

A SPATIALLY VARYING HIERARCHICAL RANDOM EFFECTS MODEL FOR LONGITUDINAL MACULAR STRUCTURAL DATA IN GLAUCOMA PATIENTS

BY ERICA SU^{1,a}, ROBERT E. WEISS^{1,b}, KOUROS NOURI-MAHDAVI^{2,d} AND ANDREW J. HOLBROOK^{1,c}

¹*Department of Biostatistics, Fielding School of Public Health, University of California, Los Angeles, ^aericasu@ucla.edu, ^brobweiss@ucla.edu, ^caholbroo@g.ucla.edu*

²*Glaucoma Division, Stein Eye Institute, David Geffen School of Medicine, University of California, Los Angeles, ^dnouri-mahdavi@jsei.ucla.edu*

We model longitudinal macular thickness measurements to monitor the course of glaucoma and prevent vision loss due to disease progression. The macular thickness varies over a 6×6 grid of locations on the retina, with additional variability arising from the imaging process at each visit. Currently, ophthalmologists estimate slopes using repeated simple linear regression for each subject and location. To estimate slopes more precisely, we develop a novel Bayesian hierarchical model for multiple subjects with spatially varying population-level and subject-level coefficients, borrowing information over subjects and measurement locations. We augment the model with visit effects to account for observed spatially correlated visit-specific errors. We model spatially varying: (a) intercepts, (b) slopes, and (c) log-residual standard deviations (SD) with multivariate Gaussian process priors with Matérn cross-covariance functions. Each marginal process assumes an exponential kernel with its own SD and spatial correlation matrix. We develop our models for and apply them to data from the Advanced Glaucoma Progression Study. We show that including visit effects in the model reduces error in predicting future thickness measurements and greatly improves model fit.

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. (2020)). 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 Berchuk, Mwanza and Warren (2019) present models accounting for spatial correlation fit to

Received February 2023; revised July 2024.

Key words and phrases. Bayesian modeling, ganglion cell complex, glaucoma, multivariate Gaussian processes, optical coherence tomography, random effects, spatially varying coefficients.

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 $20^\circ \times 20^\circ$ 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, Fan and Kong \(2014\)](#), [Liu et al. \(2019\)](#)), where regression coefficients are allowed to vary smoothly as a function 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 autoregressive (CAR) model ([Gössl, Auer and Fahrmeir \(2001\)](#), [Penny, Trujillo-Barreto and Friston \(2005\)](#), [Ge et al. \(2014\)](#)) or a Gaussian process (GP) model ([Zhang et al. \(2016a\)](#), [Castruccio, Ombao and Genton \(2018\)](#)) 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, Carlin and Gelfand \(2015\)](#)). 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, Carlin and Gelfand (2015)). 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 including 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, Kleiber and Schlather (2010) and Apanasovich, Genton and Sun (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 2 we describe the motivating data. In Section 3 we briefly review GPs and develop the SHREVE model. In Section 4 we present simulation results, evaluating the effectiveness of the SHREVE model. In Section 5 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 6.

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

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.

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 four OCT scans and a minimum of approximately two years of observed follow-up time, up to 4.25 years from baseline. Subjects returned approximately every six 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 black dot in Figure 1 and as a white

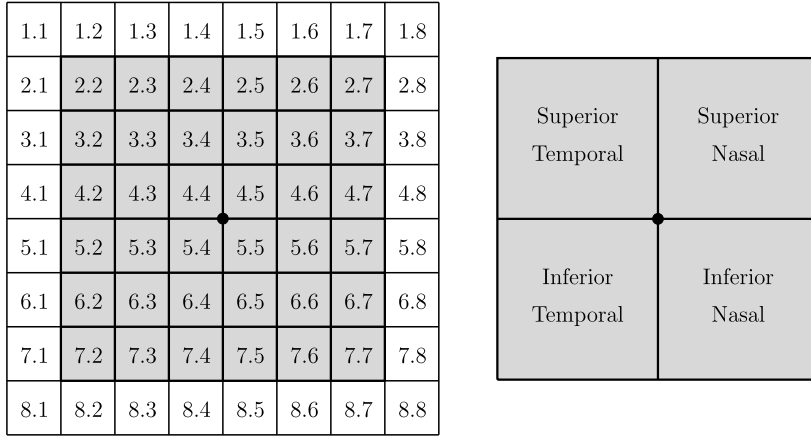


FIG. 1. Visualization of the 8×8 grid of superpixels and labels from the Spectralis posterior pole algorithm. The inner 36 superpixels included in the analysis are shaded in gray and delineated with thicker lines. Superpixels are shown in right eye orientation where rows 1–4 are located in the superior 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. For ease of reference, we divide the 36 superpixels into quadrants (superior temporal, superior nasal, inferior temporal, and inferior nasal) as shown on the right.

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 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. For ease of reference, we divide the 36 superpixels into quadrants (*superior temporal*, *superior nasal*, *inferior temporal*, and *inferior nasal*) by anatomical region. 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 1.

2.3. Data exploration. Let observation y_{ijk} be the GCC thickness measure in μm of subject $i = 1, \dots, n$ at visit $j = 1, \dots, J_i$, where J_i is the number of visits for subject i , in superpixel $k = 1, \dots, K$ observed at time t_{ij} , with $t_{i1} = 0$ for all subjects. Location $s_k = (\text{row}_k, \text{column}_k)$ 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 $j = 1, \dots, J_i$ 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 remove outliers by identifying pairs of consecutive points that have very large differences in GCC thicknesses between consecutive visits. For each pair of consecutive observations for each subject and superpixel, we calculate the consecutive-visit slope $(y_{ijk} - y_{i(j-1)k}) / (t_{ij} - t_{i(j-1)})$. The mean consecutive-visit slope across all pairs of consecutive visits for all subjects and superpixels is $-0.5 \mu\text{m}/\text{year}$. We center the consecutive-visit slopes

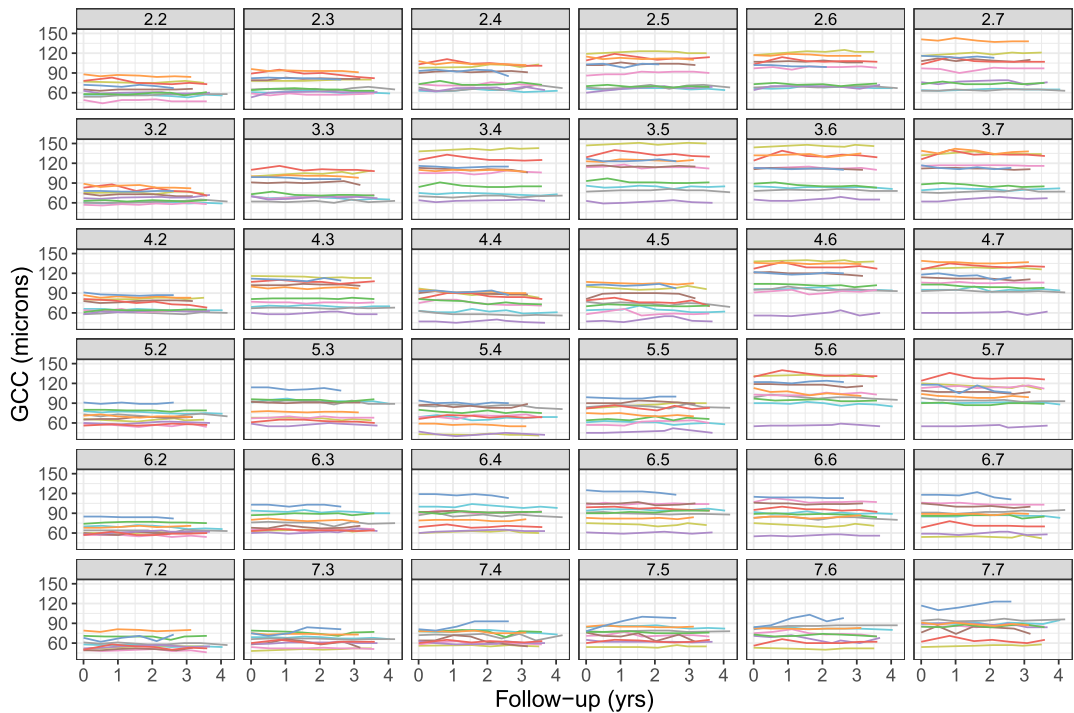


FIG. 2. Profile plots of ganglion cell complex (GCC) thickness measurements for 10 subjects across 36 superpixels against follow-up time in years since baseline visit. Each color represents a different subject. These profiles illustrate the variability in baseline GCC thickness across the 10 subjects 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).

by the mean and take the absolute value to get absolute consecutive-visit centered-slopes $|((y_{ijk} - y_{i(j-1)k})/(t_{ij} - t_{i(j-1)})) + 0.5|$. We flag pairs of observations $(y_{i(j-1)k}, y_{ijk})$ with absolute centered-slopes greater than 24 $\mu\text{m}/\text{year}$ with absolute differences greater than five μm as candidates for removal. We choose values that ensure the absolute consecutive-visit centered-slopes are unreasonably large; absolute differences greater than five μm ensure that the large slopes are not the result of short between-visit time differences. For each profile with flagged pairs of observations, we calculate the sum of absolute visit differences for the profile $\sum_{j=2}^{J_i} |y_{ij} - y_{i(j-1)k}|$ with and without either point in the flagged pair. We then remove the point in the flagged pair that results in the larger reduction in the sum of absolute visit differences. This rule ensures that the point removed is the more extreme outlier, deviating more from other observations in the profile. For each profile, if two or more observations are identified as outliers, we remove all remaining observations as well.

Eyes enrolled in the AGPS had moderate to severe glaucoma, thus exhibit a range of glaucomatous damage. Figure 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 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 show more variability both within and between subjects.

Figure 3 shows heatmaps 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

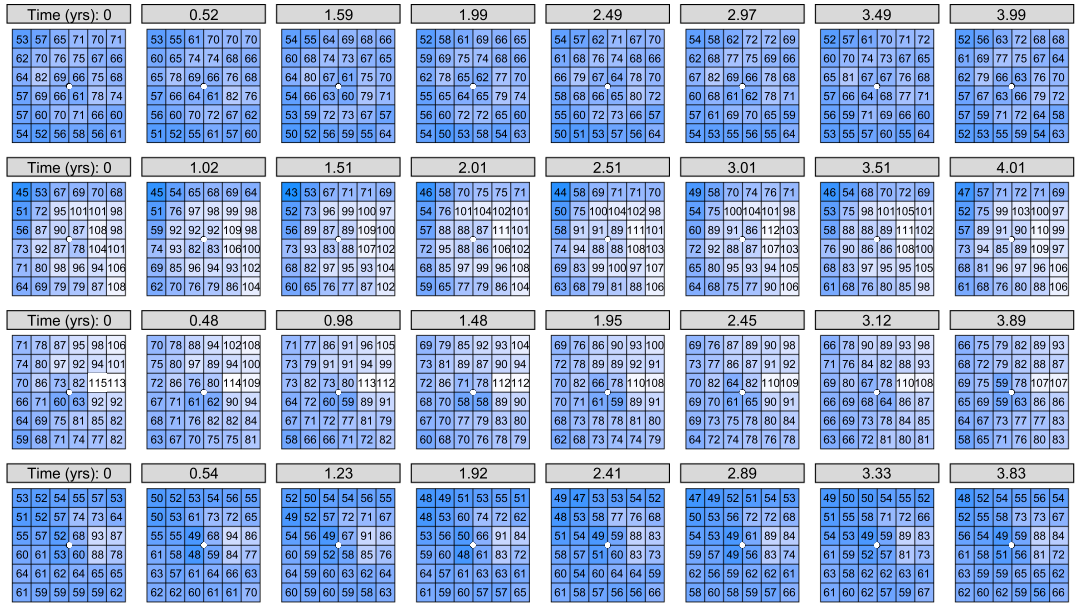


FIG. 3. Heatmaps of ganglion cell complex (GCC) thickness measurements (μm) across eight visits for four subjects for all 36 superpixels (top left 2.2 to bottom right 7.7). Each row is a different subject. The follow-up time of each visit is labeled at the top of each block. All maps share a common color 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.

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 4 shows heatmaps of estimated slopes (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, 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 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 5 shows spatially correlated slopes noticeable in the region from superpixels 3.4 to 3.7 down to 6.4 to 6.7.

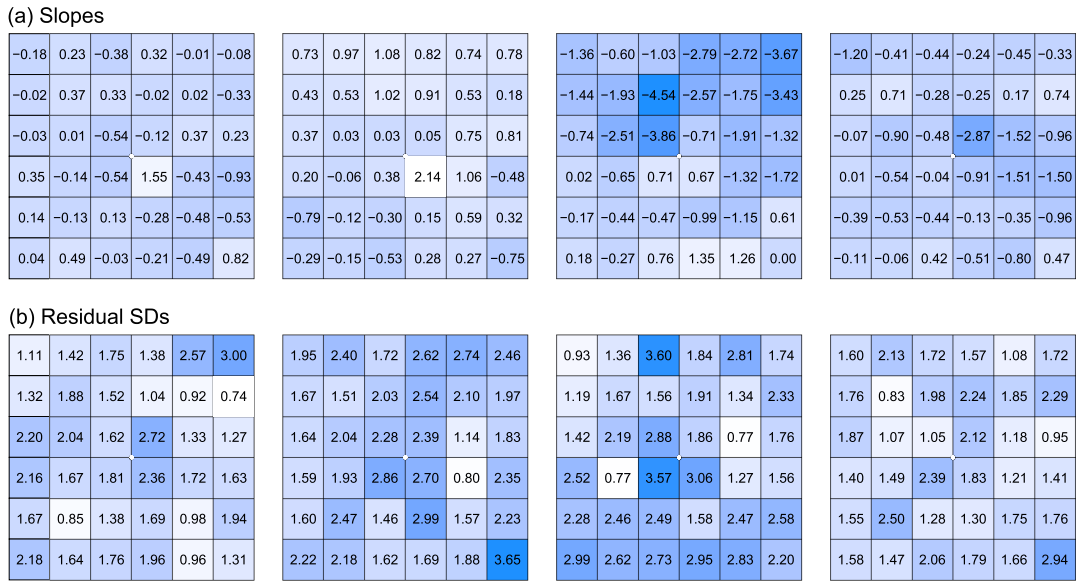


FIG. 4. Heatmaps of: (a) estimated slopes ($\mu\text{m}/\text{year}$) and (b) residual standard deviations (SD) (μm) for the same four subjects as in Figure 3 using simple linear regressions of ganglion cell complex (GCC) thicknesses on time since baseline in each superpixel. Each column is a different 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.

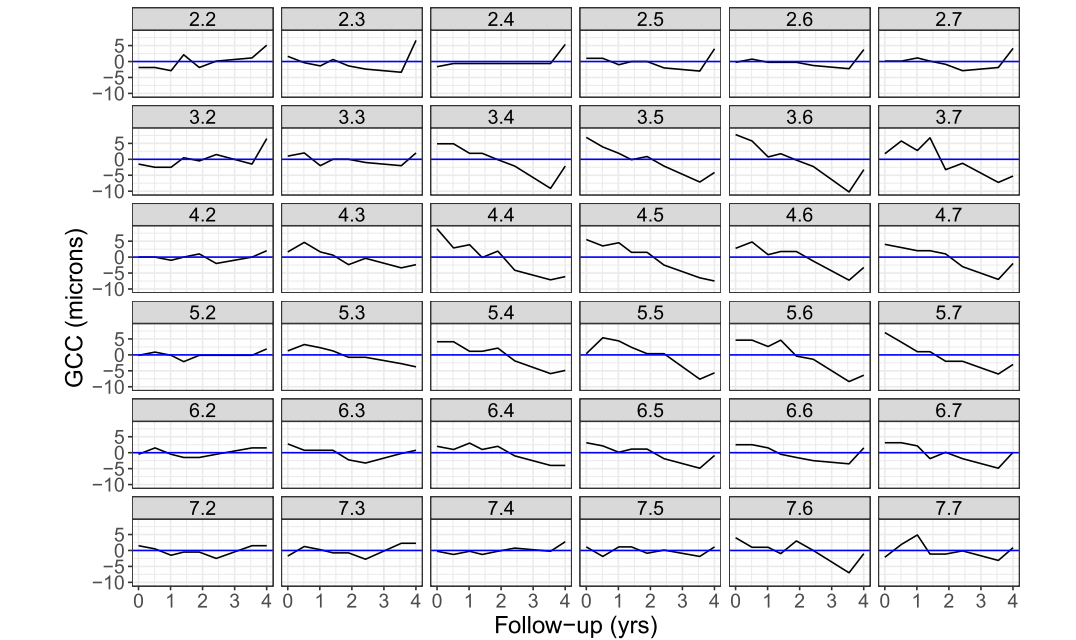


FIG. 5. Empirical residual profile plots (superpixel mean subtracted from ganglion cell complex (GCC) thickness) for a single subject across 36 superpixels. There is an increase at four years for many superpixel locations suggesting visit-specific spatially correlated errors.

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 toward their floor (Miraftabi et al. (2016)).

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.

3.1. *Gaussian processes.* A Gaussian spatial process (Rasmussen and Williams (2006), Bogachev (1998), Banerjee, Carlin and Gelfand (2015)) is a stochastic process $\{z(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^d\}$ in which any finite collection of real-valued random variables $\{z(\mathbf{s}_1), \dots, z(\mathbf{s}_K)\}$ is distributed as multivariate normal for every set of $K \geq 1$ spatial locations $\mathbf{s}_1, \dots, \mathbf{s}_K \in \mathbb{R}^d$, for dimension $d \geq 1$; we work only with $d = 2$. We denote a GP as

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

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, that is, $C(\mathbf{s}, \mathbf{s}') \equiv C(h)$.$

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 (Matérn (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 $\nu > m$ (Rasmussen and Williams (2006)). The lengthscale parameter ℓ controls how quickly the correlation decays as a function of distance with larger ℓ indicating slower correlation decay.

3.2. *Multivariate Gaussian processes.* Let $\mathbf{z}(\mathbf{s}) = (z_1(\mathbf{s}), \dots, z_P(\mathbf{s}))^T$ be a $P \times 1$ stochastic process, where each component $z_p(\mathbf{s})$ for $p = 1, \dots, P$ is a scalar random variable at location \mathbf{s} . Then $\mathbf{z}(\mathbf{s})$ is an MGP if any random vector $(\mathbf{z}(\mathbf{s}_1)^T, \dots, \mathbf{z}(\mathbf{s}_K)^T)^T$ from any set of $K \geq 1$ locations $\mathbf{s}_1, \dots, \mathbf{s}_K$ 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 $p, q =$

$1, \dots, P$, 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, $C_{pp}(h) = \sigma_{pp}^2 M(h|v_{pp}, \ell_{pp})$, for $p = 1, \dots, P$, with variance parameter $\sigma_{pp}^2 > 0$, smoothness parameter v_{pp} , and lengthscale parameter ℓ_{pp} . We model each cross-covariance function C_{pq} with a Matérn correlation function, $C_{pq}(h) = \sigma_{pq} M(h|v_{pq}, \ell_{pq})$, for $1 \leq p \neq q \leq P$, with covariance parameter σ_{pq} , smoothness parameter v_{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 v_{pp} , v_{pq} , ℓ_p , ℓ_{pq} , σ_{pp} , and σ_{pq} that result in a nonnegative definite cross-covariance function (Apanasovich, Genton and Sun (2012)). We use the simplest parameterization, where no additional parameters beyond σ_{pp}^2 , v_{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

$$(1) \quad \begin{aligned} v_{pq}(v_{pp}, v_{qq}) &= \frac{v_{pp} + v_{qq}}{2}, \\ \ell_{pq}(\ell_p, \ell_q) &= \sqrt{\frac{2}{\ell_p^{-2} + \ell_q^{-2}}}, \\ (2) \quad \sigma_{pq}(v_{pp}, v_{qq}, \ell_p, \ell_q, \sigma_{pp}, \sigma_{qq}, \mathbf{R}_{pq}) &= \sigma_{pp}\sigma_{qq} \frac{\ell_{pq}(\ell_p, \ell_q)}{\sqrt{\ell_p \ell_q}} \frac{\Gamma(v_{pq}(v_{pp}, v_{qq}))}{\Gamma^{1/2}(v_{pp}) + \Gamma^{1/2}(v_{qq})} \mathbf{R}_{pq}, \end{aligned}$$

where $\mathbf{R} = \{\mathbf{R}_{pq}\}$ 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. 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

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

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 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 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 $\boldsymbol{\theta}_\alpha = \{\sigma_{\alpha, pp}, v_{\alpha, p}, \ell_{\alpha, p}, \mathbf{R}_\alpha, p \in \{1, 2, 3\}\}$. 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 $C_{\alpha, pp}(\mathbf{s}_k, \mathbf{s}_{k'}) = \sigma_{\alpha, pp}^2 M(h|v_{\alpha, p}, \ell_{\alpha, p})$, for $p = 1, \dots, 3$, have PL marginal variances $\sigma_{\alpha, pp}^2$, PL smoothness parameters $v_{\alpha, p}$, and PL lengthscales $\ell_{\alpha, p}$. PL cross-covariance functions $C_{\alpha, pq}(\mathbf{s}_k, \mathbf{s}_{k'}) = \sigma_{\alpha, pq} M(h|v_{\alpha, pq}, \ell_{\alpha, pq})$ have covariance parameters $\sigma_{\alpha, pq}$ between processes p and q , smoothness parameters $v_{\alpha, pq}$, and lengthscales

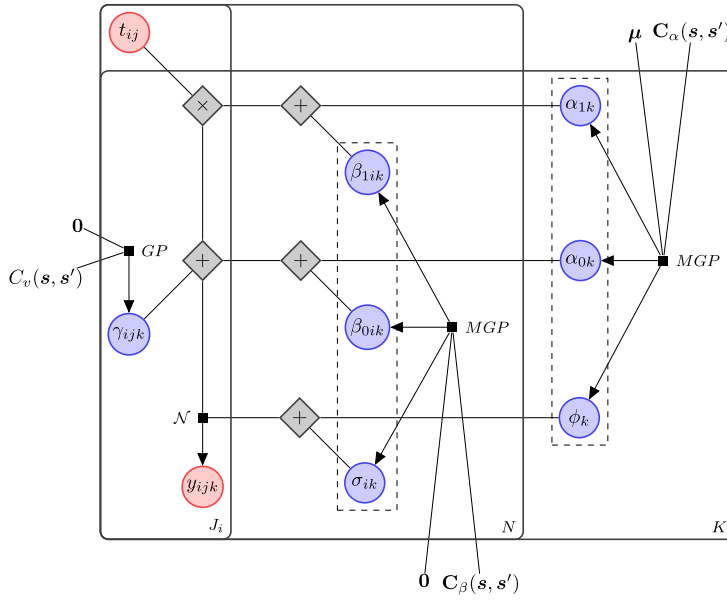


FIG. 6. Plate diagram of the proposed model. Blue nodes are latent variables, red nodes are observed variables, gray nodes are deterministic nodes, GP stands for Gaussian process, and MGP stands for multivariate Gaussian process. Plates are used to group variables repeated together over subjects, time, and space, where $i = 1, \dots, N$ indexes subjects, $j = 1, \dots, J_i$ indexes subject i 's visits, and $k = 1, \dots, K$ indexes superpixel locations.

$\ell_{\alpha,pq}$. 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 (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 (1). The 3×3 cross-correlation matrix \mathbf{R}_{α} 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 $C_{\beta,pp}(\mathbf{s}_k, \mathbf{s}_{k'}) = \sigma_{\beta,pp}^2 M(h | \nu_{\beta,p}, \ell_{\beta,p})$ for $p = 1, \dots, 3$ have RE marginal variances $\sigma_{\beta,pp}^2$, 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}_{β} as defined in (1) and (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 $C_v(\mathbf{s}_k, \mathbf{s}_{k'}) = \sigma_v^2 M(\mathbf{h} | \nu_v, \ell_v)$.

3.4. Priors. We use weakly informative priors to keep inferences within a reasonable range and allow computations to proceed satisfactorily. We rescale the distance between superpixels such that the largest distance between any two superpixels is one unit. The closest any two superpixels can be is ≈ 0.14 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 $\ell_{\alpha,1}, \ell_{\alpha,2}, \ell_{\alpha,3}, \ell_{\beta,1}, \ell_{\beta,2}, \ell_{\beta,3}, \ell_v \sim \text{IG}(3, 1)$ with mean 0.5 and SD 0.5. We evaluate the sensitivity of the SHREVE model to lengthscale priors in Supplementary Material 2 (Su et al. (2024)). Using uniform priors for lengthscales results in almost identical posterior means and intervals for the subject-level MGP lengthscale parameters but larger posterior means and intervals for the population-level MGP lengthscale parameters. For the MGP SD parameters, we wish to avoid flat priors that could pull the posterior toward extreme values.

We assign truncated-normal priors on all MGP SD parameters $\sigma_{\alpha,11}, \sigma_{\beta,11} \sim N^+(0, 15^2)$, $\sigma_{\alpha,22}, \sigma_{\alpha,33}, \sigma_{\beta,22}, \sigma_{\beta,33}, \sigma_v \sim N^+(0, 2.5^2)$, where $N^+(a, b)$ is a normal distribution with mean a and variance b restricted to the positive real line. We assign independent normal priors on the global effects $\mu_0 \sim N(77, 20^2)$, $\mu_1 \sim N(-0.47, 1.4^2)$, $\mu_\phi \sim N(0.5, 0.5^2)$. These priors cover the range of plausible values based on a review of the ophthalmology literature. Studies report GCC thicknesses ranging from 55 to 111 μm in glaucoma patients and from 55 to 71 μm in late-stage glaucoma (Tan et al. (2008), Leung et al. (2013), Nishida et al. (2022), Ghita et al. (2023)). The average rates of change range from -1.1 to -0.14 $\mu\text{m}/\text{year}$, with variability depending on the OCT system used (Leung et al. (2013), Holló and Naghizadeh (2015), Zhang et al. (2016b)). The residual SDs range from 0.8 to 2.6 μm (Tan et al. (2008), Holló and Naghizadeh (2015)).

For the correlation matrices \mathbf{R}_α and \mathbf{R}_β , 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, McCulloch and Meng (2000)). When Σ has a standard inverse-Wishart distribution, we can decompose $\Sigma = \mathbf{S}\mathbf{R}\mathbf{S}$ 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 $\nu_{\alpha,1}, \nu_{\alpha,2}, \nu_{\alpha,3}, \nu_{\beta,1}, \nu_{\beta,2}, \nu_{\beta,3}, \nu_v = \frac{1}{2}$ 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 $M(\mathbf{h}|\frac{1}{2}, \ell) = \exp(-\|\mathbf{h}\|/\ell)$.

3.5. 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 (2004)) implemented in NIMBLE v0.13.0 (de Valpine et al. (2017)). 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_v and subject-level residual SD GP SD parameter $\sigma_{\beta,33}$ together using the automated factor slice sampler (Tibbitts 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 spatial 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 spatial sub-blocks of size 3, a total of 12 parameters for each sampler. We separately sample the subject-level residual SDs in spatial sub-blocks of size 6 and the remaining visit effects in spatial 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 $(\sigma_{\beta,33}, \ell_v)$ and $(\sigma_v, \ell_{\beta,3})$. We run all models with nine chains of 250,000 iterations after a burn-in of 30,000, a thin of 100 for a total of 19,800 posterior samples. We provide the NIMBLE model code for the SHREVE model as an R script in Supplementary Material 1.

3.6. Model comparison. We fit the SHREVE model to the AGPS data and compare model fit of the SHREVE model to seven nested models, to a CAR model fit separately for each eye, and to SLR fit separately for each subject and superpixel location. The seven 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 CAR we run a separate model for each eye with intrinsic CAR priors inspired by the model developed by [Betz-Stablein et al. \(2013\)](#) for visual field data. We provide further details on the CAR model in Supplementary Material 2. 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 \(2010\)](#), [Gelman et al. \(2014\)](#)) and approximate leave-one-out cross-validation (LOO) using Pareto Smoothed Importance Sampling ([Vehtari, Gelman and Gabry \(2017\)](#)). We report WAIC

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

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 s th iteration for $s = 1, \dots, S$ posterior samples, and $V_{s=1}^S$ denotes the sample variance over S posterior samples. We report approximate LOO

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

where w_{ijk}^s , $s = 1, \dots, S$ is a vector of importance weights for data point y_{ijk} at iteration s and $w_{ijk}^s = (p(y_{ijk} | \theta^s))^{-1}$, 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

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

for $s = 1, \dots, S$ posterior MCMC samples, $i = 1, \dots, n$ 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 $\hat{y}_{iJ_{ik}}^s$, of N_{pred} total held out observations after fitting the models. In the first prediction scenario, we randomly sample and hold out seven observations $y_{iJ_{ik}}$, or approximately 20%, at the last visit for each of 110 subjects and six observations for one subject that only has 32 observations available at the last visit, for a total of $N_{\text{pred}} = 111 \times 7 - 1 = 776$ observations, and fit models with the remaining observations. In the second prediction scenario, we hold out all observations $y_{iJ_{ik}}$ at the last visit for each of the 111 subjects, for a total of $N_{\text{pred}} = 3974$ 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.

In the first prediction scenario, we define a predicted observation at each posterior sample s as

$$(3) \quad \hat{y}_{iJ_{ik}}^s = \alpha_{0k}^s + \alpha_{1k}^s t_{iJ_i} + \beta_{0ik}^s + \beta_{1ik}^s t_{iJ_i} + \gamma_{iJ_{ik}}^s,$$

where t_{iJ_i} is the time observed and $\gamma_{iJ_{ik}}^s$ is the visit effect for the held out observation at the i th subject's last visit for the SHREVE models. For the SHRE models, there is no $\gamma_{iJ_{ik}}^s$ visit effect term in (3). In the second prediction scenario, we define a predicted observation as

$$\hat{y}_{iJ_{ik}}^s = \alpha_{0k}^s + \alpha_{1k}^s t_{iJ_i} + \beta_{0ik}^s + \beta_{1ik}^s t_{iJ_i}$$

for both SHREVE and SHRE models.

4. Simulation results. We conduct a simulation study to assess the performance of the SHREVE model in comparison to seven nested models, a CAR model fit separately on data from each eye, and SLR fit separately for each subject and location as described in Section 3.6. We evaluate how well each model estimates the subject-superpixel intercepts $\alpha_{0k} + \beta_{0ik}$ and slopes $\alpha_{1k} + \beta_{1ik}$. We implement a simulation scenario with a sample size of 50 subjects and a 5×5 grid of 25 superpixel locations. We rescale the distances between superpixels such that the largest distance between any two superpixels is one unit.

To compare the estimation accuracy of subject-superpixel intercepts and slopes, we generate data with the following setup. First, we generate a set of global parameters (μ_0, μ_1, μ_ϕ) from normal distributions and GP and MGP hyperparameters from uniform distributions. Then we draw a set of population-level parameters (α_k) and subject-level parameters (β_{ik}), given the generated hyperparameters. The generated $\alpha_{0k} + \beta_{0ik}$ and $\alpha_{1k} + \beta_{1ik}$ serve as the true subject-superpixel intercepts and slopes across 100 simulation runs. For each of 100 simulated data sets, we generate GCC outcomes by introducing random measurement error ϵ_{ijk} from a normal distribution with residual SD $\exp(\phi_k + \sigma_{ik})$ and random visit effects γ_{ijk} . We provide R code for the simulation study in Supplementary Material 1 and present further details regarding data generation in Supplementary Material 2.

We evaluate the accuracy of the estimates for the $50 \times 25 = 1250$ possible subject-superpixel intercepts and slopes. For each simulation run, we take the posterior mean of $\alpha_{0k} + \beta_{0ik}$ and $\alpha_{1k} + \beta_{1ik}$ to be the model estimates of intercepts and slopes, respectively. For each intercept and slope, we record the absolute bias $|(\sum_{m=1}^{100} (\text{estimate}_m - \text{truth}))/100|$, 95% credible interval (CrI) coverage probability, 95% credible interval length (CrIL), and root mean squared error $\text{RMSE} = \sqrt{(\sum_{m=1}^{100} (\text{estimate}_m - \text{truth})^2)/100}$ where m indexes the data set.

Supplementary Material Table S1 displays results for the intercepts. We provide the mean, 2.5% quantile, and 97.5% quantile for each metric across the 1250 intercepts and slopes. For intercepts the SHREVE model has the smallest absolute bias, while SLR has the largest absolute bias (mean: 1.13 vs. 1.48). On average, all models with visits effects and SLR have appropriate 95% CrI coverage, while the models without visit effects and CAR have slightly lower coverage (range: 0.93 to 0.94). The distribution of coverage probabilities across subject-superpixels is more variable for the hierarchical models than for SLR. On average, SLR has much larger CrIL than all other models, up to 60% wider than the SHREVE model. On average, the SHREVE model has the smallest RMSE, while SLR has the largest RMSE (1.40 vs. 1.84).

Table 1 presents results for the slopes. For slopes the SHREVE model has the smallest absolute bias, while SLR has the largest absolute bias (mean: 0.44 vs. 0.71). On average, all models with visits effects and SLR have appropriate 95% CrI coverage, while the models without visit effects and CAR have slightly lower coverage (range: 0.90 to 0.93). The distribution of coverage probabilities across subject-superpixels are more variable for the hierarchical models than for SLR, suggesting some slopes having anticonservative posterior SDs. On average, SLR has much larger CrIL than all other models, up to 110% wider than the SHREVE model. The range of CrILs for SLR is also much wider than the other models; the 97.5% quantile CrIL for SLR is 12.15 while the second largest is 3.89 from the CAR model. On average, the SHREVE model has the smallest RMSE, while SLR has the largest RMSE (0.52 vs. 0.89). These results provide evidence that the SHREVE model offers markedly improved performance in estimating subject-superpixel intercepts and slopes when compared to SLR.

5. 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

TABLE 1

Summary of the mean, 2.5% quantile, and 97.5% quantile absolute bias (Abs Bias), 95 % credible interval coverage probability (Coverage), 95 % credible interval length (CrIL), and root mean squared error (RMSE) across the 1250 subject-location slopes in the simulation study. The SHREVE model has the smallest absolute bias on average, appropriate 95% credible interval coverage, and the smallest RMSE. The smallest absolute bias, largest coverage probability, smallest CrIL, and smallest RMSE are bolded

Model	Abs Bias Summary			Coverage Summary			CrIL Summary			RMSE Summary		
	Mean	2.5%	97.5%	Mean	2.5%	97.5%	Mean	2.5%	97.5%	Mean	2.5%	97.5%
SHREVE	0.44	0.26	0.96	0.95	0.68	1.00	2.13	1.43	2.91	0.52	0.32	1.05
SHREVE-(a)	0.44	0.26	0.96	0.95	0.70	1.00	2.15	1.50	2.92	0.53	0.32	1.05
SHREVE-(b)	0.46	0.24	1.07	0.94	0.68	1.00	2.21	1.50	2.96	0.54	0.30	1.15
SHREVE-(ab)	0.49	0.21	1.06	0.95	0.63	1.00	2.42	2.11	2.90	0.58	0.27	1.16
SHRE	0.45	0.26	1.01	0.90	0.55	1.00	1.89	1.27	2.62	0.54	0.33	1.11
SHRE-(a)	0.45	0.26	1.01	0.90	0.56	1.00	1.91	1.34	2.65	0.55	0.33	1.10
SHRE-(b)	0.47	0.25	1.05	0.90	0.54	1.00	1.97	1.33	2.65	0.56	0.31	1.16
SHRE-(ab)	0.50	0.23	1.05	0.92	0.56	1.00	2.21	1.90	2.72	0.59	0.29	1.18
CAR	0.52	0.25	1.06	0.93	0.68	1.00	2.49	1.66	3.89	0.64	0.31	1.31
SLR	0.71	0.30	2.03	0.95	0.90	0.99	4.38	1.76	12.15	0.89	0.37	2.54

36 superpixels. Following Vehtari et al.'s (2021) recommendation for assessing convergence, the bulk and tail effective sample sizes were all greater than 100 per chain, and the potential scale reduction factor \hat{R} were all less than 1.01. Visual assessment of model convergence show satisfactory results. We show efficiency per iteration plots of the seven parameters with the largest \hat{R} in Supplementary Material Figure S1 and summarize convergence diagnostics in Supplementary Material Table S3. The total runtime for the 250,000 iterations is approximately 26.4 hours (Apple M1 Pro 10-core CPU) for the SHREVE model.

Table 2 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 pro-

TABLE 2

Model fit comparison with widely applicable information criterion (WAIC), approximate leave-one-out cross-validation with Pareto smoothed importance sampling (LOO), mean squared prediction error (MSPE) of predictions, 95% prediction interval coverage probabilities (Cov %), and mean 95% prediction interval length (PIL). For Scenario 1 we hold out seven randomly sampled observations $y_{i,j,k}$ at the last visit of each of 110 AGPS subjects and 6 observations from one subject. For Scenario 2 we hold out all observations $y_{i,j,k}$ at the last visit of all 111 AGPS subjects. The smallest WAIC, LOO, PIL, and MSPE and largest prediction coverage values are bolded

Model	WAIC	LOO	Scenario 1			Scenario 2		
			MSPE	Cov %	PIL	MSPE	Cov %	PIL
SHREVE	107,608.9	113,303.8	6.9	77.7	4.99	9.1	77.2	5.60
SHREVE-(a)	107,942.6	113,521.5	6.9	77.8	5.00	9.0	77.3	5.62
SHREVE-(b)	110,985.2	116,907.91	6.8	81.8	5.47	9.4	80.3	6.09
SHREVE-(ab)	113,259.2	118,616.4	6.9	83.2	5.60	9.5	81.0	6.22
SHRE	124,388.8	125,293.7	7.2	67.8	4.38	9.1	65.2	4.48
SHRE-(a)	124,469.2	125,456.6	7.1	68.2	4.39	9.0	65.6	4.51
SHRE-(b)	129,357.6	129,881.6	7.5	73.3	4.92	9.7	71.1	4.99
SHRE-(ab)	130,182.1	130,708.7	7.5	75.4	5.03	9.7	71.5	5.09
CAR	126,691.9	127,390.6	7.7	71.5	5.00	10.3	69.1	5.00
SLR	128,870.2	132,916.3	39.7	83.9	9.73	52.9	83.7	9.70

cess, omitting: (a) increases WAIC (LOO) by up to 334 (218) while omitting (b) increases WAIC (LOO) by up to 4969 (4588). Omitting visit effects increases WAIC (LOO) by up to 18,372 (12,974). SLR has lower WAIC than the two SHRE models without (b), 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 1593 (984) than the difference between SHREVE and SHREVE-(b).

For predictions in the first scenario, the MSPE for SLR is 5.7 times that of the SHREVE model (39.7 vs. $6.9 \mu\text{m}^2$) and 5.5 times that of the SHRE model (39.7 vs. $7.2 \mu\text{m}^2$). Among the hierarchical models, the biggest distinction in MSPE is between models with and without visit effects. The 95% prediction interval coverage probability is lower for the SHREVE model than SLR (77.7% vs. 83.9%), although none of the models achieve appropriate 95% coverage. Models without visit effects and the CAR model have noticeably lower coverage probabilities ranging from 67.8% to 75.4%. On average, SLR has the largest prediction interval length (PIL), almost double that of the SHREVE model (9.73 vs. 4.99).

For predictions in the second scenario where the last visit for all subjects is held out, the MSPE for SLR is 5.8 times that of the SHREVE model and SHRE model (52.9 vs. $9.1 \mu\text{m}^2$). The 95% prediction interval coverage probability is lower for the SHREVE model than SLR (77.2% vs. 83.7%), although none of the models achieve appropriate 95% coverage. 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. Similar to Scenario 1, SLR has the largest PIL, almost double that of the SHREVE model (9.70 vs. 5.60).

Figure 7 plots profiles and posterior mean fitted lines from the SHREVE model and SLR for one subject for six superpixels that had the last (seventh) observation held out in the second prediction scenario. The SHREVE model better estimates slopes for noisy superpixels, like 4.3 and 5.7. All predictions of the last visit in the six superpixels by the SHREVE model are closer to the GCC observed at $t_{ij} = 3.6$ than those by SLR.

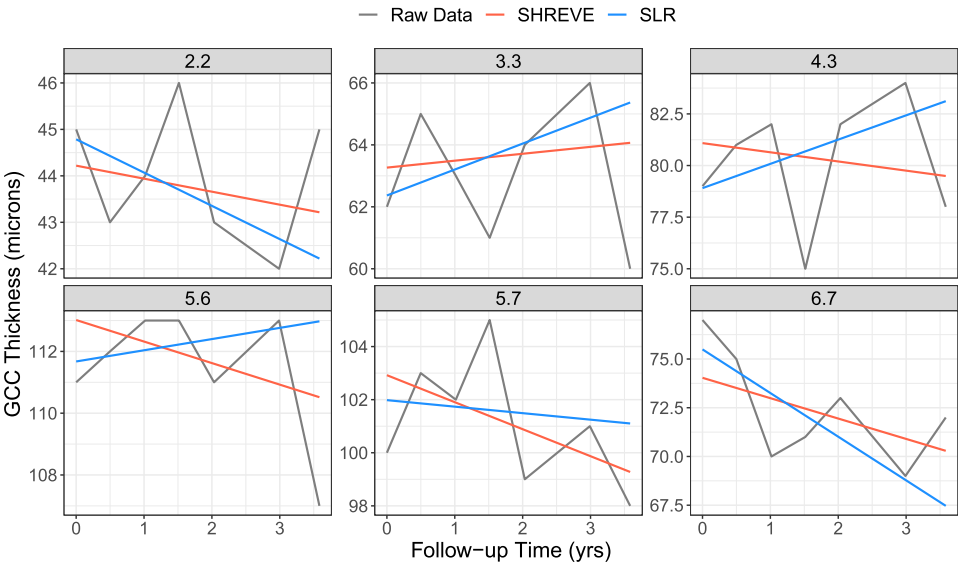


FIG. 7. Comparison of predicted observations and model fit from the SHREVE model and simple linear regression (SLR) after holding out the last observation at 3.6 years follow-up of this subject. The gray line plots the raw data, the red line is the posterior mean fitted line from the SHREVE model without 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 4.3 and 5.7, than SLR.

TABLE 3

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

Parameters	Symbols	SHREVE Model		SHRE Model	
		Mean	95% CrI	Mean	95% CrI
Global Parameters					
Intercept	μ_0	70.48	(51.43, 89.09)	71.13	(52.87, 89.48)
Slope	μ_1	−0.26	(−0.72, 0.27)	−0.26	(−0.76, 0.31)
Log Residual SD	μ_ϕ	0.25	(−0.16, 0.77)	0.63	(0.30, 0.97)
Subject-Level MGP SD Parameters					
Intercept	$\sigma_{\beta,11}$	16.21	(15.12, 17.45)	16.36	(15.27, 17.63)
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	$\ell_{\beta,1}$	0.77	(0.66, 0.90)	0.79	(0.68, 0.93)
Slope	$\ell_{\beta,2}$	0.59	(0.48, 0.73)	0.96	(0.77, 1.19)
Log Residual SD	$\ell_{\beta,3}$	0.27	(0.22, 0.32)	0.53	(0.43, 0.65)
Subject-Level MGP Correlation Parameters					
Intercept/Slope	$\rho_{\beta,12}$	−0.15	(−0.19, −0.10)	−0.13	(−0.18, −0.08)
Intercept/Log Residual SD	$\rho_{\beta,13}$	0.15	(0.10, 0.20)	0.17	(0.12, 0.23)
Slope/Log Residual SD	$\rho_{\beta,23}$	−0.24	(−0.32, −0.17)	−0.26	(−0.34, −0.18)
Visit Effect Parameters					
Lengthscale	ℓ_v	0.50	(0.44, 0.58)		
SD	σ_v	1.42	(1.37, 1.48)		

Table 3 gives posterior means and 95% CrI for parameters of interest from 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 Material Table S8 gives posterior means and 95% CrIs for the population-level MGP parameters. The population-level MGP parameters are similar between the two models.

Figure 8 plots spatial correlations $M(h)$ as a function of distance h between superpixels for the SHREVE and SHRE models. At 1.0 units distance, the spatial correlation of subject-specific slopes drops to $\exp(-1.69) \approx 0.18$ for the SHREVE model but is $\exp(-1.04) \approx 0.35$ for the SHRE model. At 1.0 units distance, the spatial correlation of subject-specific log-residual SDs is 0.02 for the SHREVE model but around 0.15 for the SHRE model. The shorter lengthscales in the SHREVE model result in markedly reduced correlations at the same distance between superpixels.

Figure 9 presents heatmaps 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, 4.5, 5.4, and 5.5) and superpixels in the seventh column have higher log-residual SDs and have smaller differences in log-residual SDs between the models. SHREVE

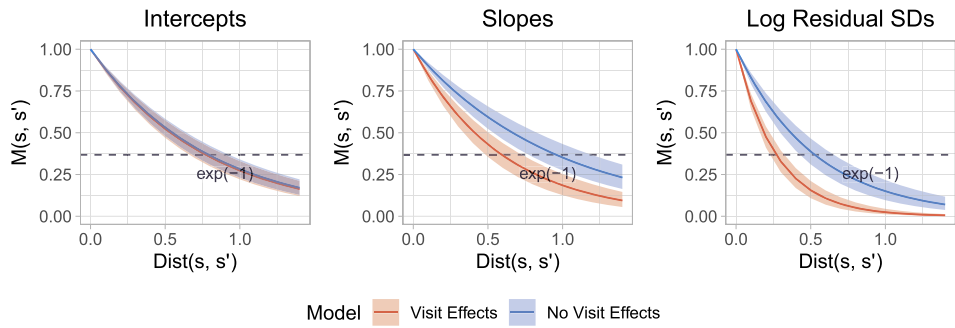


FIG. 8. Posterior mean (line) and 95% pointwise credible intervals (colored bands) of correlation as a function of the distance h between superpixels for subject-specific intercepts, slopes, and log-residual SDs from the SHREVE (Visit Effects) and SHRE (No Visit Effects) models. The correlations decay 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.

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 3990 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 10 shows the proportion of significant negative slopes by superpixel, and Supplementary Material Figure S2 shows the proportion of significant positive slopes by superpixel. The SHREVE model detects 10% more significant negative slopes in six of 36 superpixels and 5% less significant positive slopes in five 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 worsening slopes and

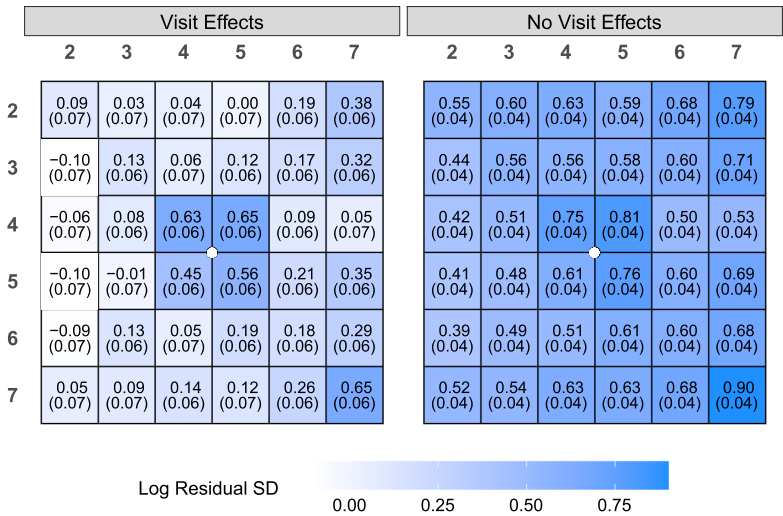


FIG. 9. Heatmap of the log-residual standard deviations (SD) comparing the SHREVE (Visit Effects) and SHRE (No Visit Effects) models. The values shown are the posterior mean (posterior SD) across the 36 superpixels. The log-residual SDs from the SHREVE model are uniformly reduced across all superpixels compared to those from the SHRE model. The white dot is the fovea.

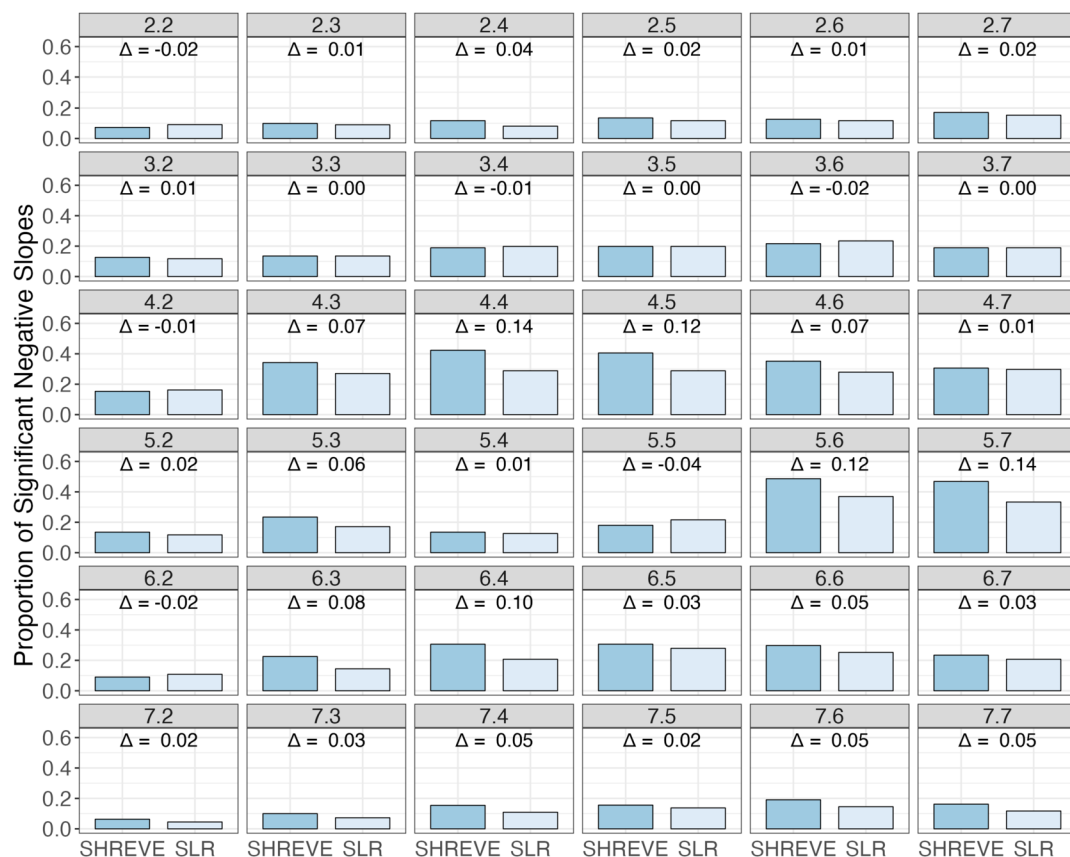


FIG. 10. Bar charts of the proportion of significant negative slopes detected by the SHREVE model and simple linear regression (SLR) across the 36 superpixels. The difference ($\Delta = \text{SHREVE} - \text{SLR}$) in proportion is labeled at the top of each subplot. Across all locations, the SHREVE model detects a higher proportion of significant negative slopes (21.4% vs. 18.0%) than SLR.

possibly reduces false positive rates as compared to SLR. Supplementary Material Figure S3 shows a heatmap of posterior means of population-level slopes from the SHREVE model. Because the SHREVE model allows for inference on population parameters, clinicians will be able to examine covariates that may influence progression across the macula. Slopes corresponding to superpixels around the fovea are significantly negative and much steeper than superpixels around the outer edge of the macular area.

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

Using information from the whole cohort to improve estimation of subject-specific slopes may be counterintuitive to clinicians. Here we show that leveraging information from multi-

ple subjects and locations facilitates more accurate estimation of subject-specific slopes. This is highlighted by the simulation study, where SLR has the highest RMSE of all models compared and is, on average, 70% higher than the SHREVE model (0.89 vs. 0.52 μm). Another benefit of using data from multiple subjects is the ability to make inferences on population parameters. This model allows clinicians to examine the associations between covariates such as age, ethnicity, gender, treatment, intraocular pressure, and blood pressure measures and loss of GCC thickness. This would not be possible using simpler models that fit data separately for each eye or for each eye-superpixel, as currently done in practice.

We make several modeling assumptions. There are many sources of error in obtaining GCC thickness measurements from OCT scans. We remove obvious outliers prior to modeling and find Gaussian errors are appropriate. We present additional model comparisons in Supplementary Material 2 to test our assumption of normality, where we show that a model with Gaussian errors gives better fit than one with t-distributed errors. Within a sensitivity analysis in Supplementary Material 2, we find our model to be robust to the removal of outliers. 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. In this way we do not assume measurement errors to be independent and identically distributed in space. In a previous study, [Mohammadzadeh et al. \(2021\)](#) found no evidence for an autoregressive variance structure, suggesting residuals are not correlated in time after accounting for random intercepts and slopes. Hence, we do not include an autoregressive variance structure and instead assume errors are independent and identically distributed in time. We treat time as a variable to estimate slopes of progression and do not assume further temporal dependence, modeling random effects as strictly spatial processes. The multivariate Matérn covariance functions we use to model the random effects have distinct lengthscales for each marginal process and are nonseparable. Alternatively, we can employ separable multivariate Matérn covariance functions by assuming all lengthscales are the same across marginal processes. Noteworthy, we do find evidence of distinct lengthscales for each marginal process, supporting our use of nonseparable MGPs. Additionally, we assume covariance functions are stationary and isotropic. To test our modeling assumptions of stationarity and isotropy, we provide additional model comparisons in Supplementary Material 2. While nonstationary covariance functions improve fit, we show that there is little practical benefit due to the small improvement. Despite these limitations, our approach does help identify progression of glaucoma for more individualized treatment plans, especially in comparison to SLR.

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. \(2013\)](#), [Berchuk, Mwanza and Warren \(2019\)](#)). One could consider multivariate CAR models as an alternative to the MGPs used in the SHREVE model. [Jin, Carlin and Banerjee \(2005\)](#) propose a generalized multivariate CAR to overcome challenges in specifying joint multivariate distributions with positive definite covariance matrices through the specification of simpler conditionals and marginal forms. However, such models suffer from the conditional specification imposing an arbitrary ordering on the variables being modeled. More recently, [MacNab \(2016a,b\)](#) introduces a framework for coregionalized multivariate CAR models, including a new class of order-free models that allow spatial interaction parameters and coregionalization coefficients to remain identifiable. Future research comparing order-free multivariate CAR models to the MGPs in the SHREVE model for this application would be of great interest.

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, [Rasmussen and Williams 2006](#)). The exponential kernel balances model fit with computational efficiency. 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 should be considered ([Banerjee et al. \(2008\)](#)). Specifically, to provide inference when the number of spatial locations is in the thousands, we could employ nearest-neighbor Gaussian processes (NNGP) as sparsity-inducing spatial priors ([Datta et al. \(2016\)](#)). The computational burden of NNGPs scales linearly with the number of locations, offering substantial scalability, while allowing fully process-based modeling. Nonetheless, our 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.

Acknowledgments. This work used computational and storage services associated with the Hoffman2 Shared Cluster provided by UCLA Office of Advanced Research Computing's Research Technology Group.

Funding. This work was supported by an NIH R01 grant (R01-EY029792), an unrestricted Departmental Grant from Research to Prevent Blindness, and an unrestricted grant from Heidelberg Engineering.

AJH was supported by NIH Grant K25 AI153816, NSF Grant DMS 2152774, and a generous gift from the Karen Toffler Charitable Trust.

SUPPLEMENTARY MATERIAL

Supplementary material 1 (DOI: [10.1214/24-AOAS1944SUPPA](https://doi.org/10.1214/24-AOAS1944SUPPA); .zip). Code for the simulation study and the SHREVE model using R package NIMBLE.

Supplementary material 2 (DOI: [10.1214/24-AOAS1944SUPPB](https://doi.org/10.1214/24-AOAS1944SUPPB); .pdf). Additional details and results for the simulation studies, sensitivity analyses, and analysis of AGPS data.

REFERENCES

- ABRAMOWITZ, M. and STEGUN, I. A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. National Bureau of Standards Applied Mathematics Series, No. 55. U. S. Government Printing Office, Washington, DC. [MR0167642](#)
- 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. *J. Amer. Statist. Assoc.* **107** 180–193. [MR2949350](#) <https://doi.org/10.1080/01621459.2011.643197>
- BANERJEE, S., CARLIN, B. P. and GELFAND, A. E. (2015). *Hierarchical Modeling and Analysis for Spatial Data*, 2nd ed. *Monographs on Statistics and Applied Probability* **135**. CRC Press, Boca Raton, FL. [MR3362184](#)
- BANERJEE, S., GELFAND, A. E., FINLEY, A. O. and SANG, H. (2008). Gaussian predictive process models for large spatial data sets. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 825–848. [MR2523906](#) <https://doi.org/10.1111/j.1467-9868.2008.00663.x>
- BARNARD, J., MCCULLOCH, R. and MENG, X.-L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statist. Sinica* **10** 1281–1311. [MR1804544](#)
- BERCHUK, S. I., MWANZA, J.-C. and WARREN, J. L. (2019). Diagnosing glaucoma progression with visual field data using a spatiotemporal boundary detection method. *J. Amer. Statist. Assoc.* **114** 1063–1074. [MR4011758](#) <https://doi.org/10.1080/01621459.2018.1537911>

- 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. *Investig. Ophthalmol. Vis. Sci.* **54** 1544–1553.
- BOGACHEV, V. I. (1998). *Gaussian Measures. Mathematical Surveys and Monographs* **62**. Amer. Math. Soc., Providence, RI. [MR1642391 https://doi.org/10.1090/surv/062](https://doi.org/10.1090/surv/062)
- 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. *Investig. Ophthalmol. Vis. Sci.* **56** 4283–4289.
- BRYAN, S. R., EILERS, P. H. C., VAN ROSMALEN, J., RIZOPOULOS, D., VERMEER, K. A., LEMIJ, H. G. and LESAFFRE, E. M. E. H. (2017). Bayesian hierarchical modeling of longitudinal glaucomatous visual fields using a two-stage approach. *Stat. Med.* **36** 1735–1753. [MR3648619 https://doi.org/10.1002/sim.7235](https://doi.org/10.1002/sim.7235)
- 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** 823–833. [MR3860703 https://doi.org/10.1111/biom.12844](https://doi.org/10.1111/biom.12844)
- DATTA, A., BANERJEE, S., FINLEY, A. O. and GELFAND, A. E. (2016). Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *J. Amer. Statist. Assoc.* **111** 800–812. [MR3538706 https://doi.org/10.1080/01621459.2015.1044091](https://doi.org/10.1080/01621459.2015.1044091)
- DE VALPINE, P., TUREK, D., PACIOREK, C. J., ANDERSON-BERGMAN, C., TEMPLE LANG, D. and BODIK, R. (2017). Programming with models: Writing statistical algorithms for general model structures with NIMBLE. *J. Comput. Graph. Statist.* **26** 403–413. [MR3640196 https://doi.org/10.1080/10618600.2016.1172487](https://doi.org/10.1080/10618600.2016.1172487)
- GARDINER, S. K. and CRABB, D. P. (2002). Examination of different pointwise linear regression methods for determining visual field progression. *Investig. Ophthalmol. Vis. Sci.* **43** 1400–1407.
- GASPARI, G. and COHN, S. E. (1999). Construction of correlation functions in two and three dimensions. *Q. J. R. Meteorol. Soc.* **125** 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. *Ann. Appl. Stat.* **8** 1095–1118. [MR3262547 https://doi.org/10.1214/14-AOAS718](https://doi.org/10.1214/14-AOAS718)
- GELFAND, A. E., DIGGLE, P. J., FUENTES, M. and GUTTORP, P., eds. (2010) *Handbook of Spatial Statistics. Chapman & Hall/CRC Handbooks of Modern Statistical Methods*. CRC Press, Boca Raton, FL. [MR2761512 https://doi.org/10.1201/9781420072884](https://doi.org/10.1201/9781420072884)
- GELFAND, A. E., KIM, H.-J., SIRMANS, C. F. and BANERJEE, S. (2003). Spatial modeling with spatially varying coefficient processes. *J. Amer. Statist. Assoc.* **98** 387–396. [MR1995715 https://doi.org/10.1198/0162145030000170](https://doi.org/10.1198/0162145030000170)
- GELFAND, A. E. and SCHLIEP, E. M. (2016). Spatial statistics and Gaussian processes: A beautiful marriage. *Spat. Stat.* **18** 86–104. [MR3573271 https://doi.org/10.1016/j.spasta.2016.03.006](https://doi.org/10.1016/j.spasta.2016.03.006)
- GELMAN, A., CARLIN, J. B., STERN, H. S., DUNSON, D. B., VEHTARI, A. and RUBIN, D. B. (2014). *Bayesian Data Analysis*, 3rd ed. *Texts in Statistical Science Series*. CRC Press, Boca Raton, FL. [MR3235677 https://doi.org/10.1201/9781420072884](https://doi.org/10.1201/9781420072884)
- GENTON, M. G. and KLEIBER, W. (2015). Cross-covariance functions for multivariate geostatistics. *Statist. Sci.* **30** 147–163. [MR3353096 https://doi.org/10.1214/14-STS487](https://doi.org/10.1214/14-STS487)
- GHITA, A. M., ILIESCU, D. A., GHITA, A. C., ILIE, L. A. and OTOBIC, A. (2023). Ganglion cell complex analysis: Correlations with retinal nerve fiber layer on optical coherence tomography. *Diagnostics* **13** 266.
- GNEITING, T., KLEIBER, W. and SCHLATHER, M. (2010). Matérn cross-covariance functions for multivariate random fields. *J. Amer. Statist. Assoc.* **105** 1167–1177. [MR2752612 https://doi.org/10.1198/jasa.2010.tm09420](https://doi.org/10.1198/jasa.2010.tm09420)
- GÖSSL, C., AUER, D. P. and FAHRMEIR, L. (2001). Bayesian spatiotemporal inference in functional magnetic resonance imaging. *Biometrics* **57** 554–562. [MR1855691 https://doi.org/10.1111/j.0006-341X.2001.00554.x](https://doi.org/10.1111/j.0006-341X.2001.00554.x)
- GUTTORP, P. and GNEITING, T. (2006). Studies in the history of probability and statistics XLIX On the Matérn correlation family. *Biometrika* **93** 989–995. [MR2285084 https://doi.org/10.1093/biomet/93.4.989](https://doi.org/10.1093/biomet/93.4.989)
- HASTIE, T. and TIBSHIRANI, R. (1993). Varying-coefficient models. *J. Roy. Statist. Soc. Ser. B* **55** 757–796. [MR1229881 https://doi.org/10.1214/aos/1176324641](https://doi.org/10.1214/aos/1176324641)
- HOLLÓ, G. and NAGHIZADEH, F. (2015). Influence of a new software version of the RTVue-100 optical coherence tomograph on the detection of glaucomatous structural progression. *Eur. J. Ophthalmol.* **25** 410–415. <https://doi.org/10.5301/ejo.5000576>
- JIN, X., CARLIN, B. P. and BANERJEE, S. (2005). Generalized hierarchical multivariate CAR models for areal data. *Biometrics* **61** 950–961. [MR2216188 https://doi.org/10.1111/j.1541-0420.2005.00359.x](https://doi.org/10.1111/j.1541-0420.2005.00359.x)
- KIM, H. and LEE, J. (2017). Hierarchical spatially varying coefficient process model. *Technometrics* **59** 521–527. [MR3740968 https://doi.org/10.1080/00401706.2017.1317290](https://doi.org/10.1080/00401706.2017.1317290)
- KINGMAN, S. (2004). Glaucoma is second leading cause of blindness globally. *Bull. World Health Organ.* **82** 887–888.
- LEUNG, C. K., YE, C., WEINREB, R. N., YU, M., LAI, G. and LAM, D. S. (2013). Impact of age-related change of retinal nerve fiber layer and macular thicknesses on evaluation of glaucoma progression. *Ophthalmology* **120** 2485–2492.

- 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. *Stat. Methods Med. Res.* **28** 1203–1215. [MR3934644 https://doi.org/10.1177/0962280217752378](https://doi.org/10.1177/0962280217752378)
- MACNAB, Y. C. (2016a). Linear models of coregionalization for multivariate lattice data: A general framework for coregionalized multivariate CAR models. *Stat. Med.* **35** 3827–3850. [MR3538050 https://doi.org/10.1002/sim.6955](https://doi.org/10.1002/sim.6955)
- MACNAB, Y. C. (2016b). Linear models of coregionalization for multivariate lattice data: Order-dependent and order-free cMCARs. *Stat. Methods Med. Res.* **25** 1118–1144. [MR3541088 https://doi.org/10.1177/0962280216660419](https://doi.org/10.1177/0962280216660419)
- MATÉRN, B. (1986). *Spatial Variation*, 2nd ed. *Lecture Notes in Statistics* **36**. Springer, Berlin. [MR0867886 https://doi.org/10.1007/978-1-4615-7892-5](https://doi.org/10.1007/978-1-4615-7892-5)
- METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. and TELLER, E. (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21** 1087–1092.
- MIRAFTEBI, 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. *Transl. Vis. Sci. Technol.* **5** 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. *Surv. Ophthalmol.* **65** 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** 888–900.
- MOHAMMADZADEH, V., SU, E., RABIOLO, A., SHI, L., ZADEH, S. H., LAW, S. K., COLEMAN, A. L., CAPRIOLI, J., WEISS, R. E. et al. (2022a). Ganglion cell complex: The optimal measure for detection of structural progression in the macula. *Am. J. Ophthalmol.* **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. *Ophthalmol. Sci.* **2** 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. *Transl. Vis. Sci. Technol.* **10** 15.
- MONTESANO, G., GARWAY-HEATH, D. F., OMETTO, G. and CRABB, D. P. (2021). Hierarchical censored Bayesian analysis of visual field progression. *Transl. Vis. Sci. Technol.* **10** 4.
- NISHIDA, T., MOGHIMI, S., MOHAMMADZADEH, V., WU, J.-H., YAMANE, M. L., KAMALIPOUR, A., MAHMOUDINEZHAD, G., MICHELETTI, E., LIEBMANN, J. M. et al. (2022). Association between ganglion cell complex thinning and vision-related quality of life in glaucoma. *JAMA Ophthalmol.* **140** 800–806.
- NOURI-MAHDAVI, K., HOFFMAN, D., RALLI, M. and CAPRIOLI, J. (2007). Comparison of methods to predict visual field progression in glaucoma. *Arch. Ophthalmol.* **125** 1176–1181. <https://doi.org/10.1001/archophth.125.9.1176>
- PENNY, W. D., TRUJILLO-BARRETO, N. J. and FRISTON, K. J. (2005). Bayesian fMRI time series analysis with spatial priors. *NeuroImage* **24** 350–362. <https://doi.org/10.1016/j.neuroimage.2004.08.034>
- 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. *Transl. Vis. Sci. Technol.* **9** 50.
- RASMUSSEN, C. E. and WILLIAMS, C. K. I. (2006). *Gaussian Processes for Machine Learning. Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. [MR2514435 https://doi.org/10.1017/9780521876223](https://doi.org/10.1017/9780521876223)
- RISER, M. D. and TUREK, D. (2020). Bayesian inference for high-dimensional nonstationary Gaussian processes. *J. Stat. Comput. Simul.* **90** 2902–2928. [MR4168232 https://doi.org/10.1080/00949655.2020.1792472](https://doi.org/10.1080/00949655.2020.1792472)
- ROBERT, C. P. and CASELLA, G. (2004). *Monte Carlo Statistical Methods*, 2nd ed. *Springer Texts in Statistics*. Springer, New York. [MR2080278 https://doi.org/10.1007/978-1-4757-4145-2](https://doi.org/10.1007/978-1-4757-4145-2)
- SCHMIDT, A. M. and GELFAND, A. E. (2003). A Bayesian coregionalization approach for multivariate pollutant data. *J. Geophys. Res., Atmos.* **108**.
- STONE, M. (1977). An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *J. Roy. Statist. Soc. Ser. B* **39** 44–47. [MR0501454 https://doi.org/10.2307/2344738](https://doi.org/10.2307/2344738)
- SU, E., WEISS, R. E., NOURI-MAHDAVI, K. and HOLBROOK, A. J. (2024). Supplement to “A spatially varying hierarchical random effects model for longitudinal macular structural data in glaucoma patients.” <https://doi.org/10.1214/24-AOAS1944SUPPA>, <https://doi.org/10.1214/24-AOAS1944SUPPB>
- TAN, O., LI, G., LU, A. T.-H., VARMA, R., HUANG, D. and ADVANCED IMAGING FOR GLAUCOMA STUDY GROUP (2008). Mapping of macular substructures with optical coherence tomography for glaucoma diagnosis. *Ophthalmology* **115** 949–956.

- TATHAM, A. J. and MEDEIROS, F. A. (2017). Detecting structural progression in glaucoma with optical coherence tomography. *Ophthalmology* **124** S57–S65. <https://doi.org/10.1016/j.ophtha.2017.07.015>
- R CORE TEAM (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- 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. et al. (2020). Comparing the rule of 5 to trend-based analysis for detecting glaucoma progression on OCT. *Ophthalmol. Glaucoma* **3** 414–420.
- TIBBITS, M. M., GROENDYKE, C., HARAN, M. and LIECHTY, J. C. (2014). Automated factor slice sampling. *J. Comput. Graph. Statist.* **23** 543–563. [MR3215824 https://doi.org/10.1080/10618600.2013.791193](https://doi.org/10.1080/10618600.2013.791193)
- VEHTARI, A., GELMAN, A. and GABRY, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* **27** 1413–1432. [MR3647105 https://doi.org/10.1007/s11222-016-9696-4](https://doi.org/10.1007/s11222-016-9696-4)
- VEHTARI, A., GELMAN, A., SIMPSON, D., CARPENTER, B. and BÜRKNER, P.-C. (2021). Rank-normalization, folding, and localization: An improved \hat{R} for assessing convergence of MCMC (with discussion). *Bayesian Anal.* **16** 667–718. [MR4298989 https://doi.org/10.1214/20-ba1221](https://doi.org/10.1214/20-ba1221)
- VER HOEF, J. M. and BARRY, R. P. (1998). Constructing and fitting models for cokriging and multivariable spatial prediction. *J. Statist. Plann. Inference* **69** 275–294. [MR1631328 https://doi.org/10.1016/S0378-3758\(97\)00162-6](https://doi.org/10.1016/S0378-3758(97)00162-6)
- WACKERNAGEL, H. (2013). *Multivariate Geostatistics*, 3rd ed. Springer, Berlin.
- WATANABE, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J. Mach. Learn. Res.* **11** 3571–3594. [MR2756194](https://doi.org/10.1016/S0140-6736(04)16257-0)
- WEINREB, R. N. and KHAW, P. T. (2004). Primary open-angle glaucoma. *Lancet* **363** 1711–1720. [https://doi.org/10.1016/S0140-6736\(04\)16257-0](https://doi.org/10.1016/S0140-6736(04)16257-0)
- WICKHAM, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer, New York.
- ZHANG, F., JIANG, W., WONG, P. and WANG, J.-P. (2016a). A Bayesian probit model with spatially varying coefficients for brain decoding using fMRI data. *Stat. Med.* **35** 4380–4397. [MR3554969 https://doi.org/10.1002/sim.6999](https://doi.org/10.1002/sim.6999)
- ZHANG, X., FRANCIS, B. A., DASTIRIDOU, A., CHOPRA, V., TAN, O., VARMA, R., GREENFIELD, D. S., SCHUMAN, J. S., HUANG, D. et al. (2016b). Longitudinal and cross-sectional analyses of age effects on retinal nerve fiber layer and ganglion cell complex thickness by Fourier-domain OCT. *Transl. Vis. Sci. Technol.* **5** 1–1.
- ZHU, H., FAN, J. and KONG, L. (2014). Spatially varying coefficient model for neuroimaging data with jump discontinuities. *J. Amer. Statist. Assoc.* **109** 1084–1098. [MR3265682 https://doi.org/10.1080/01621459.2014.881742](https://doi.org/10.1080/01621459.2014.881742)