1})$) except for the subject-level residual SDs and visit effects where opposites were paired together $(\sigma_{\beta,33},\ell_\nu)$ and $(\sigma_\nu,\ell_{\beta,3})$. We run all models with 9 chains of 250,000 iterations after a burn-in of 30,000, a thin of 100 for a total of 19,800 posterior samples. Following Vehtari et al. (2021)'s recommendation for assessing convergence, the bulk and tail effective sample sizes were all greater than 100 per chain and the potential scale reduction factor $\widehat{R}$ were all less than 1.01. Visual assessment of model convergence show satisfactory results. We show efficiency per

iteration plots of the 7 parameters with the largest $\widehat{R}$ in Supplementary Figure 3.11 and summarize convergence diagnostics in Supplementary Table 3.3. The NIMBLE model code for the SHREVE model is available as an R script at https://github.com/erica-su/SHREVE.

3.3.6 Model comparison

We fit the SHREVE model to the AGPS data and compare model fit of the SHREVE model to 7 nested models and to SLR fit separately for each subject and superpixel location. The 7 submodels were SHREVE omitting (a) the population-level residual SD process ϕ_k, (b) the subject-specific residual SD process σ_{ik}, (c) the spatially varying visit effects γ_{ijk}, and all combinations (ab), (ac), (bc), and (abc). We call the SHREVE model without visit effects the spatially varying hierarchical random effects (SHRE) model. For SLR, we run a separate model for each eye and superpixel using flat priors with results equivalent to classical least squares.

We compare models with the Watanabe-Akaike (or widely applicable) information criterion (WAIC) (Watanabe and Opper, 2010; Gelman et al., 2013) and approximate leave-one-out cross-validation (LOO) using Pareto Smoothed Importance Sampling (Vehtari et al., 2017). We report WAIC

$$\text{WAIC} = -2\left[\sum_{i=1}^{n}\sum_{j=1}^{J_i}\sum_{k=1}^{K}\log\left(\frac{1}{S}\sum_{s=1}^{S}p(y_{ijk}|\theta^s)\right) - \sum_{i=1}^{n}\sum_{j=1}^{J_i}\sum_{k=1}^{K}V_{s=1}^{S}\left(\log p(y_{ijk}|\theta^s)\right)\right]$$

summing over all data points y_{ijk}, where $p(y_{ijk}|\theta)$ is the pointwise predictive density, θ are the model parameters, superscript s denotes parameters drawn at the sth iteration for $s = 1,\ldots,S$ posterior samples, and $V_{s=1}^{S}$ denotes the sample variance over S posterior samples. We report approximate LOO

$$\text{LOO} = -2\sum_{i=1}^{n}\sum_{j=1}^{J_i}\sum_{k=1}^{K}\log\left(\frac{\sum_{s=1}w_{ijk}^{s}p(y_{ijk}|\theta^s)}{\sum_{s=1}w_{ijk}^{s}}\right)$$

where w_{ijk}^{s}, $s = 1,\ldots,S$ is a vector of importance weights for data point y_{ijk} at iteration s and $w_{ijk}^{s} = \left(p(y_{ijk}|\theta^s)\right)^{-1}$ except for extreme weights. Approximate LOO estimates the out-of-sample

predictive accuracy of the model (Stone, 1977). Lower WAIC and LOO indicate better fit.

To assess predictive accuracy of the proposed model, we compare models on mean squared prediction error

$$\text{MSPE} = \frac{\sum_{s=1}^{S}\sum_{i=1}^{n}\sum_{k\in\mathcal{K}_i}(y_{iJ_ik} - \hat{y}^s_{iJ_ik})^2}{SN_{pred}}$$

for $s = 1,\ldots,S$ posterior MCMC samples, $i = 1,\ldots,n$ subjects, $k \in \mathcal{K}_i$ held out superpixels for subject i, held out observations y_{iJ_ik}, and predicted observations for each posterior sample $\hat{y}^s_{iJ_ik}$, of N_{pred} total held out observations after fitting the models. We randomly sample and hold out 7 observations y_{iJ_ik}, or approximately 20%, at the last visit for each of 110 subjects and 6 observations for one subject because they only had 32 observations available at the last visit, for a total of $N_{pred} = 111 \times 7 - 1 = 776$ observations, and fit models with the remaining observations. Not all observations are available at all superpixels because we remove some observations in the data cleaning step. For the SHREVE models, we define a predicted observation at each posterior sample s as

$$\hat{y}^s_{iJ_ik} = \alpha^s_{0k} + \alpha^s_{1k}t_{iJ_i} + \beta^s_{0ik} + \beta^s_{1ik}t_{iJ_i} + \gamma_{iJ_ik}, \tag{3.3}$$

where t_{iJ_i} is the time observed and γ_{iJ_ik} is the visit effect for the held out observation at the ith subject's last visit. For the SHRE models, there is no γ_{iJ_ik} visit effect term in (3.3).

3.4 Advanced Glaucoma Progression Study

After identifying and removing approximately 0.5% of the data as outliers, we analyze 29,179 observations from 111 subjects over 36 superpixels. Table 3.1 gives the WAIC, LOO, and MSPE of models considered. The SHREVE model has the lowest WAIC and LOO. Comparing pairs of SHREVE and SHRE models with and without the (a) population-level residual SD process and (b) subject-level residual SD process, omitting (a) increases WAIC (LOO) by up to 421 (238) while omitting (b) increases WAIC (LOO) by up to 4,964 (4,573). Omitting visit effects increases WAIC (LOO) by up to 18,361 (12,899). SLR has lower WAIC than the two SHRE models without (b),

Table 3.1: Model fit comparison with widely applicable information criterion (WAIC), approximate leave-one-out cross-validation with Pareto Smoothed Importance Sampling (LOO), and mean squared prediction error (MSPE) of predictions. For predictions, we hold out 7 randomly sampled observations y_{iJ_ik} at the last visit of each of 110 AGPS subjects and 6 observations from one subject. Models with visit effects perform better than models without visit effects. SLR performs noticeably worse compared to the hierarchical models. The smallest WAIC, LOO, and MSPE values are bolded.

Model	**Joint Model**	**Visit Effects**	**Superpixel Residual SD**	**Subject Residual SD**	**WAIC**	**LOO**	**MSPE** (μm^2)
SHREVE	✓	✓	✓	✓	**107,581.6**	**113,323.1**	6.6
SHREVE-(a)	✓	✓	✗	✓	108,002.2	113,560.7	**6.5**
SHREVE-(b)	✓	✓	✓	✗	110,992.3	116,978.1	6.8
SHREVE-(ab)	✓	✓	✗	✗	113,238.3	118,647.5	6.9
SHRE	✓	✗	✓	✓	124,389.5	125,304.7	7.2
SHRE-(a)	✓	✗	✗	✓	124,468.8	125,461.2	7.1
SHRE-(b)	✓	✗	✓	✗	129,353.2	129,877.3	7.5
SHRE-(ab)	✓	✗	✗	✗	130,188.4	130,732.1	7.5
SLR	✗	✗	✗	✓	128,870.2	132,916.3	39.7

but SLR still has higher LOO. Having subject-specific residual SDs is more important for models without a visit effect component, as the difference in WAIC (LOO) between SHRE and SHRE-(b) is larger by 1553 (918) than the difference between SHREVE and SHREVE-(b). For predictions, the MSPE for SLR is 6.0 times that of the SHREVE model (39.7 vs. 6.6 μm^2) and 5.5 times that of the SHRE model (39.7 vs. 7.2 μm^2). Among the hierarchical models, the biggest distinction in MSPE is between models with and without visit effects. Comparing pairs of SHREVE and SHRE models, omitting the subject-level residual SD process consistently increases the MSPE, while omitting the population-level residual SD process has a negligible effect on MSPE. Figure 3.7 plots profiles and posterior mean fitted lines from the SHREVE model and SLR for one subject for superpixels that had the last (7th) observation held out. The SHREVE model better estimates slopes for noisy superpixels like 2.3 and 5.6. All predictions of the last visit in the 6 superpixels by the SHREVE model are closer to the GCC observed at $t_{ij} = 3.6$ than those by SLR.

Table 3.2 gives posterior means and 95% credible intervals (CrI) for parameters of interest from

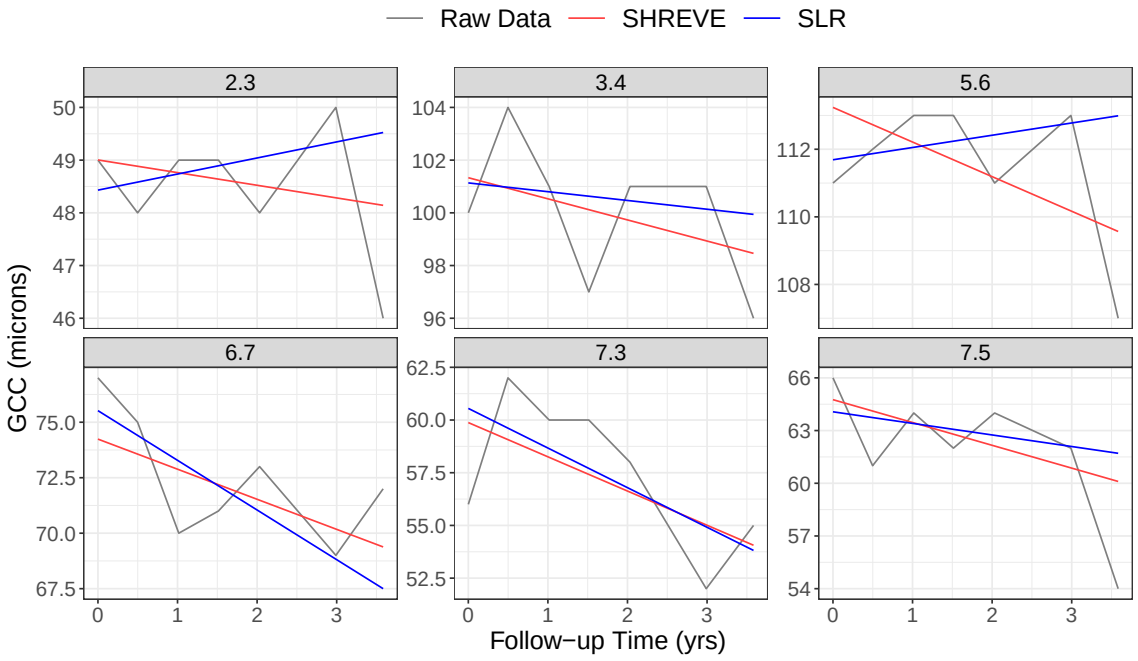

Figure 3.7: Comparison of predicted observations and model fit from the SHREVE model and simple linear regression (SLR) after holding out the last observation at 3.6 years follow-up of this subject. The gray line plots the raw data, the red line is the posterior mean fitted line from the SHREVE model without adding in the visit effects, and the blue line shows the fitted line from SLR. The SHREVE model is able to better estimate slopes and predict the last observation in noisy superpixels like 2.3 and 5.6 than SLR.

the SHREVE and SHRE models. The SHREVE global log residual SD parameter has a smaller posterior mean than SHRE (0.35 vs 0.66 μm), although CrIs overlap; global intercepts and slopes have similar posterior means and CrIs. The SHREVE subject-level slopes and log residual SDs MGP lengthscales are shorter than for the SHRE model, implying that the spatial correlation of subject-level slopes and log residual SDs decays faster after including visit effects, allowing random effects to vary more across the macula. The SHREVE subject-level MGP SD parameter is larger than from SHRE, meaning the variability of subject-specific residual SDs is higher within a superpixel for the SHREVE model. All other subject-level MGP parameters are similar between the models. Supplementary Table 3.4 gives posterior means and 95% CrIs for the population-level MGP parameters. The population-level MGP parameters are similar between the two models.

Figure 3.8 plots spatial correlations $M(h)$ as a function of distance h between superpixels for

Table 3.2: Posterior mean and 95% credible interval (CrI) for global parameters and subject-level multivariate Gaussian process (MGP) parameters comparing the SHREVE and SHRE models.

		SHREVE Model		SHRE Model	
Parameters	Symbols	Mean	95% CrI	Mean	95% CrI
Global Parameters					
Intercept	μ_0	70.02	(54.47, 84.21)	71.22	(56.83, 84.80)
Slope	μ_1	-0.30	$(-0.59, 0.02)$	-0.30	$(-0.60, 0.04)$
Log Residual SD	μ_ϕ	0.35	(0.05, 0.86)	0.66	(0.39, 0.97)
Subject-Level MGP SD Parameters					
Intercept	$\sigma_{\beta,11}$	16.17	(15.11, 17.39)	16.33	(15.24, 17.57)
Slope	$\sigma_{\beta,22}$	0.94	(0.87, 1.03)	1.00	(0.92, 1.09)
Log Residual SD	$\sigma_{\beta,33}$	0.45	(0.42, 0.49)	0.34	(0.32, 0.37)
Subject-Level MGP Lengthscale Parameters					
Intercept	$\ell_{\beta,1}$	5.42	(4.67, 6.32)	5.58	(4.80, 6.51)
Slope	$\ell_{\beta,2}$	4.20	(3.41, 5.16)	6.79	(5.48, 8.46)
Log Residual SDs	$\ell_{\beta,3}$	1.87	(1.57, 2.24)	3.71	(3.02, 4.61)
Subject-Level MGP Correlation Parameters					
Intercepts/Slopes	$\rho_{\beta,12}$	-0.14	$(-0.19, -0.10)$	-0.13	$(-0.18, -0.08)$
Intercepts/Log Residual SDs	$\rho_{\beta,13}$	0.12	(0.08, 0.16)	0.17	(0.11, 0.22)
Slopes/Log Residual SDs	$\rho_{\beta,23}$	-0.21	$(-0.28, -0.14)$	-0.24	$(-0.31, -0.17)$
Visit Effect Parameters					
Lengthscale	ℓ_v	3.54	(3.07, 4.10)		
SD	σ_v	1.42	(1.37, 1.48)		

the SHREVE and SHRE models. At 4.2 units distance, the spatial correlation of subject-specific slopes drops to $\exp(-1) \approx 0.37$ for the SHREVE model but is $\exp(-0.62) \approx 0.54$ for the SHRE model. At 1.9 units distance, the spatial correlation of subject-specific log residual SDs is 0.37 for the SHREVE model but around 0.60 for the SHRE model. The shorter lengthscales in the SHREVE model result in reduced correlation at similar distances between superpixels.

Figure 3.9 presents heat maps of the posterior means and SDs of the log residual SDs from the SHREVE and SHRE models. For most superpixels, the SHREVE model uniformly reduces log residual SDs by approximately 0.5 compared to the SHRE model. The four central superpixels (4.4,

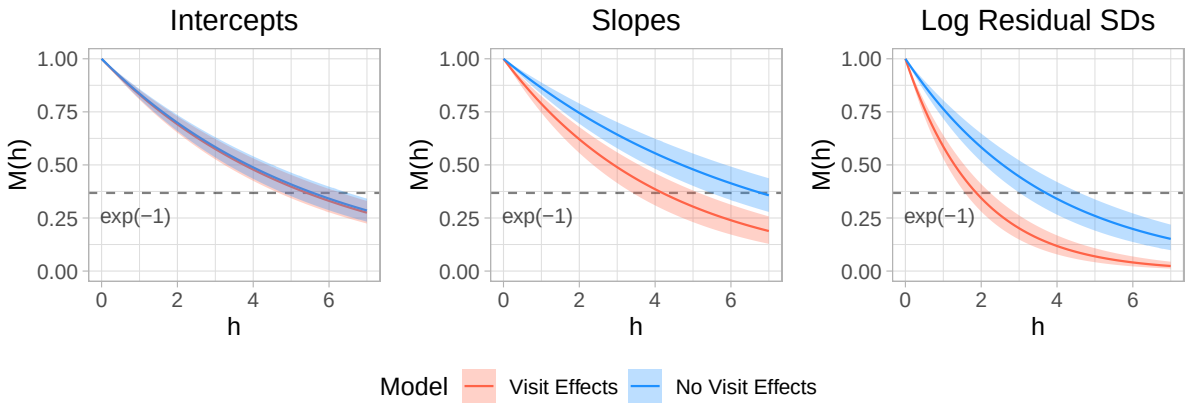

Figure 3.8: Posterior mean (line) and 95% pointwise credible intervals (colored bands) of correlation as a function of distance h between superpixels for subject-specific intercepts, slopes, and log residual SDs from the SHREVE (Visit Effects) and SHRE (No Visit Effects) models. The correlations decay faster in the SHREVE model with shorter lengthscales for slopes and log residual SDs. The dashed line indicates where the correlation is $\exp(-1)$ and the distance between superpixels is equal to the lengthscale in the exponential kernel.

4.5, 5.4, and 5.5) and superpixels in the 7th column have higher log residual SDs and have smaller differences in log residual SDs between the models. SHREVE breaks down measurement error into two components, spatially correlated errors due to the imaging process and general measurement noise. By accounting for visit effects, we reduce residual variance, leading to substantial improvement in model fit.

We compare subject-specific slopes estimated from the SHREVE model to those estimated using SLR. We declare a slope to be significantly negative or positive when the upper bound or lower bound of the 95% CrI is less than or greater than 0, respectively. Across the 3,990 subject-superpixel profiles, the SHREVE model detects a higher proportion of significant negative slopes (21.4% vs 18.0%) and lower proportion of significant positive slopes (3.1% vs 4.3%) as compared to SLR. Figure 3.10 shows the proportion of significant negative slopes by superpixel, and Supplementary Figure 3.12 shows the proportion of significant positive slopes by superpixel. The SHREVE model detects 10% more significant negative slopes in 6 of 36 superpixels and 5% less significant positive slopes in 5 of 36 superpixels. Because glaucoma is an irreversible disease, GCC thicknesses are not expected to increase over time. These findings indicate SHREVE is more sensitive in detecting

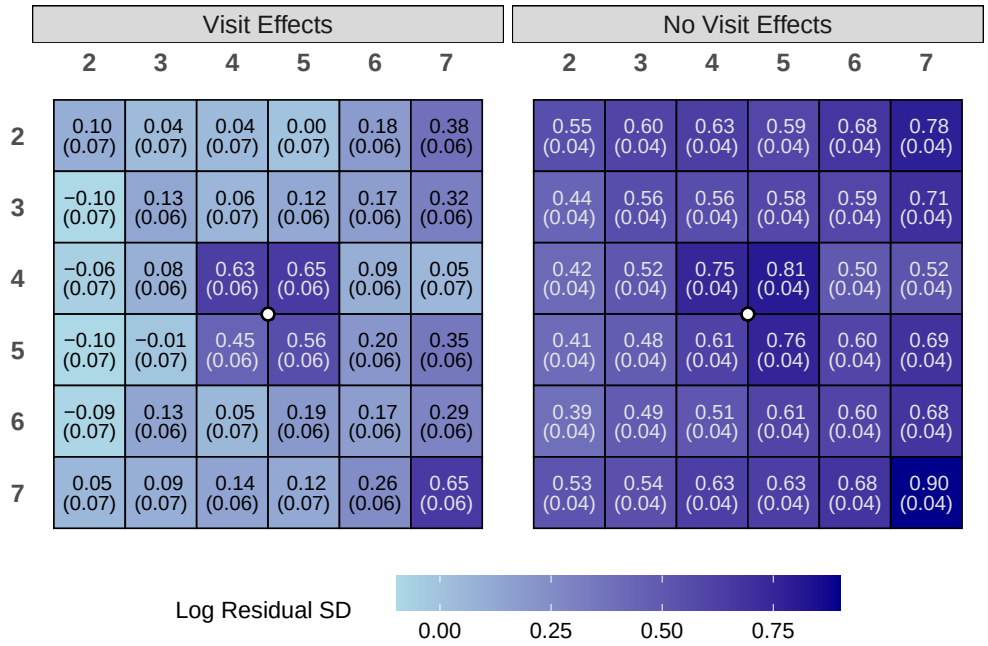

Figure 3.9: Heat map of the log residual standard deviations (SD) comparing the SHREVE (Visit Effects) and SHRE (No Visit Effects) models. The values shown are the posterior mean (posterior SD) across the 36 superpixels. The log residual SDs from the SHREVE model are uniformly reduced across all superpixels compared to those from the SHRE model. The white dot is the fovea.

worsening slopes and possibly reduces false positive rates as compared to SLR.

3.5 Discussion

We motivate and develop a Bayesian hierarchical model with population- and subject-level spatially varying coefficients and show that including visit effects reduces error in predicting future observations and greatly improves model fit. In current practice, ophthalmologists use SLR to assess slopes for individual subject-superpixel profiles, using information from only a single subject and location at a time. To better estimate subject-specific slopes, we include information from the whole cohort; explicitly model the correlations between subject-specific intercepts, slopes, and log residual SDs; allow population parameters and random effects to be spatially correlated; and account for visit-specific spatially correlated errors. Using information from the entire cohort, our proposed model

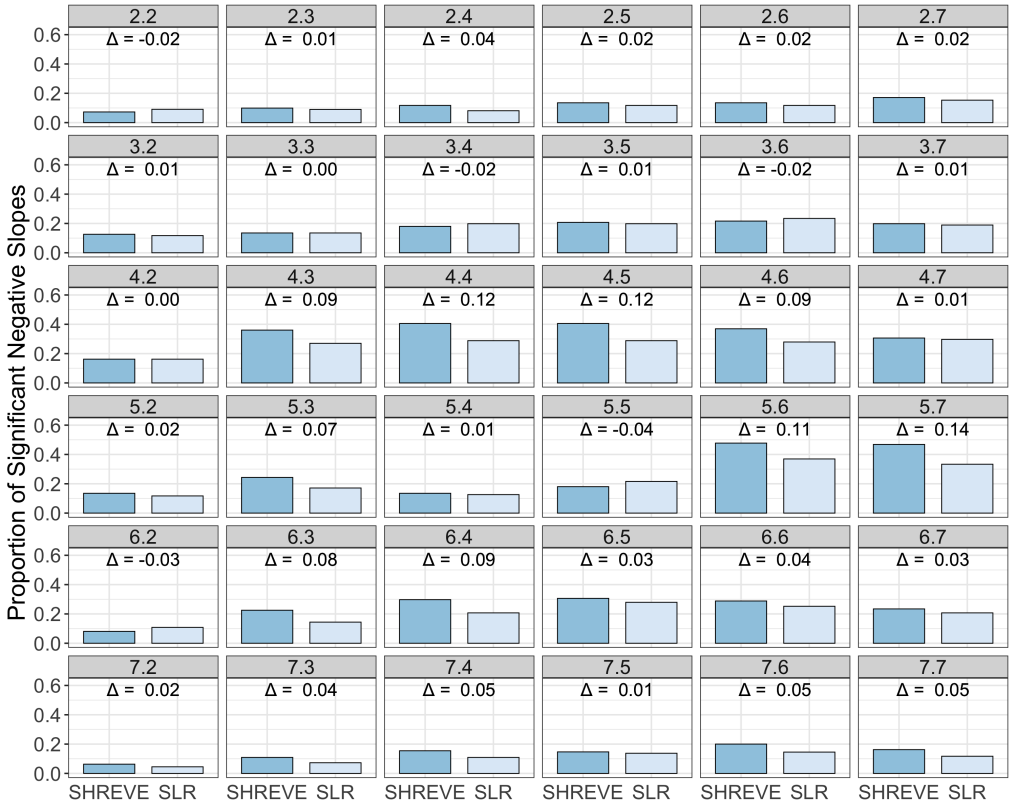

Figure 3.10: Bar charts of the proportion of significant negative slopes detected by the SHREVE model and simple linear regression (SLR) across the 36 superpixels. The difference ($\Delta =$ SHREVE $-$ SLR) in proportion is labeled at the top of each subplot. Across all locations, the SHREVE model detects a higher proportion of significant negative slopes (21.4% vs 18.0%) than SLR.

leads to decreased noise in estimating subject-specific slopes, having smaller posterior SDs in 79% of subject-superpixel slopes as compared to SLR.

There are many sources of error in obtaining the GCC thickness measurements from OCT scans. By separating measurement errors into visit-specific spatially correlated errors and other measurement noise, we are better able to detect eye-superpixels where GCC thicknesses are progressing most rapidly. Our approach will help identify progression of glaucoma for more individualized treatment plans.

Other methods for modeling spatial variation over discrete locations include CAR models, where random effect distributions are conditional on some neighboring values (Betz-Stablein et al., 2013;

Berchuck et al., 2019). Instead, we model spatial correlation between all locations with GPs, where the spatial correlation depends only on the distance between any two locations. In addition to our a priori specification of $\nu = \frac{1}{2}$, we fit our model using Matérn correlation functions with $\nu = \frac{3}{2}$, $\nu = \frac{5}{2}$, and $\nu = \infty$ (squared exponential kernel, Williams and Rasmussen 2006). These early exploratory analyses had difficulty in MCMC convergence. One limitation of using GPs is the increasing difficulty in fitting when the number of locations is large. Fitting GP models involves matrix inversion which increases computational complexity in cubic order with the number of locations. When the number of locations is too large, approximations for the processes could be considered (Banerjee et al., 2008). Nonetheless, we expect these model developments will benefit ophthalmologists as they seek to better estimate subject-specific slopes from structural thickness measurements.

We developed the current model specifically for GCC macular thickness measurements. Of further interest is to simultaneously model all the inner retinal layers that make up GCC to identify which sublayers may be worsening faster than others while accounting for between-layer correlations. Future extensions of the SHREVE model could include working with multivariate outcomes, which may pose additional computational challenges.

3.6 Supplementary material

3.6.1 Convergence assessment of the SHREVE model

We provide more details on convergence of the SHREVE model as mentioned in Section 3.3.5. Following Vehtari et al. (2021)'s recommendation on assessing convergence, we monitor the potential scale reduction factor $\widehat{R}$ and the bulk and tail effective sample sizes (ESS) for all model parameters. We obtain $\widehat{R}$ less than 1.01 and bulk and tail ESS all greater than 100 per chain for all parameters, indicating convergence. Figure 3.11 shows the efficiency per iteration of the bulk ESS and potential reduction factor $\widehat{R}$ of 7 model parameters with the largest $\widehat{R}$ in the SHREVE model. The bulk ESS increases linearly with iterations indicating that the relative efficiency is constant over different numbers of draws. $\widehat{R}$ decreases exponentially with increasing iterations and are all less than 1.01.

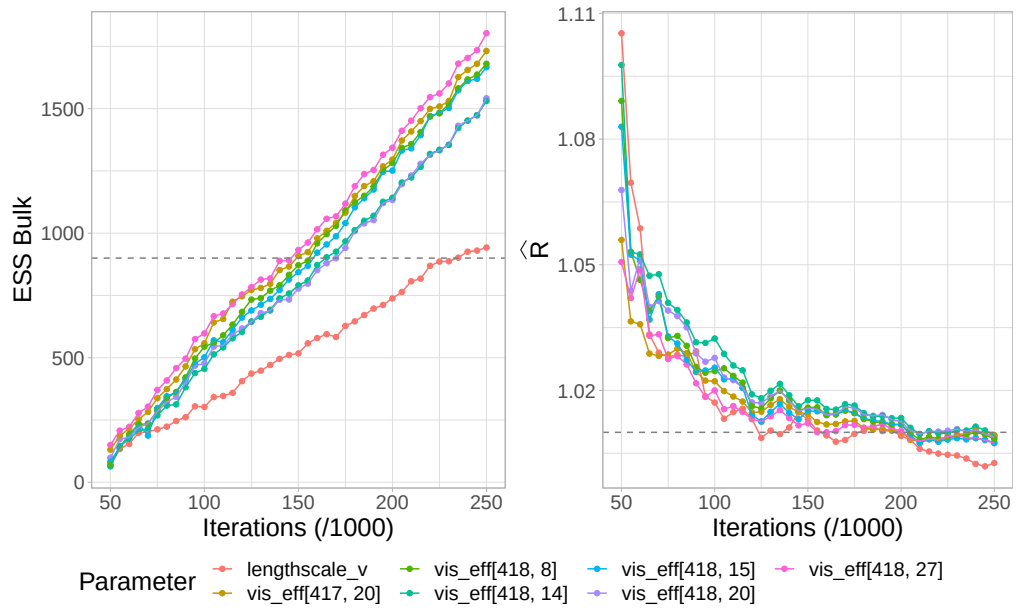

Figure 3.11: Plots of efficiency per iteration of the bulk effective sample size (ESS) on the left and potential scale reduction factor $\widehat{R}$ on the right for the 7 parameters with the largest $\widehat{R}$ from the SHREVE model. The bulk ESS increases linearly with increasing iterations while $\widehat{R}$ decreases exponentially with increasing iterations. The bulk ESSs were all greater than 100 per chain and $\widehat{R}$ were all less than 1.01.

Table 3.3 gives the sampling efficiency of all model parameters in terms of bulk and tail ESS and $\widehat{R}$.

3.6.2 Additional results for AGPS analysis

We provide additional results mentioned in Section 3.4. Table 3.4 presents the posterior mean and 95% CrIs for population-level MGP parameters for the SHREVE and SHRE models. All population-level MGP parameter posterior means and 95% CrIs are similar between the models. Figure 3.12 plots the proportions of significant positive slopes for the SHREVE model and SLR in each of the 36 superpixels. Across all locations, the SHREVE model detects a lower proportion of significant positive slopes (3.1% vs 4.3%) than SLR.

Table 3.3: The mean minimum/maximum bulk effective sample size (ESS), tail ESS, and potential scale reduction factor $\widehat{R}$ for the SHREVE model broken down by parameter types. For convergence, $\widehat{R}$ should be < 1.01 and ESS should be > 100 per chain. The first column gives the parameter types: Hyperparameters include global parameters and multivariate Gaussian process hyperparameters; Population-level includes the population-level intercepts, slopes, and log residual standard deviations (SD) across all locations; Intercepts, Slopes, and Log Residual SDs include the subject-level intercepts, slopes, and log residual SDs across all locations, respectively; and Visit Effects include the visit effects across all locations. The second column gives the number of parameters summarized.

		Bulk ESS		Tail ESS		$\widehat{R}$	
Parameter	#	Mean	Min	Mean	Min	Mean	Max
Hyperparameters	23	12545.8	942.3	14152.8	1951.3	1.001	1.003
Population-level	108	11519.7	3706.8	15610.8	7478.1	1.001	1.002
Intercepts	3990	4406.5	1663.8	9644.9	3585.9	1.002	1.007
Slopes	3990	3757.7	1602.4	8559.5	3597.6	1.002	1.007
Log Residual SDs	3990	16268.4	9311.0	18271.6	14123.2	1.000	1.002
Visit Effects	29179	4963.8	1505.0	10572.5	3475.9	1.002	1.009

Table 3.4: Posterior mean and 95% credible interval (CrI) for population-level multivariate Gaussian process (MGP) parameters for the SHREVE and SHRE models. SD stands for standard deviation.

		SHREVE Model		SHRE Model	
Parameters	Symbols	Mean	95% CrI	Mean	95% CrI
Population-Level MGP SD Parameters					
Intercept	$\sigma_{\alpha,11}$	13.68	(9.27, 21.13)	13.31	(9.06, 20.55)
Slope	$\sigma_{\alpha,22}$	0.31	(0.20, 0.54)	0.32	(0.21, 0.56)
Log Residual SD	$\sigma_{\alpha,33}$	0.36	(0.19, 0.79)	0.22	(0.11, 0.47)
Population-Level MGP Lengthscale Parameters					
Intercept	$\ell_{\alpha,1}$	3.56	(1.32, 8.83)	3.27	(1.20, 8.16)
Slope	$\ell_{\alpha,2}$	2.66	(0.88, 8.33)	2.82	(0.94, 8.90)
Log Residual SD	$\ell_{\alpha,3}$	4.68	(0.75, 19.68)	6.43	(0.98, 26.16)
Population-Level MGP Correlation Parameters					
Intercepts/Slopes	$\rho_{\alpha,12}$	-0.42	$(-0.68, -0.13)$	-0.42	$(-0.67, -0.12)$
Intercepts/Log Residual SDs	$\rho_{\alpha,13}$	-0.30	$(-0.57, -0.02)$	-0.28	$(-0.55, 0.01)$
Slopes/Log Residual SDs	$\rho_{\alpha,23}$	-0.11	$(-0.42, 0.20)$	-0.06	$(-0.37, 0.24)$

Figure 3.12: Bar charts of the proportion of significant positive slopes detected by the SHREVE model and simple linear regression (SLR) across the 36 superpixels. The difference ($\Delta = \text{SHREVE} - \text{SLR}$) in proportion is labeled at the top of each subplot. Across all locations, the SHREVE model detects a lower proportion of significant positive slopes (3.1% vs 4.3%) than SLR.

CHAPTER 4

A hierarchical spatially varying coefficients model for longitudinal visual field data in glaucoma subjects

4.1 Introduction

Glaucoma is one of the leading causes of blindness worldwide and projected to affect more than 110 million people by 2040 (Tham et al., 2014). Because vision loss due to glaucoma is irreversible, timely treatment is crucial to slow down disease progression and prevent further vision loss. Ophthalmologists assess glaucoma progression by monitoring decreases in visual acuity over time. Visual field (VF) testing evaluates functional changes in vision by measuring the sensitivity to presented light stimuli at multiple test locations within the VF. The Humphrey Field Analyzer (Carl Zeiss Meditec, Dublin, CA) is a commonly used tool for VF testing. The standard testing pattern for glaucoma is called the 24-2, which measures visual sensitivity at 52 locations, excepting two blind spots, with 6° of spacing between points. However, many studies show that the 24-2 pattern can miss damage in the central, macular area while using the 10-2 testing pattern is more likely to detect progression (De Moraes et al., 2017; Traynis et al., 2014; Wu et al., 2019). The 10-2 testing pattern measures visual sensitivity at 68 locations in the central 10° of the visual field.

During VF testing, the machine presents stimuli of varying intensity and the patient presses a button when the stimulus is perceived. The test continues until the sensitivity threshold is determined. Monitoring decreases in VF sensitivity thresholds over time provides a means to monitor VF loss due to glaucoma progression. VF threshold sensitivities are reported on the log scale in decibels (dB) of attenuation, ranging from 50 dB to 0 dB, with 50 dB corresponding to the dimmest

stimulus. Due to the limitations of the measurement scale, sensitivities are censored at 0 dB, the maximal, brightest stimulus. VF data is heteroskedastic with increasing variability in measurements as the level of damage increases. Due to the complex nature of VF data, analysis of VF progression is difficult.

Currently, physicians estimate rates of change in VF sensitivities using simple linear regression (SLR) for each subject and test location separately (Gardiner and Crabb, 2002; Nouri-Mahdavi et al., 2007). Shortcomings of SLR include not accounting for the heteroskedastic nature, censoring, and spatial structure of VF data, and not borrowing information across subjects. Recently, some models accounting for the hierarchical structure or spatial dependence have been proposed to model 24-2 VF data. Betz-Stablein et al. (2013) propose a model using conditional autoregressive priors to account for spatial correlations between neighboring test locations and VF clusters corresponding to the optic nerve head (Garway-Heath et al., 2000). Berchuck et al. (2019) present a model with localized spatio-temporal smoothing, allowing the spatial correlation to adapt over time. Montesano et al. (2021) develop a hierarchical model accounting for location and VF cluster levels, censoring, and heteroskedasticity. However, these methods are fit to data from individual eyes and ignore population information. Alternatively, Bryan et al. (2017) introduce a two-stage approach to fit a multilevel hierarchical model using data from the entire cohort, but do not account for spatial correlations. While there has been much work on modeling 24-2 VF data, these methods do not directly apply to 10-2 VF data, which has a higher density of locations in the central area of the retina.

We analyze longitudinal 10-2 VF data from the Advanced Glaucoma Progression Study (AGPS), a cohort of eyes with central or moderate to severe glaucoma. Our desired model should include data from all subjects and account for the spatial structure of VF locations, data censoring, and heteroskedasticity. In this study, we motivate and develop the Hierarchical Spatially varying Longitudinal Visual Field (HSLVF) model, a novel Bayesian hierarchical model accounting for censoring with spatially varying population- and subject-level coefficients and spatially varying population- and subject-level residual standard deviations (SD). We analyze data from the entire AGPS cohort

using the HSLVF model and demonstrate how careful construction of the spatial covariance structure allows additional flexibility in estimating intercepts and slopes.

In spatially varying coefficient models, regression coefficients vary over space to account for the spatial dependence in the response variable (Hastie and Tibshirani, 1993; Gelfand et al., 2010). With data observed over a grid of locations, regression coefficients may vary over space in either the discrete, areal setting or continuous, point-referenced setting (Gelfand et al., 2010). For our data, we allow the intercepts, slopes, and residual SDs to vary smoothly over space using a point-referenced method in the Bayesian framework. We assume Gaussian process (GP) models for the spatially varying coefficients. In a GP model, coefficients from any finite set of locations have a multivariate normal distribution with a mean function and valid covariance function specifying the expected value at each location and covariance between coefficients at any two locations, respectively (Gelfand et al., 2010). Gelfand et al. (2003) first described the use of GPs to model spatially varying coefficients in a Bayesian framework. We assign GP priors at the population and subject level within our hierarchical model. For simplicity and computational ease, the covariance functions of GPs are often assumed to be isotropic and stationary, meaning the covariance depends only on the distance between locations while the variance is constant over space (Gelfand et al., 2010). To better accommodate spatial irregularities within the data, we allow the subject-level intercepts and slopes to have GPs with nonstationary covariance functions. We demonstrate the benefits of nonstationarity by comparing models with stationary covariance functions to models with nonstationary covariance functions.

In Section 4.2, we describe the motivating data and study cohort. In Section 4.3, we briefly review GPs and develop the HSLVF model. In Section 4.4, we highlight the importance of using nonstationary covariance functions and analyze AGPS data using the HSLVF model. We conclude with a discussion in Section 4.5.

4.2 Visual field data

This section describes the study subjects and goals. We highlight data characteristics that drive model development.

4.2.1 Advanced Glaucoma Progression Study

The AGPS is an ongoing longitudinal study at the University of California, Los Angeles (Mohammadzadeh et al., 2021). The study adheres to the tenets of the Declaration of Helsinki and conforms to Health Insurance Portability and Accountability Act policies. At the time of enrollment, all study participants provided written informed consent. AGPS evaluated the central 10° of the VF using the 10-2 testing pattern of the Humphrey Field Analyzer approximately every 6 months. The data include 10-2 VF threshold sensitivities from 137 eyes with a range of 1 year to 6.25 years of observed follow-up time. The device measures threshold sensitivities at 68 locations within a 10° radius from the point of fixation, with testing locations 2° apart. We export data with the right eye format for ease of presentation and analysis; left eyes are mirror images of right eyes and are flipped left-right.

4.2.2 Visual field threshold sensitivity

The Humphrey Field Analyzer tests light intensities ranging from 10,000 apostilbs (asb) to 0.1 asb, the brightest to dimmest stimuli. These intensities are converted to the log scale (dB), where dB $= 40 - 10\log_{10}(\text{asb})$. Ability to see a 35 dB stimulus indicates good, healthy vision, while failure to see a 0 dB stimulus indicates blindness. In our cohort, we remove sensitivities of dB > 35, as these values are too good to be true for our cohort of moderate to severe glaucoma. Due to limitations of the measurement scale, the maximum intensity is 10,000 asb or 0 dB. There is no distinction between actual 0 dB sensitivities and values censored at 0 dB; a reading of 0 dB is recorded in both cases.

Censored 0 dB VF values are often treated as actual 0 dB in the analysis of progression, introducing positive bias in estimating rates of change. Bryan et al. (2013) report SLR not taking

censoring into account at 0 dB performs similarly to SLR taking censoring into account and to other robust models fit on each subject-location separately, but note that all models they compared result in predictions too large to be of clinical value. Montesano et al. (2021) find that hierarchical models accounting for censoring at 0 dB fit separately on individual eyes result in the smallest bias in estimating slopes when compared to SLR and hierarchical models that do not account for censoring. To reduce bias and properly model the true sensitivities, we incorporate censoring into our HSLVF model. We take a different approach to censoring and assume that values are interval censored in the interval $(-5, 0]$. Values of -10 dB or -20 dB would imply the stimuli are 10 or 20 times brighter than the current maximum of the Humphrey Field Analyzer and seem implausible.

4.2.3 Heteroskedasticity

Previous research suggests that variability in VF measurements tends to increase with the level of damage (Chauhan et al., 1993; Henson et al., 2000; Russell et al., 2012a). Variability tends to increase linearly as sensitivity drops from 35 dB to around 10 dB, then decreases as sensitivity approaches the measurement floor (Artes et al., 2005; Wyatt et al., 2007). The relationship between variability and sensitivity is often modeled by the function $\log(\text{SD}) = A \times \text{sensitivity (dB)} + B$, where the logarithm of the SD and true sensitivity are assumed to be linearly related (Henson et al., 2000; Russell et al., 2012a; Bryan et al., 2017). For eyes with glaucoma, Henson et al. (2000) estimate A and B to be -0.098 and 3.62 dB while Bryan et al. (2017) find A and B to be -0.08 and 2.82 dB. Montesano et al. (2021) obtain a good fit using a third-degree polynomial, but approximate the relationship with a linear broken stick model in practice. Alternatively, Zhu et al. (2014b) model measurement variability using a mixture of Weibull distributions instead of normal distributions.

We illustrate the relationship between log residual SD and sensitivity in Figure 4.1, We plot estimated intercepts against estimated log residual SDs from SLR, treating censored 0 dB as actual 0 dB values, fitting data from each subject-location separately. The blue lowess curve demonstrates how variability tends to increase as the intercepts drop from 30 dB to around 10 dB and then decreases rapidly as intercepts drop to 0 dB. There is noticeable variability in log residual SDs across subjects

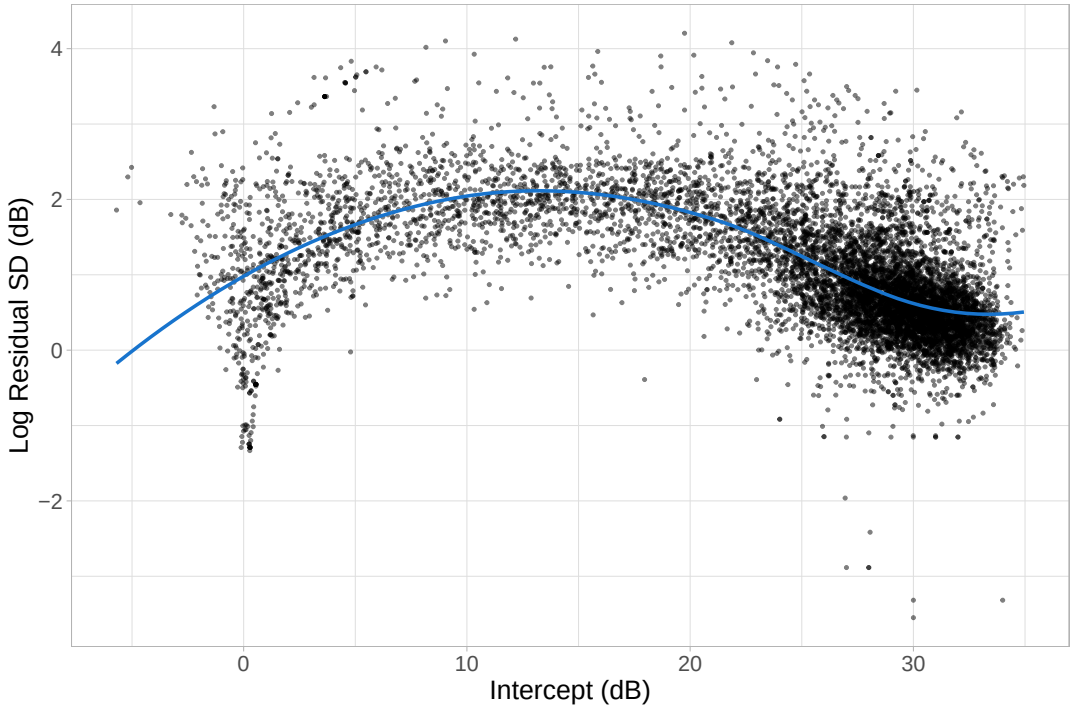

Figure 4.1: Scatter plot of intercept against log residual standard deviation (SD) posterior means from simple linear regression of sensitivity in decibels (dB) against time from each subject-location combination. The blue line is a lowess curve. We leave out 4 points with intercepts greater than 35 dB and 450 points with constant 0 dB sensitivities over time.

and locations. VF testing is subjective and responses are highly dependent on the cooperation of the subject (Broadway, 2012). To better accommodate this variability, we want to allow residual SDs to differ between subjects and locations in a principled fashion, offering more flexibility than assuming a linear relationship between $\log(\text{SD})$ and true sensitivity applied to all subjects and locations.

4.2.4 Spatial structure of test locations

Glaucoma is characterized by loss of retinal ganglion cells over time and this loss can be monitored using optical coherence tomography (OCT). In addition to functional VF measures, structural thickness measurements of retinal cells from OCT provide another means to monitor glaucoma progression. To combine both structural and functional information, the VF locations are displaced to where the underlying retinal ganglion cells are connected to their photoreceptors (Drasdo et al.,

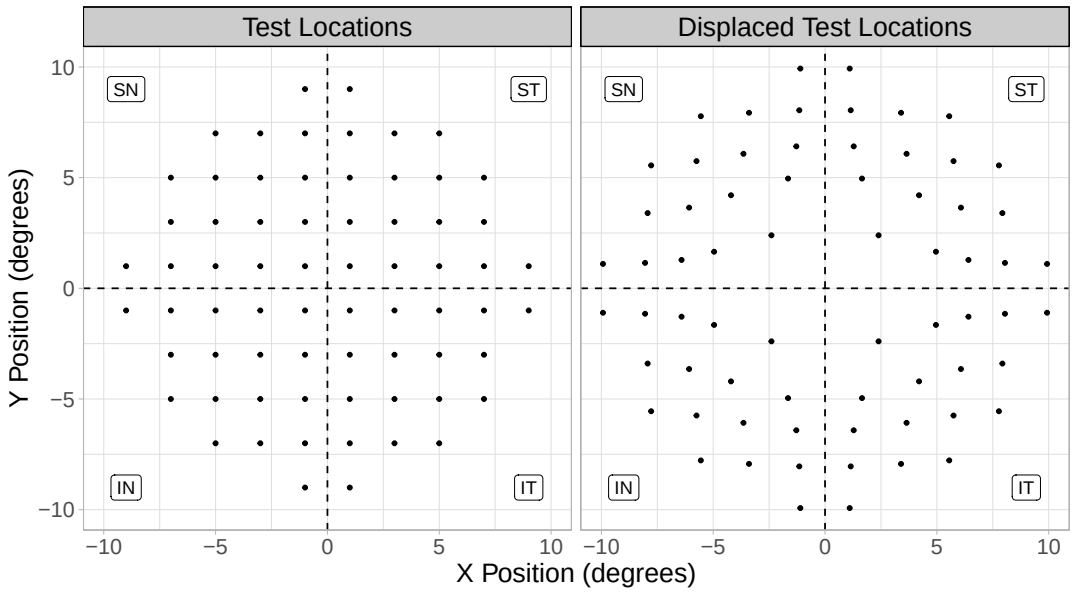

Figure 4.2: Plot of the 10-2 visual field test locations before (left) and after (right) accounting for ganglion cell displacement. The distances are measured in degrees from the center. We analyze and visualize data in right eye orientation, where SN is superior nasal, ST is superior temporal, IN is inferior nasal, and IT is inferior temporal.

2007; Montesano et al., 2020). We show the position of the 68 test locations before and after this displacement in Figure 4.2 (Drasdo et al., 2007; Raza et al., 2011). We denote the coordinates of each location $(\text{X position}, \text{Y position})$ in two dimensional space and label locations by their original planar positions as shown on the left in Figure 4.2. The positions are measured in degrees from the center $(0,0)$ of the VF, which corresponds to the fovea, the center of the retina. The displacement is more prominent in the center and decreases with distance from the center. We use the displaced test locations as shown on the right in Figure 4.2 to define distances between locations in our model. The smallest and largest Euclidean distance between any two displaced locations is 1.5 and 20.0 degrees, respectively.

4.2.5 Modeling criteria

In the analysis of VF progression, we are interested in estimating location-specific rates of change for each subject to identify eyes in need of treatment to prevent further vision loss. To achieve this,

we model data from all subjects in AGPS and account for data censoring, measurement variability, and the spatial structure of test locations. Ophthalmology still lacks a model that accommodates all aspects of VF data simultaneously. The Bayesian hierarchical framework allows us to accomplish this in a principled manner, shrinking intercepts and slopes towards the population averages at each location in exchange for reducing uncertainty. In assigning priors, we can incorporate our knowledge of the pathophysiology of glaucoma to guide this shrinkage.

4.3 Methods

This section introduces the GP priors we use to model the spatially varying coefficients, formulates the single eye and HSLVF models, and describes model implementation.

4.3.1 Gaussian processes

A spatial stochastic process $z(\mathbf{s})$ for $\mathbf{s} \in \mathbb{R}^d$, dimension $d \geq 1$, is Gaussian if for any $K \geq 1$ spatial locations $\mathbf{s}_1, \ldots, \mathbf{s}_K \in \mathbb{R}^d$, any finite collection of real-valued random variables $\{z(\mathbf{s}_1), \ldots, z(\mathbf{s}_K)\}$ is multivariate normal (Williams and Rasmussen, 2006; Bogachev, 1998; Banerjee et al., 2014). We denote a GP as

$$z(\mathbf{s}) \sim \text{GP}(m(\mathbf{s}), C(\mathbf{s}, \mathbf{s}')),$$

with mean function $m(\mathbf{s}) = \mathbb{E}[z(\mathbf{s})]$ and covariance function $C(\mathbf{s}, \mathbf{s}') = \text{cov}[z(\mathbf{s}), z(\mathbf{s}')]$ for two locations $\mathbf{s}$ and $\mathbf{s}'$, which may be the same or distinct. The covariance function $C(\mathbf{s}, \mathbf{s}')$ describes the association between outcomes $z(\mathbf{s})$ and $z(\mathbf{s}')$. In general, we expect values closer together in the process to be more highly correlated than values far apart.

The form of the covariance function encodes assumptions on the GP such as stationarity and isotropy. Stationarity means $C(\mathbf{s}, \mathbf{s}')$ depends only on the spatial separation vector $\mathbf{s} - \mathbf{s}'$ between points, and isotropy means $C(\mathbf{s}, \mathbf{s}')$ depends only on the distance between locations $h = \|\mathbf{s} - \mathbf{s}'\|$, where $\|\cdot\|$ is the Euclidean norm, i.e., $C(\mathbf{s}, \mathbf{s}') \equiv C(h)$.

We define stationary Matérn covariance functions as

$$C^S(\mathbf{s},\mathbf{s}') = \sigma^2 M(h|\nu,\ell), \tag{4.1}$$

where $\sigma^2 > 0$ is the variance and $M(h|\nu,\ell)$ is the Matérn correlation function (Matern, 1986)

$$M(h|\nu,\ell) = \frac{2^{1-\nu}}{\Gamma(\nu)}(\sqrt{2\nu}h/\ell)^\nu K_\nu(\sqrt{2\nu}h/\ell),$$

where $\nu > 0$ is the smoothness parameter, $\ell > 0$ is the lengthscale, and K_ν is the modified Bessel function of the second kind of order ν (Abramowitz and Stegun, 1964). In general, the process is m times mean square differentiable if and only if $\nu > m$ (Williams and Rasmussen, 2006). The lengthscale parameter ℓ controls how quickly the correlation decays as a function of distance with larger ℓ indicating slower correlation decay.

4.3.2 Nonstationary covariance functions

Nonstationary covariance functions depend on the referenced locations and are not invariant under spatial shifts. Paciorek and Schervish (2006) and Risser and Turek (2020) provide a nonstationary version of the Matérn covariance function as

$$C(\mathbf{s},\mathbf{s}') = \sigma(\mathbf{s})\sigma(\mathbf{s}')\frac{|\boldsymbol{\Sigma}(\mathbf{s})|^{1/4}|\boldsymbol{\Sigma}(\mathbf{s}')|^{1/4}}{\left|\frac{\boldsymbol{\Sigma}(\mathbf{s})+\boldsymbol{\Sigma}(\mathbf{s}')}{2}\right|^{1/2}} M(\sqrt{Q(\mathbf{s},\mathbf{s}')}|\nu)$$

where

$$Q(\mathbf{s},\mathbf{s}') = (\mathbf{s}-\mathbf{s}')^T\left(\frac{\boldsymbol{\Sigma}(\mathbf{s})+\boldsymbol{\Sigma}(\mathbf{s}')}{2}\right)^{-1}(\mathbf{s}-\mathbf{s}'),$$

where $\sigma(\mathbf{s})$ is the spatially varying SD process and $\boldsymbol{\Sigma}(\mathbf{s})$ is the matrix-valued anisotropy process. The SD process $\sigma(\mathbf{s})$ allows the variance of $z(\mathbf{s})$ to vary as a function of $\mathbf{s}$ and the anisotropy process $\boldsymbol{\Sigma}(\mathbf{s})$ controls the lengthscale and directional spatial dependence. We define a spatially constant anisotropy process $\boldsymbol{\Sigma}(\mathbf{s}) \equiv \Sigma$ for all $\mathbf{s}$, since we do not expect there to be directional spatial dependence. We use

nonstationary Matérn covariance functions of the form

$$C^{NS}(\mathbf{s},\mathbf{s}') = \sigma(\mathbf{s})\sigma(\mathbf{s}')M(h|\nu,\ell) \tag{4.2}$$

where the SD process $\sigma(\mathbf{s})$ is spatially varying while the spatial correlation $M(h|\nu,\ell)$ depends on the distance between locations h.

4.3.3 Single eye model

To highlight the importance of using nonstationary covariance functions, we first develop models fit on data from individual eyes using either stationary or nonstationary Matérn covariance functions. We refer to the model with only stationary covariance functions as the Single Eye Stationary (SES) model and the model with nonstationary covariance functions as the Single Eye Nonstationary (SENS) model. For a single subject, let observation y_{jk} be the VF threshold sensitivity measured in dB at visit $j = 1,\ldots,J$, where J is the number of visits for the subject, in test location $k = 1,\ldots,K$, observed at time t_j, with $t_1 = 0$. Location $s_k = (\text{X position}_k, \text{Y position}_k)$ denotes the two-dimensional spatial coordinates for test location k. Let y^*_{jk} denote the true, latent sensitivity for observations, where we observe

$$y_{jk} = \begin{cases} 0, & \text{if } -5 \le y^*_{jk} \le 0 \\ y^*_{jk}, & \text{otherwise} \end{cases}$$

and $y^*_{jk} < -5$ is assumed impossible. We model y^*_{jk} as linear in time

$$\begin{aligned} y^*_{jk} &= \gamma_0 + \gamma_1 t_j + \beta_{0k} + \beta_{1k} t_j + \varepsilon_{jk} \\ \varepsilon_{jk}|\tau^2_k &\sim \mathrm{N}(0,\tau^2_k), \\ \log \tau_k &= \gamma_2 + \sigma_k, \end{aligned}$$

where γ_0, γ_1, and γ_2 are the subject's global eye intercept, slope, and log residual SD, respectively, β_{0k}, β_{1k}, and σ_k are the intercept, slope, and log residual SD processes, respectively, in test location k. We assign independent normal priors on the global effects $\gamma_0 \sim N(20, 6^2)$, $\gamma_1 \sim N(-0.2, 0.5^2)$, $\gamma_2 \sim N(0.9, 0.5^2)$.

For the SES model, we model the intercept process with $\beta_{0k}|\boldsymbol{\theta}_{\beta 0} \sim \text{GP}(0, C^S_{\beta_0}(\mathbf{s}_k, \mathbf{s}_{k'}))$, slope process with $\beta_{1k}|\boldsymbol{\theta}_{\beta 1} \sim \text{GP}(0, C^S_{\beta 1}(\mathbf{s}_k, \mathbf{s}_{k'}))$, log residual SD process with $\log \sigma_k|\boldsymbol{\theta}_\sigma \sim \text{GP}(0, C^S_\sigma(\mathbf{s}_k, \mathbf{s}_{k'}))$ where covariance functions $C^S_{\beta 0}$, $C^S_{\beta 1}$, and C^S_σ are stationary Matérn covariance functions as in (4.1). The superscripts S and NS denote stationary and nonstationary covariance functions, respectively. The hyperparameters for the intercept, slope, and log residual processes are $\boldsymbol{\theta}_{\beta 0} = \{\sigma_{\beta 0}, \nu_{\beta 0}, \ell_{\beta 0}\}$, $\boldsymbol{\theta}_{\beta 1} = \{\sigma_{\beta 1}, \nu_{\beta 1}, \ell_{\beta 1}\}$, and $\boldsymbol{\theta}_\sigma = \{\sigma_\sigma, \nu_\sigma, \ell_\sigma\}$, respectively. For the GP SD and lengthscale parameters, we wish to avoid flat priors that could pull the posterior towards extreme values. We assign truncated-normal priors on all GP parameters $\sigma_{\beta 0} \sim N^+(0, 10^2)$, $\sigma_{\beta 1}, \sigma_\sigma \sim N^+(0, 2^2)$, and $\ell_{\beta 0}, \ell_{\beta 1}, \ell_\sigma \sim N^+(0, 5^2)$ where $N^+(a, b)$ is a normal distribution with mean a and variance b restricted to the positive real line.

For the SENS model, we use nonstationary covariance functions $C^{NS}_{\beta 0}$ and $C^{NS}_{\beta 1}$ as described in (4.2) for more flexible processes on the intercepts and slopes, respectively. The hyperparameters for the intercept and slope processes are $\boldsymbol{\theta}_{\beta 0} = \{\lambda_{\beta 0k}, \nu_{\beta 0}, \ell_{\beta 0}\}$ and $\boldsymbol{\theta}_{\beta 1} = \{\lambda_{\beta 1k}, \nu_{\beta 1}, \ell_{\beta 1}\}$, respectively. We model the spatially varying SD process for the intercept process with $\log \lambda_{\beta 0k}|\mu_{\lambda 0}, \sigma_{\lambda 0}, \nu_{\lambda 0}, \ell_{\lambda 0} \sim \text{GP}(\mu_{\lambda 0}, C^S_{\lambda 0}(\mathbf{s}_k, \mathbf{s}_{k'}))$ and for the slope process with $\log \lambda_{\beta 1k}|\mu_{\lambda 1}, \sigma_{\lambda 1}, \nu_{\lambda 1}, \ell_{\lambda 1} \sim \text{GP}(\mu_{\lambda 1}, C^S_{\lambda 1}(\mathbf{s}_k, \mathbf{s}_{k'}))$ where $C^S_{\lambda 0}$ and $C^S_{\lambda 1}$ take the stationary Matérn form. We model the log residual SD process with $\log \sigma_k|\boldsymbol{\theta}_\sigma \sim \text{GP}(0, C^S_\sigma(\mathbf{s}_k, \mathbf{s}_{k'}))$ in parallel with the SES model. We assign truncated-normal priors to all GP parameters $\sigma_{\beta 0} \sim N^+(0, 10^2)$, $\sigma_{\beta 1}, \sigma_\sigma \sim N^+(0, 2^2)$, $\sigma_{\lambda 0}, \sigma_{\lambda 1} \sim N^+(0, 3^2)$, and $\ell_{\beta 0}, \ell_{\beta 1}, \ell_\sigma, \ell_{\lambda 0}, \ell_{\lambda 1} \sim N^+(0, 5^2)$. We assign normal priors on the means of the log SD processes $\log \lambda_{\beta 0k}$ and $\log \lambda_{\beta 1k}$ with $\mu_{\lambda 0} \sim N(2.3, 0.25^2)$ and $\mu_{\lambda 1} \sim N(-0.2, 0.25^2)$, respectively. Finally, we assume all hyperparameters are independent *a priori*.

4.3.4 Model specification for a hierarchical spatially varying longitudinal visual field model

The proposed HSLVF model is an extension of the SENS model to accommodate data from the entire cohort, allowing the subject-specific intercept and slope processes to have nonstationary Matérn covariance functions. We augment the notation from the single eye model to include data from all subjects. Let observation y_{ijk} be the VF threshold sensitivity measured in dB of subject $i = 1, \ldots, n$ at visit $j = 1, \ldots, J_i$, at time t_{ij}, where J_i is the number of visits for subject i, in test location $k = 1, \ldots, K$, with $t_{i1} = 0$ for all subjects. Let y^*_{ijk} denote the true, latent sensitivity for observations. As with the single eye models, we observe

$$y_{ijk} = \begin{cases} 0, & \text{if } -5 \leq y^*_{ijk} \leq 0 \\ y^*_{ijk}, & \text{otherwise.} \end{cases}$$

We model y^*_{ijk} as

$$\begin{aligned} y^*_{ijk} &= \mu_0 + \mu_1 t_{ij} + \alpha_{0k} + \alpha_{1k} t_{ij} + \beta_{0ik} + \beta_{1ik} t_{ij} + \varepsilon_{ijk} \\ \varepsilon_{ijk} | \tau^2_{ik} &\sim \mathrm{N}(0, \tau^2_{ik}), \\ \log \tau_{ik} &= \mu_2 + \phi_k + \sigma_{ik}, \end{aligned}$$

where μ_0, μ_1, and μ_2 are the population global eye intercept, slope, and log residual SD, respectively; α_{0k}, α_{1k}, and ϕ_k are the test location k population-level intercept, slope, and log residual SD processes, respectively; and β_{0ik}, β_{1ik}, and σ_{ik} are subject-specific intercept, slope, and log residual SD processes, respectively, in location k. We use weakly informative priors to encourage reasonable ranges of values and facilitate computing. We assign normal priors on the global effects $\mu_0 \sim N(20, 6^2)$, $\mu_1 \sim N(-0.2, 0.5^2)$, $\mu_2 \sim N(0.9, 0.5^2)$.

We model the population-level intercept process with $\alpha_{0k} | \sigma_{\alpha 0}, \nu_{\alpha 0}, \ell_{\alpha 0} \sim \mathrm{GP}(0, C^S_{\alpha 0}(\mathbf{s}_k, \mathbf{s}_{k'}))$, slope process with $\alpha_{1k} | \sigma_{\alpha 1}, \nu_{\alpha 1}, \ell_{\alpha 1} \sim \mathrm{GP}(0, C^S_{\alpha 1}(\mathbf{s}_k, \mathbf{s}_{k'}))$, log residual SD process with $\phi_k | \sigma_\phi, \nu_\phi, \ell_\phi \sim \mathrm{GP}(0, C^S_\phi(\mathbf{s}_k, \mathbf{s}_{k'}))$ where covariance functions $C^S_{\alpha 0}$, $C^S_{\alpha 1}$, and C^S_ϕ take the station-

ary form of the Matérn covariance function. We model the subject-specific log residual SD process with $\sigma_{ik}|\sigma_\sigma, \nu_\sigma, \ell_\sigma \sim \text{GP}(0, C^S_\sigma(\mathbf{s}_k, \mathbf{s}_{k'}))$ where C^S_σ takes the same stationary Matérn form. For the subject-specific intercept and slope processes, we assign nonstationary Matérn covariance functions with $\beta_{0ik}|\lambda_{\beta 0k}, \nu_{\beta 0}, \ell_{\beta 0} \sim \text{GP}(0, C^{NS}_{\beta 0}(\mathbf{s}_k, \mathbf{s}_{k'}))$ and $\beta_{1ik}|\lambda_{\beta 1k}, \nu_{\beta 1}, \ell_{\beta 1} \sim \text{GP}(0, C^{NS}_{\beta 1}(\mathbf{s}_k, \mathbf{s}_{k'}))$, respectively. We model the spatially varying SD process for β_{0ik} with $\log \lambda_{\beta 0k}|\mu_{\lambda 0}, \sigma_{\lambda 0}, \nu_{\lambda 0}, \ell_{\lambda 0} \sim \text{GP}(\mu_{\lambda 0}, C^S_{\lambda 0}(\mathbf{s}_k, \mathbf{s}_{k'}))$ and for β_{1ik} with $\log \lambda_{\beta 1k}|\mu_{\lambda 1}, \sigma_{\lambda 1}, \nu_{\lambda 1}, \ell_{\lambda 1} \sim \text{GP}(\mu_{\lambda 1}, C^S_{\lambda 1}(\mathbf{s}_k, \mathbf{s}_{k'}))$.

We assign normal priors on the means of the spatially varying log SD processes $\log \lambda_{\beta 0k}$ and $\log \lambda_{\beta 1k}$ with $\mu_{\lambda 0} \sim N(2.3, 0.25^2)$ and $\mu_{\lambda 1} \sim N(-0.2, 0.25^2)$, respectively. We set all GP smoothness parameters $\nu_{\alpha 0}$, $\nu_{\alpha 1}$, ν_ϕ, ν_σ, $\nu_{\beta 0}$, $\nu_{\beta 1}$, $\nu_{\lambda 0}$, $\nu_{\lambda 1} = \frac{1}{2}$ since we obtain measurements from a coarse grid of test locations and expect the processes to be rough. We assign truncated-normal priors on GP SD parameters $\sigma_{\alpha 0} \sim N^+(0, 10^2)$, $\sigma_\phi, \sigma_\sigma \sim N^+(0, 2^2)$, $\sigma_{\lambda 0}, \sigma_{\lambda 1} \sim N^+(0, 3^2)$, and an inverse gamma prior on the SD parameter for the population-level slope process $\sigma_{\alpha 1} \sim IG(3, 2)$ to avoid infinitesimal SDs. For GP lengthscales, we expect values to fall plausibly between 1.5 and 20.0 degrees, the range of distances between any two test locations. We therefore assign identical truncated-normal priors on all GP lengthscales $\ell_{\alpha 0}, \ell_{\alpha 1}, \ell_\phi, \ell_\sigma, \ell_{\beta 0}, \ell_{\beta 1}, \ell_{\lambda 0}, \ell_{\lambda 1} \sim N^+(0, 5^2)$. Finally, we assume all top-level hyperparameters are independent *a priori*.

4.3.5 Computation and inference

We use the R programming language (R Core Team, 2021) for all analyses and GGPLOT2 (Wickham, 2016) for visualization. To fit the models, we use Markov Chain Monte Carlo (MCMC) methods (Metropolis et al., 1953; Robert and Casella, 2005) implemented in NIMBLE v0.13.1 (de Valpine et al., 2017, 2022). To sample from the posteriors, we use Gibbs sampling and update specific parameters using Metropolis-Hastings within Gibbs. We update global effects μ_0, μ_1, μ_2, μ_{λ_0}, and μ_{λ_1} using scalar Metropolis-Hastings random walk samplers; each pair of SD and lengthscale parameters from GPs together using multivariate Metropolis-Hastings random walk samplers; and the subject-level random effects β_{0ik}, β_{1ik}, and σ_{ik} and spatially varying log SD processes $\log \lambda_{\beta_0 k}$ and $\log \lambda_{\beta_1 k}$ using multivariate Metropolis-Hastings random walk samplers in sub-blocks after testing various

schemes to improve sampling efficiency (Risser and Turek, 2020). We jointly sample subject-level intercepts and slopes in sub-blocks of size 4. We separately sample the subject-level residual SDs and spatially varying log SD processes in sub-blocks of size 8 or 10. To improve mixing of chains, we repeat samplers for 13 subject-level intercepts and slopes with poor mixing and the spatially varying log SD processes 3 times in each MCMC iteration.

We follow Vehtari et al. (2021)'s recommendation for assessing convergence, monitoring the potential scale reduction factor $\widehat{R}$, bulk effective sample sizes (ESS), and tail ESS for all model parameters. We use the R package RSTAN (Stan Development Team, 2022) to provide ESS measures and $\widehat{R}$ convergence diagnostics.

4.4 Advanced Glaucoma Progression Study

We apply our methodology to 10-2 VF measurements from 137 eyes (137 subjects) in the AGPS. After removing sensitivities greater than 35 dB, we analyze 86,819 observations from 9,316 subject-location VF series.

4.4.1 Single eye model results

First, we fit the SES and SENS models on data from each eye separately as part of an exploratory analysis before applying the HSLVF model. Comparing the 9,316 subject-location parameters between the SES and SENS models, we find that the mean (10th, 90th percentile) difference (SES − SENS) in posterior means is $-0.01(-0.56, 0.46)$ and $0.03(-0.08, 0.12)$ for intercepts and slopes, respectively. Figure 4.3 plots the posterior means of subject-location intercepts and slopes from SES against SENS. There is close agreement between the models for the majority of intercepts and slopes. However, there are some intercepts and slopes with very large differences as highlighted by the lighter colored dots.

Figure 4.4 provides results from two subjects where the intercepts and slopes differ greatly between the SES and SENS models at multiple locations. For heat maps of intercepts and slopes,

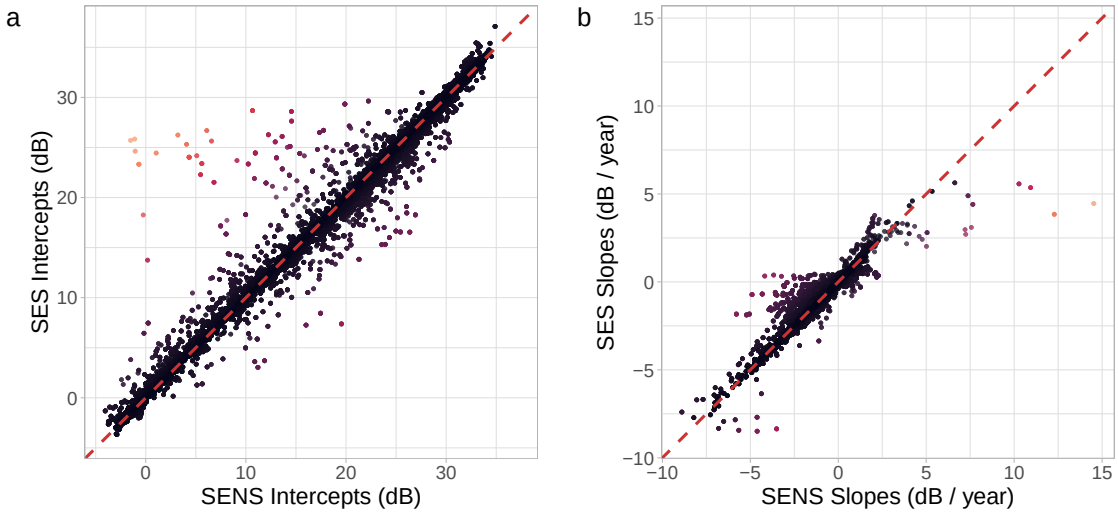

Figure 4.3: Scatter plot of (a) intercept posterior means in decibels (dB) and (b) slope posterior means (dB / year) from the Single Eye Stationary (SES, y-axis) vs. Single Eye Nonstationary (SENS, x-axis) models. The red dashed line is the $x = y$ diagonal. Some subject-location intercepts and slopes that differ greatly between the models are highlighted by the lighter colored dots.

the coordinates on the top and left indicate the original X and Y coordinate positions of test locations, respectively. The posterior mean fitted lines in Figure 4.4 (c, f) show locations where the SENS model follows the raw data more closely than the SES model. For Subject 1 (Figure 4.4 a-c), the SENS model allows for more flexibility in intercepts and slopes in locations (-1, 5) to (7, 5), where intercepts and slopes deviate greatly from the population average. The SES model does not allow for large differences between neighboring locations. For Subject 2 (Figure 4.4 d-f), the SENS model more closely follows the data in locations $(-5,-7)$, $(-3,-7)$, $(-1,-9)$, and $(1,-9)$ where intercepts differ up to 8 dB between the models.

4.4.2 HSLVF results

We run the HSLVF model with 9 chains of 360,000 iterations after a burn-in of 60,000, a thin of 200 for a total of 13,500 posterior samples. We summarize convergence diagnostics in Table 4.1. Despite the complexity of the model, customized sub-block random walk samplers in NIMBLE prove to be efficient at sampling highly correlated posteriors. The total runtime for the 360,000 iterations

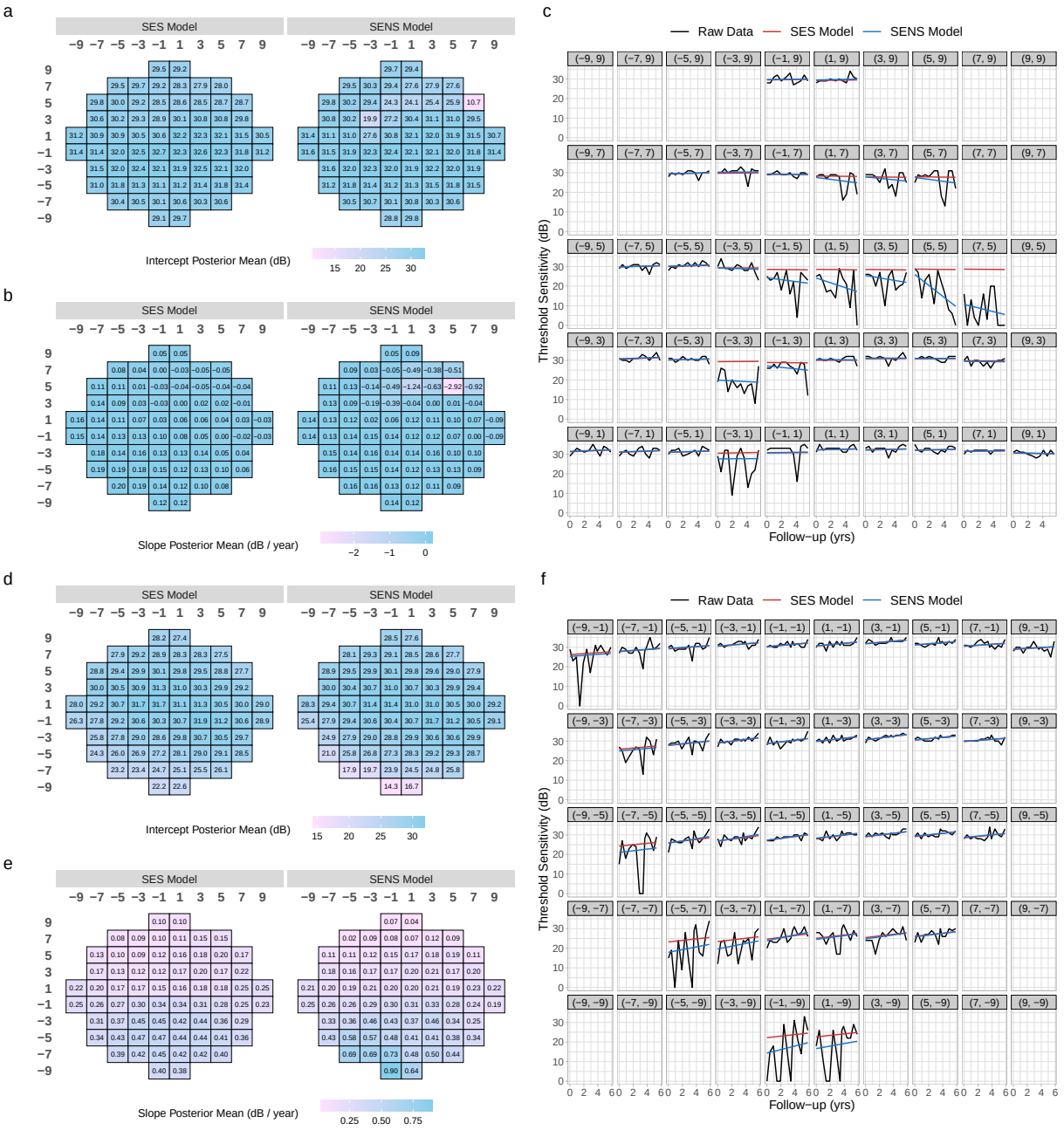

Figure 4.4: Comparison of intercepts, slopes, and model fit between the Single Eye Stationary (SES) and Single Eye Nonstationary (SENS) models on two subjects. Heat maps of intercept posterior means in decibels (dB) across the 68 locations for (a) Subject 1 and (d) Subject 2. Heat maps of slope posterior means (dB / year) across the 68 locations for (b) Subject 1 and (e) Subject 2. Posterior mean fitted lines from the SES (red) and SENS (blue) models for the (c) superior hemiretina of Subject 1 and (f) inferior hemiretina of Subject 2. Posterior mean fitted lines from the SENS model follow the raw data more closely than SES in multiple locations.

Table 4.1: The mean minimum/maximum bulk effective sample size (ESS), tail ESS, and potential scale reduction factor $\widehat{R}$ for the HSLVF model broken down by parameter types. For convergence, $\widehat{R}$ should be < 1.01 and ESS should be > 100 per chain. The first column gives the parameter types: Hyperparameters include global parameters and Gaussian process hyperparameters; Population-level includes the population-level intercepts, slopes, and log residual standard deviations (SD) across all locations; Intercepts, Slopes, and Log Residual SDs include the subject-level intercepts, slopes, and log residual SDs across all locations, respectively; and Log SD Process includes the spatially varying log SD for the subject-level intercepts and slopes across all locations. The second column gives the number of parameters summarized.

		Bulk ESS		Tail ESS		$\widehat{R}$	
Parameter	#	Mean	Min	Mean	Min	Mean	Max
Hyperparameters	19	11532.7	3406.3	12186.8	7632.6	1.000	1.002
Population-Level	204	11944.7	7766.0	12782.2	10984.9	1.000	1.001
Intercepts	9316	12708.3	2134.6	13041.8	5604.4	1.000	1.005
Slopes	9316	10487.1	1972.9	11905.2	4389.8	1.000	1.005
Log Residual SDs	9316	13255.3	7472.8	13183.4	9546.0	1.000	1.002
Log SD Process	136	5048.4	2493.1	8400.7	4986.4	1.001	1.004

is approximately 36 hours (Apple M1 Pro 10-core CPU). The chains achieve $\widehat{R}$ less than 1.01 with bulk and tail ESS greater than 100 per chain for all parameters, indicating convergence. Visual assessment of model convergence was satisfactory.

Table 4.2 gives posterior means and 95% credible intervals (CrI) for global and GP parameters from the HSLVF model. The lengthscales for the intercept and slope GPs are longer for the subject-level processes than the population-level, suggesting that spatial correlation decays much slower or noise is higher so smoothing is more appropriate for the subject-level intercepts and slopes. In contrast, the lengthscales for the log residual SD processes are not longer for the subject-level process than the population-level process. The posterior mean and 95% CrI difference in population-level and subject-level lengthscales $(\ell_\phi - \ell_\sigma)$ for the log residual SD processes is 2.24 $(-2.10, 8.30)$.

Figure 4.5 displays heat maps of posterior means of population-level intercepts and slopes. There is more damage in the superior hemiretina and inferior nasal areas, while the central locations have higher intercepts. To assess glaucoma progression, ophthalmologists are most interested in the slopes. We declare a slope to be significantly negative or positive when the upper bound or lower

Table 4.2: Posterior mean and 95% credible interval (CrI) for global parameters and Gaussian process (GP) parameters for the HSLVF model. SD stands for standard deviation.

Parameters	Symbols	Level	Mean	95% CrI
Global Parameters				
Intercept	μ_0		20.59	(16.37, 24.43)
Slope	μ_1		-0.24	$(-0.40, -0.08)$
Log Residual SD	μ_2		1.04	(0.81, 1.30)
GP SD Parameters				
Intercept	$\sigma_{\alpha 0}$	Population	3.06	(2.40, 3.90)
Slope	$\sigma_{\alpha 1}$	Population	0.10	(0.08, 0.14)
Log Residual SD	σ_{ϕ}	Population	0.20	(0.15, 0.28)
Log Residual SD	σ_{σ}	Subject	0.73	(0.70, 0.76)
GP Lengthscale Parameters				
Intercept	$\ell_{\alpha 0}$	Population	13.06	(8.42, 18.87)
Slope	$\ell_{\alpha 1}$	Population	14.91	(9.11, 21.83)
Log Residual SD	ℓ_{ϕ}	Population	9.01	(4.75, 15.01)
Intercept	$\ell_{\beta 0}$	Subject	32.11	(28.07, 36.55)
Slope	$\ell_{\beta 1}$	Subject	32.49	(28.08, 37.15)
Log Residual SD	ℓ_{σ}	Subject	6.77	(6.12, 7.50)
Spatially Varying Log SD GP Parameters				
Intercept Mean	$\mu_{\lambda 0}$		2.36	(2.04, 2.70)
Slope Mean	$\mu_{\lambda 1}$		-0.52	$(-0.74, -0.26)$
Intercept SD	$\sigma_{\lambda 0}$		0.41	(0.31, 0.55)
Slope SD	$\sigma_{\lambda 1}$		0.22	(0.15, 0.30)
Intercept Lengthscale	$\ell_{\lambda 0}$		9.01	(5.04, 14.79)
Slope Lengthscale	$\ell_{\lambda 1}$		9.74	(5.31, 15.85)

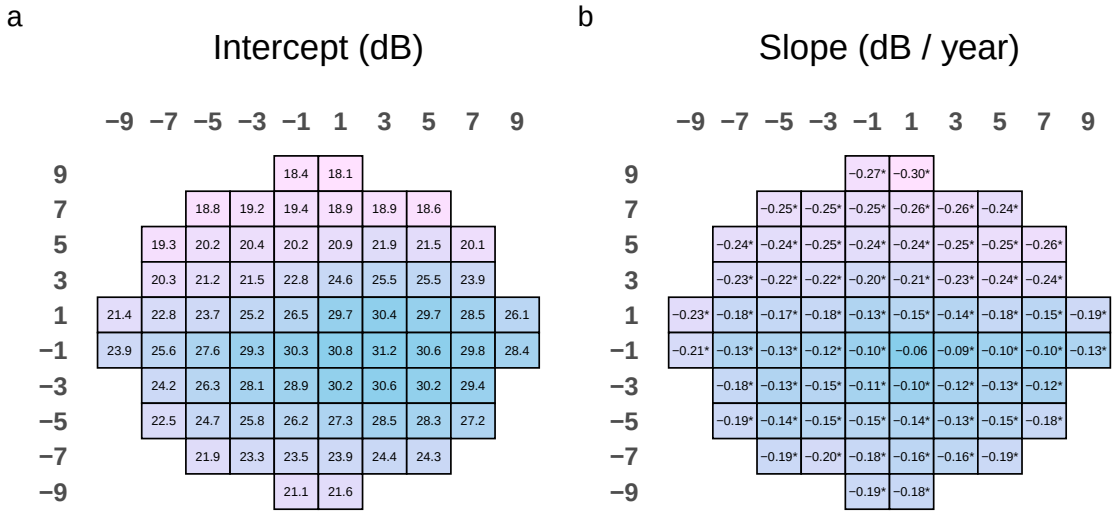

Figure 4.5: Heat maps of posterior means of population-level (a) intercepts in decibels (dB) and (b) slopes (dB / year) across the 68 test locations from the HSLVF model. The asterisk indicates a significant negative slope, where we define significance when the upper bound of the 95% credible interval is less than 0. We use a significantly negative slope to indicate progression.

bound of the 95% CrI is less than or greater than 0, respectively. We use a significantly negative slope to indicate progression. All locations except $(1, -1)$ have significant negative population-level slopes, with the steepest slopes in the superior locations. Across all 9,316 subject-location series, the HSLVF model detects 29.9% of slopes as significantly negative and 6.4% as significantly positive.

Figure 4.6 shows heat maps of posterior means of the spatially varying SD parameters for the nonstationary Matérn covariance functions of the subject-level intercepts and slopes. The SDs for intercepts are larger in the superior hemiretina and inferior nasal areas, indicating random intercepts are more variable. The SDs for slopes are larger in the superior hemiretina, indicating more variability in random slopes.

4.5 Discussion

We motivate and develop a Bayesian hierarchical model to analyze VF data from all study subjects, accommodating data censoring, heteroskedasticity, and the spatial structure of test locations.

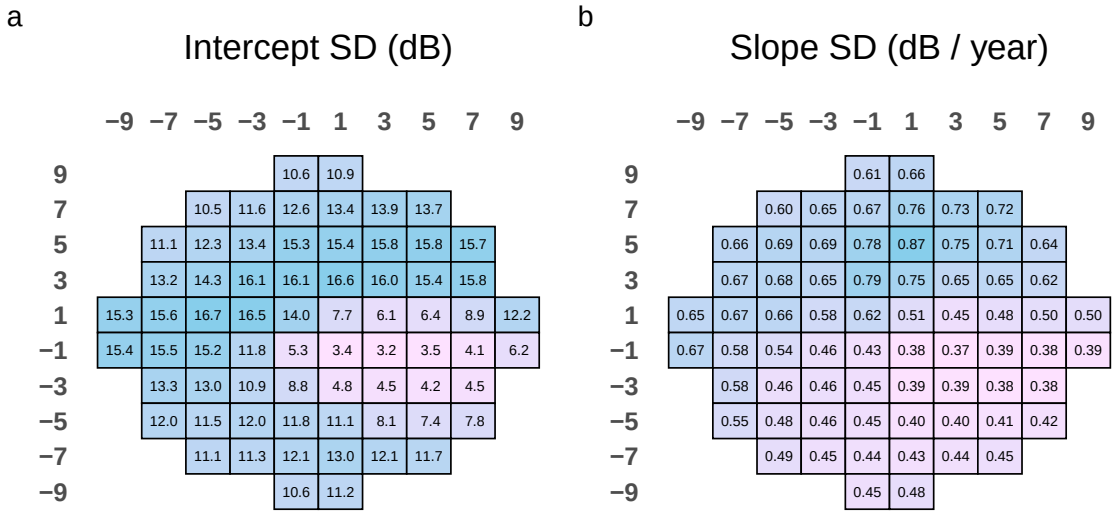

Figure 4.6: Heat maps of posterior means of the spatially varying standard deviation (SD) parameters $\lambda_{\beta 0k}$ and $\lambda_{\beta 1k}$ for the nonstationary Matérn covariance functions of the subject-level (a) intercepts in decibels (dB) and (b) slopes (dB / year), respectively, in the HSLVF model. There is more variability in random intercepts and slopes in the superior hemiretina than inferior.

Given that measurement variability is known to increase with glaucoma damage, we highlight the performance of HSLVF in identifying progressing locations in our cohort of moderate to severe glaucoma subjects. Our model borrows information across subjects, facilitating inference of population-level parameters and reducing uncertainty in estimating subject-level parameters. Most notably, the HSLVF model detects significant negative slopes in 29.9% of the 9,316 subject-location VF series. In contrast, SLR detects only 10.3% of the subject-location VF series as significantly negative.

We take a different approach in handling censored measurements of 0 dB by treating them as interval censored. While the cutoff of 0 dB is arbitrary in nature, we opt to provide a lower bound to keep estimated sensitivities within a reasonable range. It is imperative to include data censoring in our model, especially for subject-specific repeated measures with constant values of 0 dB. Without censoring, the model would not be able to provide appropriate inference for subject-location residual variances in constant repeated measures. Other studies have modeled heteroskedasticity by estimating the relationship between log(SD) and true sensitivity using frequency-of-seeing and test-retest datasets (Chauhan et al., 1993; Artes et al., 2002; Henson et al., 2000; Russell et al., 2012a;

Montesano et al., 2021). In the absence of test-retest data, we instead allow residual SDs to vary across subject-locations in a principled fashion through our GP prior. This approach accounts for subject-specific variability arising from the subjective nature of the VF testing procedure.

We account for spatial structure in intercepts, slopes, and log residual SDs through our GP covariance functions. By using nonstationary covariance functions with spatially varying SDs to model subject-specific intercepts and slopes, our model appropriately captures the variation in random intercept and slope SDs across different test locations. The spatially varying SDs provide insight into which locations are more heterogeneous with more variability in baseline damage and rates of change, indicating regions of potential progression. There are a variety of approaches to specify the anisotropy processes as in Risser and Turek (2020). As part of exploratory analyses, we fit models using nonstationary covariance functions with spatially varying lengthscales. These early analyses did not convergence, possibly indicating that there is not enough information to estimate varying lengthscales or that the process is highly spatially correlated. We define our covariances using displaced distances between locations, combining both functional and structural information. Our approach facilitates the development towards combined structure-function analyses to further improve detection of progression.

Bibliography

Abramowitz, M. and Stegun, I. A. (1964). *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*, volume 55. US Government Printing Office.

Abu, S. L., Marín-Franch, I., and Racette, L. (2020). A framework for assessing glaucoma progression using structural and functional indices jointly. *PLoS One*, 15(7):e0235255.

Allison, K., Patel, D., and Alabi, O. (2020). Epidemiology of glaucoma: the past, present, and predictions for the future. *Cureus*, 12(11).

Apanasovich, T. V., Genton, M. G., and Sun, Y. (2012). A valid Matérn class of cross-covariance functions for multivariate random fields with any number of components. *Journal of the American Statistical Association*, 107(497):180–193.

Artes, P. H., Hutchison, D. M., Nicolela, M. T., LeBlanc, R. P., and Chauhan, B. C. (2005). Threshold and variability properties of matrix frequency-doubling technology and standard automated perimetry in glaucoma. *Investigative Ophthalmology & Visual Science*, 46(7):2451–2457.

Artes, P. H., Iwase, A., Ohno, Y., Kitazawa, Y., and Chauhan, B. C. (2002). Properties of perimetric threshold estimates from full threshold, SITA Standard, and SITA Fast strategies. *Investigative Ophthalmology & Visual Science*, 43(8):2654–2659.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data*. Chapman and Hall/CRC, 2nd edition.

Banerjee, S., Gelfand, A. E., Finley, A. O., and Sang, H. (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(4):825–848.

Barnard, J., McCulloch, R., and Meng, X.-L. (2000). Modeling covariance matrices in terms of

standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, 10(4):1281–1311.

Berchuck, S. I., Mwanza, J.-C., and Warren, J. L. (2019). Diagnosing glaucoma progression with visual field data using a spatiotemporal boundary detection method. *Journal of the American Statistical Association*, 114(527):1063–1074.

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

Bogachev, V. I. (1998). *Gaussian Measures*, volume 62 of *Mathematical Surveys and Monographs*. American Mathematical Society.

Broadway, D. C. (2012). Visual field testing for glaucoma–a practical guide. *Community Eye Health*, 25(79-80):66.

Bryan, S. R., Eilers, P. H., Lesaffre, E. M., Lemij, H. G., and Vermeer, K. A. (2015). Global visit effects in point-wise longitudinal modeling of glaucomatous visual fields. *Investigative Ophthalmology & Visual Science*, 56(8):4283–4289.

Bryan, S. R., Eilers, P. H., Rosmalen, J. v., Rizopoulos, D., Vermeer, K. A., Lemij, H. G., and Lesaffre, E. M. (2017). Bayesian hierarchical modeling of longitudinal glaucomatous visual fields using a two-stage approach. *Statistics in Medicine*, 36(11):1735–1753.

Bryan, S. R., Vermeer, K. A., Eilers, P. H., Lemij, H. G., and Lesaffre, E. M. (2013). Robust and censored modeling and prediction of progression in glaucomatous visual fields. *Investigative Ophthalmology & Visual Science*, 54(10):6694–6700.

Castruccio, S., Ombao, H., and Genton, M. G. (2018). A scalable multi-resolution spatio-temporal model for brain activation and connectivity in fMRI data. *Biometrics*, 74(3):823–833.

Chauhan, B. C., Garway-Heath, D. F., Goñi, F. J., Rossetti, L., Bengtsson, B., Viswanathan, A. C., and Heijl, A. (2008). Practical recommendations for measuring rates of visual field change in glaucoma. *British Journal of Ophthalmology*, 92(4):569–573.

Chauhan, B. C., Tompkins, J. D., LeBlanc, R. P., and McCormick, T. A. (1993). Characteristics of frequency-of-seeing curves in normal subjects, patients with suspected glaucoma, and patients with glaucoma. *Investigative Ophthalmology & Visual Science*, 34(13):3534–3540.

De Moraes, C. G., Hood, D. C., Thenappan, A., Girkin, C. A., Medeiros, F. A., Weinreb, R. N., Zangwill, L. M., and Liebmann, J. M. (2017). 24-2 visual fields miss central defects shown on 10-2 tests in glaucoma suspects, ocular hypertensives, and early glaucoma. *Ophthalmology*, 124(10):1449–1456.

de Valpine, P., Paciorek, C., Turek, D., Michaud, N., Anderson-Bergman, C., Obermeyer, F., Wehrhahn Cortes, C., Rodrìguez, A., Temple Lang, D., and Paganin, S. (2022). *NIMBLE: MCMC, Particle Filtering, and Programmable Hierarchical Modeling*. R package version 0.13.1.

de Valpine, P., Turek, D., Paciorek, C., Anderson-Bergman, C., Temple Lang, D., and Bodik, R. (2017). Programming with models: writing statistical algorithms for general model structures with NIMBLE. *Journal of Computational and Graphical Statistics*, 26:403–413.

Drasdo, N., Millican, C. L., Katholi, C. R., and Curcio, C. A. (2007). The length of Henle fibers in the human retina and a model of ganglion receptive field density in the visual field. *Vision Research*, 47(22):2901–2911.

Gardiner, S. K. and Crabb, D. P. (2002). Examination of different pointwise linear regression methods for determining visual field progression. *Investigative Ophthalmology & Visual Science*, 43(5):1400–1407.

Garway-Heath, D. F., Poinoosawmy, D., Fitzke, F. W., and Hitchings, R. A. (2000). Mapping the visual field to the optic disc in normal tension glaucoma eyes. *Ophthalmology*, 107(10):1809–1815.

Gaspari, G. and Cohn, S. E. (1999). Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125(554):723–757.

Ge, T., Müller-Lenke, N., Bendfeldt, K., Nichols, T. E., and Johnson, T. D. (2014). Analysis of multiple sclerosis lesions via spatially varying coefficients. *The Annals of Applied Statistics*, 8(2):1095–1118.

Gelfand, A. E., Diggle, P., Guttorp, P., and Fuentes, M. (2010). *Handbook of Spatial Statistics*. CRC Press.

Gelfand, A. E., Kim, H.-J., Sirmans, C., and Banerjee, S. (2003). Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association*, 98(462):387–396.

Gelfand, A. E. and Schliep, E. M. (2016). Spatial statistics and Gaussian processes: A beautiful marriage. *Spatial Statistics*, 18:86–104.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis*. Chapman & Hall/CRC, 3rd edition.

Genton, M. G. and Kleiber, W. (2015). Cross-covariance functions for multivariate geostatistics. *Statistical Science*, 30(2):147–163.

Gneiting, T., Kleiber, W., and Schlather, M. (2010). Matérn cross-covariance functions for multivariate random fields. *Journal of the American Statistical Association*, 105(491):1167–1177.

Gordon, M. O., Beiser, J. A., Brandt, J. D., Heuer, D. K., Higginbotham, E. J., Johnson, C. A., Keltner, J. L., Miller, J. P., Parrish, R. K., Wilson, M. R., et al. (2002). The Ocular Hypertension Treatment Study: baseline factors that predict the onset of primary open-angle glaucoma. *Archives of Ophthalmology*, 120(6):714–720.

Gössl, C., Auer, D. P., and Fahrmeir, L. (2001). Bayesian spatiotemporal inference in functional magnetic resonance imaging. *Biometrics*, 57(2):554–562.

Gracitelli, C. P., Tatham, A. J., Zangwill, L. M., Weinreb, R. N., Liu, T., and Medeiros, F. A. (2014). Estimated rates of retinal ganglion cell loss in glaucomatous eyes with and without optic disc hemorrhages. *PLoS One*, 9(8):e105611.

Guttorp, P. and Gneiting, T. (2006). Studies in the history of probability and statistics XLIX On the Matérn correlation family. *Biometrika*, 93(4):989–995.

Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 55(4):757–779.

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

Kass, M. A., Heuer, D. K., Higginbotham, E. J., Johnson, C. A., Keltner, J. L., Miller, J. P., Parrish, R. K., Wilson, M. R., Gordon, M. O., Group, O. H. T. S., et al. (2002). The Ocular Hypertension Treatment Study: a randomized trial determines that topical ocular hypotensive medication delays or prevents the onset of primary open-angle glaucoma. *Archives of Ophthalmology*, 120(6):701–713.

Kim, H. and Lee, J. (2017). Hierarchical spatially varying coefficient process model. *Technometrics*, 59(4):521–527.

Kingman, S. (2004). Glaucoma is second leading cause of blindness globally. *Bulletin of the World Health Organization*, 82:887–888.

Kline, P. (2014). *An Easy Guide to Factor Analysis*. Routledge.

Liu, Y.-Y., Ishikawa, H., Chen, M., Wollstein, G., Schuman, J. S., and Rehg, J. M. (2013). Longitudinal modeling of glaucoma progression using 2-dimensional continuous-time hidden markov model. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013: 16th International Conference, Nagoya, Japan, September 22-26, 2013, Proceedings, Part II 16*, pages 444–451. Springer.

Liu, Z., Bartsch, A. J., Berrocal, V. J., and Johnson, T. D. (2019). A mixed-effects, spatially varying coefficients model with application to multi-resolution functional magnetic resonance imaging data. *Statistical Methods in Medical Research*, 28(4):1203–1215.

Mansouri, K., Leite, M. T., Medeiros, F., Leung, C., and Weinreb, R. (2011). Assessment of rates of structural change in glaucoma using imaging technologies. *Eye*, 25(3):269–277.

Matern, B. (1986). *Spatial Variation*. Springer, 2nd edition.

Medeiros, F. A., Leite, M. T., Zangwill, L. M., and Weinreb, R. N. (2011). Combining structural and functional measurements to improve detection of glaucoma progression using Bayesian hierarchical models. *Investigative Ophthalmology & Visual Science*, 52(8):5794–5803.

Medeiros, F. A., Zangwill, L. M., Bowd, C., Mansouri, K., and Weinreb, R. N. (2012). The structure and function relationship in glaucoma: implications for detection of progression and measurement of rates of change. *Investigative Ophthalmology & Visual Science*, 53(11):6939–6946.

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

Miraftabi, A., Amini, N., Gornbein, J., Henry, S., Romero, P., Coleman, A. L., Caprioli, J., and Nouri-Mahdavi, K. (2016). Local variability of macular thickness measurements with SD-OCT and influencing factors. *Translational Vision Science & Technology*, 5(4):5.

Mohammadzadeh, V., Fatehi, N., Yarmohammadi, A., Lee, J. W., Sharifipour, F., Daneshvar, R., Caprioli, J., and Nouri-Mahdavi, K. (2020a). Macular imaging with optical coherence tomography in glaucoma. *Survey of Ophthalmology*, 65(6):597–638.

Mohammadzadeh, V., Rabiolo, A., Fu, Q., Morales, E., Coleman, A. L., Law, S. K., Caprioli, J., and Nouri-Mahdavi, K. (2020b). Longitudinal macular structure–function relationships in glaucoma. *Ophthalmology*, 127(7):888–900.

Mohammadzadeh, V., Su, E., Rabiolo, A., Shi, L., Zadeh, S. H., Law, S. K., Coleman, A. L., Caprioli, J., Weiss, R. E., and Nouri-Mahdavi, K. (2022a). Ganglion cell complex: The optimal measure for detection of structural progression in the macula. *American Journal of Ophthalmology*, 237:71–82.

Mohammadzadeh, V., Su, E., Shi, L., Coleman, A. L., Law, S. K., Caprioli, J., Weiss, R. E., and Nouri-Mahdavi, K. (2022b). Multivariate longitudinal modeling of macular ganglion cell complex: spatiotemporal correlations and patterns of longitudinal change. *Ophthalmology Science*, 2(3):100187.

Mohammadzadeh, V., Su, E., Zadeh, S. H., Law, S. K., Coleman, A. L., Caprioli, J., Weiss, R. E., and Nouri-Mahdavi, K. (2021). Estimating ganglion cell complex rates of change with Bayesian hierarchical models. *Translational Vision Science & Technology*, 10(4):15.

Montesano, G., Garway-Heath, D. F., Ometto, G., and Crabb, D. P. (2021). Hierarchical censored Bayesian analysis of visual field progression. *Translational Vision Science & Technology*, 10(12):4.

Montesano, G., Ometto, G., Hogg, R. E., Rossetti, L. M., Garway-Heath, D. F., and Crabb, D. P. (2020). Revisiting the Drasdo model: implications for structure-function analysis of the macular region. *Translational Vision Science & Technology*, 9(10):15.

Nouri-Mahdavi, K., Hoffman, D., Ralli, M., and Caprioli, J. (2007). Comparison of methods to predict visual field progression in glaucoma. *Archives of Ophthalmology*, 125(9):1176–1181.

Paciorek, C. J. and Schervish, M. J. (2006). Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics: The Official Journal of the International Environmetrics Society*, 17(5):483–506.

Penny, W. D., Trujillo-Barreto, N. J., and Friston, K. J. (2005). Bayesian fMRI time series analysis with spatial priors. *NeuroImage*, 24(2):350–362.

Phu, J. and Kalloniatis, M. (2021). The Frontloading Fields Study (FFS): detecting changes in mean deviation in glaucoma using multiple visual field tests per clinical visit. *Translational Vision Science & Technology*, 10(13):21.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rabiolo, A., Mohammadzadeh, V., Fatehi, N., Morales, E., Coleman, A. L., Law, S. K., Caprioli, J., and Nouri-Mahdavi, K. (2020). Comparison of rates of progression of macular OCT measures in glaucoma. *Translational Vision Science & Technology*, 9(7):50.

Rao, H., Kumbar, T., Kumar, A., Babu, J., Senthil, S., and Garudadri, C. (2013). Agreement between event-based and trend-based glaucoma progression analyses. *Eye*, 27(7):803–808.

Raza, A. S., Cho, J., de Moraes, C. G., Wang, M., Zhang, X., Kardon, R. H., Liebmann, J. M., Ritch, R., and Hood, D. C. (2011). Retinal ganglion cell layer thickness and local visual field sensitivity in glaucoma. *Archives of Ophthalmology*, 129(12):1529–1536.

Risser, M. D. and Turek, D. (2020). Bayesian inference for high-dimensional nonstationary Gaussian processes. *Journal of Statistical Computation and Simulation*, 90(16):2902–2928.

Robert, C. and Casella, G. (2005). *Monte Carlo Statistical Methods*. Springer, 2nd edition.

Rui, C., Montesano, G., Crabb, D. P., Brusini, P., Chauhan, B. C., Rossetti, L. M., Fogagnolo, P., Giraud, J.-M., Fenolland, J.-R., and Oddone, F. (2021). Improving event-based progression analysis in glaucomatous visual fields. *Scientific Reports*, 11(1):16353.

Russell, R. A., Crabb, D. P., Malik, R., and Garway-Heath, D. F. (2012a). The relationship between variability and sensitivity in large-scale longitudinal visual field data. *Investigative Ophthalmology & Visual Science*, 53(10):5985–5990.

Russell, R. A., Malik, R., Chauhan, B. C., Crabb, D. P., and Garway-Heath, D. F. (2012b). Improved estimates of visual field progression using bayesian linear regression to integrate structural

information in patients with ocular hypertension. *Investigative Ophthalmology & Visual Science*, 53(6):2760–2769.

Schmidt, A. M. and Gelfand, A. E. (2003). A Bayesian coregionalization approach for multivariate pollutant data. *Journal of Geophysical Research: Atmospheres*, 108(D24).

Stan Development Team (2022). RStan: the R interface to Stan. R package version 2.21.7.

Stone, M. (1977). An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):44–47.

Su, Y.-S. and Yajima, M. (2021). *R2jags: Using R to Run 'JAGS'*. R package version 0.7-1.

Swaminathan, S. S., Berchuck, S. I., Jammal, A. A., Rao, J. S., and Medeiros, F. A. (2022). Rates of glaucoma progression derived from linear mixed models using varied random effect distributions. *Translational Vision Science & Technology*, 11(2):16.

Tatham, A. J. and Medeiros, F. A. (2017). Detecting structural progression in glaucoma with optical coherence tomography. *Ophthalmology*, 124(12):S57–S65.

Tham, Y.-C., Li, X., Wong, T. Y., Quigley, H. A., Aung, T., and Cheng, C.-Y. (2014). Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology*, 121(11):2081–2090.

Thompson, A. C., Jammal, A. A., Berchuck, S. I., Mariottoni, E. B., Wu, Z., Daga, F. B., Ogata, N. G., Urata, C. N., Estrela, T., and Medeiros, F. A. (2020a). Comparing the rule of 5 to trend-based analysis for detecting glaucoma progression on OCT. *Ophthalmology Glaucoma*, 3(6):414–420.

Thompson, A. C., Jammal, A. A., and Medeiros, F. A. (2020b). A review of deep learning for screening, diagnosis, and detection of glaucoma progression. *Translational Vision Science & Technology*, 9(2):42–42.

Tibbits, M. M., Groendyke, C., Haran, M., and Liechty, J. C. (2014). Automated factor slice sampling. *Journal of Computational and Graphical Statistics*, 23(2):543–563.

Traynis, I., De Moraes, C. G., Raza, A. S., Liebmann, J. M., Ritch, R., and Hood, D. C. (2014). Prevalence and nature of early glaucomatous defects in the central 10° of the visual field. *JAMA Ophthalmology*, 132(3):291–297.

Vehtari, A., Gelman, A., and Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27(5):1413–1432.

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., and Bürkner, P.-C. (2021). Rank-normalization, folding, and localization: an improved $\widehat{R}$ for assessing convergence of MCMC (with discussion). *Bayesian Analysis*, 16(2):667–718.

Ver Hoef, J. M. and Barry, R. P. (1998). Constructing and fitting models for cokriging and multivariable spatial prediction. *Journal of Statistical Planning and Inference*, 69(2):275–294.

Wackernagel, H. (2013). *Multivariate Geostatistics*. Springer, 3rd edition.

Watanabe, S. and Opper, M. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11(12):3571–3594.

Weinreb, R. N., Aung, T., and Medeiros, F. A. (2014). The pathophysiology and treatment of glaucoma: a review. *JAMA*, 311(18):1901–1911.

Weinreb, R. N. and Khaw, P. T. (2004). Primary open-angle glaucoma. *The Lancet*, 363(9422):1711–1720.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Williams, C. K. and Rasmussen, C. E. (2006). *Gaussian Processes for Machine Learning*, volume 2. MIT Press.

Wu, Z. and Medeiros, F. A. (2018). Comparison of visual field point-wise event-based and global trend-based analysis for detecting glaucomatous progression. *Translational Vision Science & Technology*, 7(4):20–20.

Wu, Z., Medeiros, F. A., Weinreb, R. N., Girkin, C. A., and Zangwill, L. M. (2019). Comparing 10-2 and 24-2 visual fields for detecting progressive central visual loss in glaucoma eyes with early central abnormalities. *Ophthalmology Glaucoma*, 2(2):95–102.

Wyatt, H. J., Dul, M. W., and Swanson, W. H. (2007). Variability of visual field measurements is correlated with the gradient of visual sensitivity. *Vision Research*, 47(7):925–936.

Zhang, F., Jiang, W., Wong, P., and Wang, J.-P. (2016). A Bayesian probit model with spatially varying coefficients for brain decoding using fMRI data. *Statistics in Medicine*, 35(24):4380–4397.

Zhu, H., Fan, J., and Kong, L. (2014a). Spatially varying coefficient model for neuroimaging data with jump discontinuities. *Journal of the American Statistical Association*, 109(507):1084–1098.

Zhu, H., Russell, R. A., Saunders, L. J., Ceccon, S., Garway-Heath, D. F., and Crabb, D. P. (2014b). Detecting changes in retinal function: analysis with non-stationary Weibull error regression and spatial enhancement (ANSWERS). *PloS One*, 9(1):e85654.