Bayesian Semiparametric Functional Mixed Models for Serially Correlated Functional Data, with Application to Glaucoma Data

Wonyul Lee,

Correspondence to: Jeffrey S. Morris.

Wonyul Lee is Postdoctoral Researcher, University of Texas M.D. Anderson Cancer Center, Houston, TX 77230 (freshlwy@gmail.com); Michelle Miranda is Postdoctoral Researcher, University of Texas M.D. Anderson Cancer Center, Houston, TX 77230 (MFMiranda@gmail.com); Philip Rausch is Researcher, Humboldt-Universität zu Berlin (philip.rausch@gmail.com); Veera Baladandayuthapani is Associate Professor, University of Texas M.D. Anderson Cancer Center, Houston, TX 77230 (veera@mdanderson.org); Massimo Fazio is Assistant Professor, University of Alabama at Birmingham, Birmingham, AL 35294 (massimof@uab.edu); and J. Crawford Downs is Professor, University of Alabama at Birmingham, Birmingham, AL 35294 (cdowns@uab.edu); Jeffrey S. Morris is Del and Dennis McCarthy Distinguished Professor, University of Texas M.D. Anderson Cancer Center, Houston, TX 77230 (jefmorris@mdanderson.org).

SUPPLEMENTARY MATERIALS

Supplement.pdf: This document is organized as follows. Section 1 describes details of MCMC update steps. Section 2 provides derivation of DF(t) for Model (23). Sections 3–4 include details of model selection results. In Section 5, we assess whether the smoothing parameter for the nonparametric age effect should be constant or vary around the scleral surface. In Section 6, we provide description of overall procedure to fit our model and obtain inferential results with a simulation dataset. Section 7 contains sensitivity analyses to various modeling assumptions, including results for other basis (wavelet-regularized PC), other model (model that also includes left vs. right eye effect), and other regularization hyperparameters (choice of values with no additional shrinkage provided by prior). Section 8 describes some supplementary files presenting additional results discussed in the paper, and Section 9 describes the simulated pseudo-data and demonstrates that it appears to capture the features of the real MPS scleral strain data reasonably well.

RawMPScurves.zip: It includes plots of raw MPS curves, results after tensor wavelet compression, and results after robust filtering to remove spiky artifacts.

movies.zip: Includes various .mp4 movie files illustrating various detailed results from the paper, including the following:

MPSvsAge-wave.mp4: Movie of MPS vs. age for each IOP based on the tensor wavelet basis function.

Combo plots.mp4: Movie of key summary results based on the tensor wavelet basis functions.

Intrafunctional correlations.mp4 Movie showing intrafunctional correlations induced by tensor wavelet basis functions.

Intra IOP corr.mp4 Movie showing interfunctional variance and serial correlation across IOP from same eye based on the tensor wavelet basis functions.

MPSvsAge-pc.mp4: Movie of MPS vs. age for each IOP based on the principal component basis functions.

AUCvsAge-pc.mp4: Movie of AUC vs. age based on the principal component basis functions.

Intrafunctional correlations-pc.mp4 Movie showing intrafunctional correlations induced by the principal component basis functions.

MPSvsAge-eye.mp4: Movie of MPS vs. age for each IOP based on tensor wavelet basis functions and model including left vs. right eye effect.

AUCvsAge-eye.mp4: Movie of AUC vs. age based on tensor wavelet basis functions and model including left vs. right eye effect.

Intrafunctional correlations-eye.mp4 Movie showing intrafunctional correlations induced by the tensor wavelet basis functions and model including left vs. right eye effect

MPSvsAge-nosmooth.mp4: Movie of MPS vs. age for each IOP based on tensor wavelet basis functions and model with no smoothing ($\pi. = 1$, $\tau. = 10^6$).

AUCvsAge-nosmooth.mp4: Movie of AUC vs. age based on tensor wavelet basis functions and model with no smoothing ($\pi. = 1$, $\tau. = 10^6$).

Intrafunctional correlations-nosmooth.mp4 Movie showing intrafunctional correlations induced by the tensor wavelet basis functions and model with no smoothing ($\pi. = 1$, $\tau. = 10^6$).

EYE toolbox.zip Contains all of the files necessary to run the methods presented in the paper, including the raw glaucoma data, pseudo data and full scripts to run the analyses and produce the plots contained in the paper. This includes wfmm_install.pdf that contains step-by-step instructions on how to install the R package *wfmm* and assocoiated executable, and Analysis_of_Pseudo_Data.pdf that contains detailed step-by-step instructions for running a complete analysis on pseudo data generated to mimic the real data in the application in Matlab, including the basis transform, model selection heuristic, MCMC in basis space, projection of posterior samples back to data space, MCMC convergence diagnostics, and producing all inferential summaries and plots contained in this paper that present results and illustrate properties of the model, with run time estimates for each step. We also include Producing Plots for Main Data Analysis in Paper.pdf that gives instructions for running scripts to reproduce the figures for the real data analysis for the main model used for the MPS scleral strain data analysis contained in the paper. This toolbox and data are available on our github (https://github.com/MorrisStatLab/SemiparametricFMM).

Department of Biostatistics, University of Texas M.D. Anderson Cancer Center, Houston, TX 77230

Michelle F. Miranda,
Department of Biostatistics, University of Texas M.D. Anderson Cancer Center, Houston, TX 77230

Philip Rausch,
Department of Psychology, Institut für Psychologie, Humboldt-Universität zu Berlin, Germany

Veerabhadran Baladandayuthapani,
Department of Biostatistics, University of Texas M.D. Anderson Cancer Center, Houston, TX 77230

Massimo Fazio,
Department of Ophthalmology, University of Alabama at Birmingham, Birmingham, AL 35294

J. Crawford Downs, and
Department of Ophthalmology, University of Alabama at Birmingham, Birmingham, AL 35294

Jeffrey S. Morris
Department of Biostatistics, University of Texas M.D. Anderson Cancer Center, Houston, TX 77230

Abstract

Glaucoma, a leading cause of blindness, is characterized by optic nerve damage related to intraocular pressure (IOP), but its full etiology is unknown. Researchers at UAB have devised a custom device to measure scleral strain continuously around the eye under fixed levels of IOP, which here is used to assess how strain varies around the posterior pole, with IOP, and across glaucoma risk factors such as age. The hypothesis is that scleral strain decreases with age, which could alter biomechanics of the optic nerve head and cause damage that could eventually lead to glaucoma. To evaluate this hypothesis, we adapted Bayesian Functional Mixed Models to model these complex data consisting of correlated functions on spherical scleral surface, with nonparametric age effects allowed to vary in magnitude and smoothness across the scleral surface, multi-level random effect functions to capture within-subject correlation, and functional growth curve terms to capture serial correlation across IOPs that can vary around the scleral surface. Our method yields fully Bayesian inference on the scleral surface or any aggregation or transformation thereof, and reveals interesting insights into the biomechanical etiology of glaucoma. The general modeling framework described is very flexible and applicable to many complex, high-dimensional functional data.

Keywords

Bayesian models; Functional data analysis; Functional mixed models; Functional regression; Glaucoma; Longitudinal Functional Data; Nonparametric effects; Smoothing Splines; Spherical data; Wavelets

1. INTRODUCTION

Glaucoma is one of the leading causes of blindness in the world. While its etiology is not fully understood, it is known to be caused by damage to the optic nerve head (ONH) that can be induced by intraocular pressure (IOP). Researchers hypothesized that the biomechanics of the peripapillary (PP) scleral region close to the ONH, shown to be an important determinant of ONH biomechanics, may play a major role in glaucoma pathogenesis and progression (Sigal et al., 2005; Burgoyne and Downs, 2008). Recently, novel custom instrumentation was developed that can induce a fixed level of IOP and precisely measure the mechanical strain in the posterior human sclera (Fazio, Bruno, Reynaud, Poggialini and Downs, 2012; Fazio, Grytz, Bruno, Girard, Gardiner, Girkin and Downs, 2012). A commercial laser speckle interferometer (ESPI) is used to measure the IOP-induced scleral displacement continuously around the posterior eye. The scleral displacement is processed to estimate the mechanical strain tensor on a grid of points on the scleral surface, and the largest eigenvalue of this strain tensor computed to yield the maximum principal strain (MPS), a scalar measurement for each location on the scleral surface that summarizes the magnitude of tensile strain at that location. Intuitively, scleral regions with higher MPS for a given level of IOP are more pliable, which in principle could either relieve IOP and reduce the potential for ONH damage or focus strain at the ONH and increase glaucoma damage. Age is a primary glaucoma risk factor and age-related stiffening occurs in other load-bearing soft tissues. Thus, the researchers hypothesized that age-related changes in scleral stiffness might contribute to the known age-related increase in glaucoma risk.

Utilizing this custom instrumentation, Fazio et al. (2014) conducted a study to test the hypothesis that scleral stiffness increases with age, with the expectation that future work will follow to elucidate the specific role of scleral stiffness in glaucoma. They obtained twenty pairs of eyes from normal human donors in the Lions Eye Bank of Oregon in Portland, Oregon, and the Alabama Eye Bank in Birmingham, Alabama. From each subject, the MPS was measured in the posterior globes of both left and right eyes on a partial spherical domain with 120 circumferential locations $\phi \in (0°, 360°)$ and 120 meridional locations $\theta \in (9°, 24°)$, where $\theta = 0°$ corresponds to the ONH. For each eye, MPS was measured under nine different IOP levels (7, 10, 15, 20, 25, 30, 35, 40, and 45 mmHg). Further scientific and technical details can be found in Section 5. Figure 1 plots a polar azimuthal projection of MPS functions for one subject under 45 mmHg IOP level. The center of each panel corresponds to the ONH of each eye. The hypothesis is that MPS will decrease with age, especially in scleral regions closer to the ONH.

The resulting data set is complex and high dimensional with many layers of structure. First, for each eye at each IOP level, the MPS measurements constitute a function on a two-dimensional partial spherical domain for which one needs to account for within-function (intrafunctional) correlations. Second, there are multiple sources of between-function (interfunctional) correlation. The measurements from the left and right eye from the same subject are expected to be correlated, and measurements from the same eye across the IOP levels should be serially correlated. The strength of these nested and serial correlations may potentially vary around the scleral surface. Based on preliminary looks at the data, it appears that the age effect on MPS may not be linear, and also appears to vary around the scleral

surface. This data set is also enormous, with over 4.5 million measurements, which poses practical problems for model building and fitting.

Many researchers, when faced with such complexities, would use one of several strategies to simplify the data so they can analyze them. Some researchers would get rid of the complexities of the functional data by computing summaries and modeling only those, while discarding the original functional data. In this application area, some researchers create summaries in the peripapillary (PP) and mid-peripheral (MP) scleral regions by integrating over the region closest to the ONH ($9° – 17°$) and further out ($17° – 24°$), respectively, or over several circumferential sections. This practice would miss any scientific insights not captured by these summaries. Alternatively, some researchers would only model a subset of the data (e.g. only a single IOP or single eye per subject) to avoid having to deal with potentially complex interfunctional correlation, or ignore this correlation entirely by modeling IOP within eyes or eyes within subject as independent. Some would ignore intrafunctional correlation by modeling individual pixels on the image independently. Many researchers would also only consider parametric or linear effects for covariates such as age without considering whether more complex nonparametric covariate effects might be necessary to capture the true relationship. All of these simplification strategies have potential statistical downsides.

In our opinion, it would be far preferable to model this complex function data set in its entirety using a statistical model that is flexible enough to capture all of its potentially complex intrafunctional and interfunctional structure. However, to the best of our knowledge, no statistical model has been presented in existing literature that simultaneously handles all of this structure. In this paper, we use the Bayesian functional mixed model (BayesFMM) framework, introduced in Morris and Carroll (2006) and further developed in subsequent papers, to model these data, which has sufficient flexibility to capture nonparametric covariate effects, serial and nested interfunctional correlation, functions on a fine 2d partial spherical domain, and is computationally efficient enough to scale up to this enormous size and produce Bayesian inference for functional parameters as well as any desired summaries. This requires a careful description of how to utilize the BayesFMM framework to model smooth nonparametric effects and serial interfunctional correlation, which has not been done in any existing literature to date.

While motivated by and applied to the glaucoma scleral strain data, the BayesFMM framework we present here is more generally applicable to many types of complex, high dimensional functional data of modern interest including wearable computing data, genome-wide data, proteomics data, geospatial time series data and neuroimaging data. Our hope is that besides revealing scientific insights into glaucoma, this paper can serve as a template for performing a state-of-the-art functional response regression analysis for other complex functional data sets.

The rest of the paper is organized as follows. Section 2 contains a brief review of some relevant literature in functional regression, including existing methods for modeling nested or serial interfunctional correlation and nonparametric smooth covariate functional effects. Section 3 contains methods, overviewing the BayesFMM framework and providing

methodological details for how to fit serially correlated functions and nonparametric smooth functional effects in this framework and discussing its properties. Section 4 describes a model selection heuristic that can be used to assess which of the various potential complex modeling structures are needed for the current data before having to run a full MCMC. Section 5 presents the details of our analysis of the glaucoma scleral strain data, including details on basis function and modeling components chosen, a summary of results, and various sensitivity analyses. Section 6 contains a discussion of general scientific conclusions from our analysis and an assessment of the strengths and weaknesses of the BayesFMM framework for this setting of complex functional regression models. Supplementary materials provide additional computational details and graphical results of the analysis, as well as code to fit the models contained in the paper.

2. LITERATURE REVIEW

There is a rich and rapidly expanding literature on methods to perform functional regression, which includes *functional predictor regression* (scalar-on-function), *functional response regression* (function-on-scalar), and *function-on-function regression*. For example, see Morris (2015) for an extensive review of work in this area, which has exploded in the past decade. In this paper, our goal is to regress the scleral strain MPS functions on predictors age and IOP, so we are interested in functional response regression. Here, we will summarize some key existing literature relevant to the structures present in the glaucoma scleral strain data set, including methods that account for nested and serial interfunctional correlation and smooth nonparametric covariate effects.

As summarized in Morris (2015), much of the work on functional response regression assumes independently sampled functions, but a number of published models can handle interfunctional correlation. Many are focused on nested or crossed sampling designs that induce compound symmetry covariance structures among functions sampled within the same cluster, including Brumback and Rice (1998), Morris et al. (2003), Berhane and Molitor (2008), Aston et al. (2010), and Goldsmith and Kitago (2016). There are relatively few that model serially correlated functions, for which functions are observed at multiple levels of some continuous variable for the same subject. Most commonly, the functions occur along a grid of time points, which could be called *longitudinally correlated functional data*, but can also occur over other continuous variables such as IOP for our glaucoma data. There are a number of papers that focus on functional predictor regression (Goldsmith et al., 2012; Gertheiss, Goldsmith, Crainiceanu and Greven, 2013; Kundu et al., 2016; Islam et al., 2016) or estimate multi-level principal components (Greven et al., 2010; Zipunnikov et al., 2011; Chen and Müller, 2012; Li and Guan, 2014; Zipunnikov et al., 2014; Park and Staicu, 2015; Shou et al., 2015; Hasenstab et al., 2017) for serially correlated functions, but these papers do not deal directly with functional response regression, i.e. do not regress the functions on covariates while accounting for this serial correlation in the error structure. Any methods built for a general functional mixed modeling framework containing multiple levels of general random effect functions, including the BayesFMM framework first introduced by Morris and Carroll (2006) and the functional additive mixed model (FAMM) framework first introduced by Scheipl et al. (2015), can be used for functional response regression while accounting for serial interfunctional correlation if an acceptable parametric form for the

serial variable can be found. However, no specific models or examples in existing papers have presented such a case. In this paper, we will demonstrate in detail how to account for serial correlation within the BayesFMM framework.

Most functional response regression work, while modeling coefficients as nonparametric in the functional domain *t*, has focused on models linear in covariates *x*, e.g. with terms such as $x\beta(t)$. A number of methods allow functional coefficients that are smooth and nonparametric in covariate *x* as well, e.g. $f(x, t)$. The last chapter of Wood (2006) describes additive mixed models (AMMs) that extend Laird and Ware (1982) to include nonparametric fixed effects and parametric random effects. After specifying a parametric form for *t* in fixed and random effects, this framework can fit terms $f(x, t)$ that are nonparametric in *x* but parametric in *t* using mgcv in R. Scheipl et al. (2015) describe a generalization of AMMs to the functional regression settting, yielding *functional additive mixed models (FAMM)* that can fit functional response, functional predictor, or function-on-function regression models, and allowing terms $f(x, t)$ that are nonparametric in both *x* and *t*, using mgcv to fit the underlying models. Initially designed for splines, this work was extended to allow the use of functional principal components (fPC) for the random effect functions to handle sparse, irregularly sampled outcomes in Cederbaum et al. (2015), to model generalized outcomes in Scheipl et al. (2016), and to utilize a new boosting-based fitting procedure for some models that leads to other computational and modeling benefits in Brockhaus et al. (2015). This series of work is summarized in a review article Greven and Scheipl (2017*a*). These works utilize the fact, going way back to Wahba (1978), that penalized splines can be represented using a linear mixed model framework. Using similar representations, the BayesFMM framework first introduced in Morris and Carroll (2006) can also be used to fit nonparametric terms $f(x, t)$ with appropriate specification of the design matrices *X* and *Z* in the model, but this has not been done in any existing paper. In this paper, we will describe in detail how to accommodate smooth nonparametric terms $f(x, t)$ in the BayesFMM framework.

To our knowledge, the only methods in existing literature with the flexibility to both fit smooth nonparametric functional covariate terms $f(x, t)$ and simultaneously account for nested and serial correlation are the FAMM framework (Scheipl et al., 2015; Greven and Scheipl, 2017*a*) and the BayesFMM framework (Morris and Carroll, 2006). Both of these frameworks are extremely flexible, based on functional extensions of linear mixed models, but have important differences, which are discussed in detail in Morris (2017) and Greven and Scheipl (2017*b*). The FAMM framework has several limitations that prevent its application to our glaucoma scleral strain data, including the requirement of functions to be on 1d Euclidean domains, use of spline bases for fixed effects with L2 penalties, and use of a computational approach that may not scale up well to enormous data sets like this. Thus, to fit these data, we adapt the BayesFMM framework to include smooth nonparametric age effects and serial correlation across IOP, and demonstrate how we can use it to reveal insights into the biomechanical etiology of glaucoma.

3. METHODS

We first overview the BayesFMM framework in Section 3.1, and then for ease of exposition we build up the necessary components of our model separately before presenting our final

general model. In Section 3.2, we demonstrate how to capture serial interfunctional correlation through functional growth curve models, and then demonstrate how to accommodate smooth nonparametric covariate functional effects in Section 3.3. Besides presenting the models, we also describe their properties and contrast with existing alternative methods. In Section 3.4 we present a general alternative form of the BayesFMM that includes smooth nonparametric terms that we use to fit our glaucoma scleral strain data.

3.1 Bayesian Functional Mixed Models (BayesFMM)

Suppose we have a sample of functions $Y_i(\mathbf{t})$, $i = 1, \ldots, N$ observed on a common fine grid of size T on a domain $\mathcal{T}$, which is potentially multi-dimensional and/or non-Euclidean (Morris et al., 2011). The FMM introduced by Morris and Carroll (2006) is a functional response regression model given by

$$Y_i(\mathbf{t}) = \sum_{a=1}^{A} X_{ia} B_a(\mathbf{t}) + \sum_{h=1}^{H} \sum_{m=1}^{M_h} Z_{ihm} U_{hm}(\mathbf{t}) + E_i(\mathbf{t}) ., \quad (1)$$

where $B_a(\mathbf{t})$ are fixed effect functions that model the effect of covariate X_{ia} on the response Y at position $\mathbf{t}$ for each covariate $a = 1,\ldots, A$, and the $U_{hm}(\mathbf{t})$ are random effect functions at level $h = 1, \ldots, H$ corresponding to design matrix Z_{ihm}, with $m = 1, \ldots, M_h$ being the number of random effects at the respective levels. As in linear mixed models for scalar data, the fixed and random effect predictors can be discrete or continuous, and can involve individual covariates or interactions of multiple covariates. Although not explicitly portrayed in (1), this modeling framework also accommodates functional predictors to perform function-on-function regression (Meyer et al., 2015).

Distributional and Covariance Assumptions—Here, for simplicity, we first describe the Gaussian FMM with conditionally independent random effect and residual error functions, and then mention some other alternatives available in this framework. In this case, the random effect functions $U_{hm}(\mathbf{t})$ are iid mean zero Gaussian Processes with intrafunctional covariance cov$\{ U_{hm}(\mathbf{t}_1), U_{hm}(\mathbf{t}_2)\} = Q_h(\mathbf{t}_1, \mathbf{t}_2)$ and the residual error functions $E_i(\mathbf{t})$ are iid mean zero Gaussian Processes with intrafunctional covariance cov$\{E_i(\mathbf{t}_1),E_i(\mathbf{t}_2)\} = S(\mathbf{t}_1, \mathbf{t}_2)$. Other extensions of this framework allow the option of conditional autoregressive (CAR) (Zhang et al., 2014) or Matern spatial covariance or AR(p) temporal interfunctional correlation structures in the residual errors (Zhu et al., 2014). Although focusing on Gaussian regression here, a robust version of this framework assuming heavier tailed distributions on the random effects or residuals is available (Zhu et al., 2011) if robustness to outliers is desired, and can also be utilized with any other features or modeling components in the BayesFMM framework.

Basis Transform Modeling Approach—A *basis transform modeling approach* is used to fit model (1). This first involves representing the observed functions with a basis expansion with a set of basis functions $\psi_k(\mathbf{t})$, $k = 1, \ldots, K$:

$$Y_i(\mathbf{t}) = \sum_{k=1}^{K} Y_{ik}^* \psi_k(\mathbf{t}) \quad (2)$$

While initially developed for wavelets (Morris and Carroll, 2006), as first discussed in Morris et al. (2011), the modeling approach can be used with any basis. It is meant to be used with *lossless* transforms with $Y_i(\mathbf{t}) \equiv \sum_k Y_{ik}^* \psi_k(\mathbf{t})$ for all observed $\mathbf{t}$, so that the basis coefficients $\{ Y_{ik}^*; k = 1, \ldots, K\}$ contain all information within the observed functional data $\{ Y_i(\mathbf{t}); \mathbf{t} = \mathbf{t}_1, \ldots, \mathbf{t}_T \}$, or at least *near-lossless* with

$$\left\| Y_i(\mathbf{t}) - \sum_{k=1}^{K} Y_{ik}^* \psi_k(\mathbf{t}) \right\| < \varepsilon \quad \forall i = 1, \ldots, N \quad (3)$$

for some small value ε and measure $\|\bullet\|$. This assures that the chosen basis is sufficiently rich such that for practical purposes it can recapitulate the observed functional data, and visual inspection of the raw functions and basis transformation should reveal virtually no difference. Any basis functions can be used, including commonly used choices splines, wavelets, Fourier bases, PCs or creatively constructed custom bases, and can be defined on multi-dimensional or non-Euclidean domains $\mathcal{T}$.

Rather than including the bases in a design matrix and using scalar regression methods to fit the model, our approach is to transform the observed functions into the basis space to obtain the basis coefficients $\mathbf{Y}^*$, a $N \times K$ matrix for which element (i, k) contains the basis coefficient k for observed function i, fit a basis-space version of the FMM to these coefficients, and then transform results back to the data space model (1) for estimation and inference. With basis representation written in matrix form $\mathbf{Y} = \mathbf{Y}^* \mathbf{\Psi}$ with $\mathbf{\Psi}$ a $K \times T$ matrix of basis functions evaluated on the observational grid with $\mathbf{\Psi}_{kj} = \psi_k(\mathbf{t}_j)$, the coefficients can be computed by $\mathbf{Y}^* = \mathbf{Y} \mathbf{\Psi}'$ with $\mathbf{\Psi}' = \mathbf{\Psi}'(\mathbf{\Psi} \mathbf{\Psi}')^{-1}$ as long as rank$(\mathbf{\Psi}) = K$, or for certain basis functions including wavelets and Fourier bases their special structure enables fast algorithms for computing these coefficients.

Basis Space Model—Thus, rather than fitting model (1) directly, the basis-space version of the model is fit for each basis coefficient $k = 1, \ldots, K$:

$$Y_{ik}^* = \sum_{a=1}^{A} X_{ia} B_{ak}^* + \sum_{h=1}^{H} \sum_{m=1}^{M_h} Z_{ihm} U_{hmk}^* + E_{ik}^*, \quad (4)$$

where B_{ak}^*, U_{hmk}^*, and E_{ik}^* are basis coefficients for the functional fixed effects $B_a(\mathbf{t}) = \sum_k B_{ak}^* \psi_k(\mathbf{t})$, functional random effects $U_{hm}(\mathbf{t}) = \sum_k U_{hmk}^* \psi_k(\mathbf{t})$, and functional

residuals $E_i(\mathbf{t}) = \sum_k E_{ik}^* \psi_k(\mathbf{t})$, respectively. While in principle correlation across basis coefficients can be accommodated, for complex, high-dimensional functions, it may be beneficial to model these basis coefficients independently. For the Gaussian FMM with conditionally independent random effect and residual error functions, it is assumed $U_{hmk}^* \sim N(0, q_{hk})$ and $E_{ik}^* \sim N(0, s_k)$ with q_{hk} and s_k scalar variance components. Although modeling independently in the basis space, this structure induces intrafunctional correlation according to the chosen basis functions. For example, for the residual error functions, the induced $T \times T$ intrafunctional correlation $\mathbf{S}$ with $\mathbf{S}_{ij} = \mathbf{S}(\mathbf{t}_i, \mathbf{t}_j)$ is given by:

$$\mathbf{S} = \mathbf{\Psi}' \mathbf{S}^* \mathbf{\Psi}, \quad (5)$$

where $\mathbf{S}^* = \text{diag}\{s_k; k = 1, \ldots, K\}$, and $\mathbf{Q}_h$ defined likewise. For suitably chosen basis functions that effectively capture the characteristic structure of the observed functions $Y_i(\mathbf{t})$, this can allow a flexible class of covariance structures indexed by K covariance parameters. One can assess the suitability of these assumptions by taking the basis-space variance components, computing (5), and plotting this covariance matrix to see if it appears to capture the salient structure. Figure 3 panels (e) and (f) plot the intrafunctional correlation structure induced by the chosen tensor wavelet basis for the scleral strain MPS data set for a particular scleral location for the eye-to-eye random intercepts and residual error functions, and the file `intrafunctional_correlation.mp4` in the supplement shows a more extensive summary of all random levels across scleral locations.

Shrinkage Priors for Regularization of Fixed Effects—If fitting this model using a frequentist approach, L1 or L2 penalties could be imposed on the basis space fixed effects to induce regularization/smoothing of the fixed effect functions $B_a(\mathbf{t})$, as described in Morris (2015). However, this framework was designed to use a Bayesian modeling approach, in which case regularization of the fixed effect functions $B_a(\mathbf{t})$ is accomplished through specification of shrinkage priors for the corresponding basis coefficients:

$$B_{ak}^* \sim g(\boldsymbol{\gamma}_{aj}) \quad (6)$$

for some mean zero distribution $g(\bullet)$ with corresponding regularization parameters $\boldsymbol{\gamma}_{aj}$ indexed by $j = 1, \ldots, J$ that define a partitioning of the basis coefficients $k = 1, \ldots, K$ into *regularization sets*, which are subsets of basis coefficients sharing the same regularization parameters. Morris and Carroll (2006) used the spike-slab prior (George and McCulloch, 1993) for $g(\bullet)$ and we also use that here, but other alternatives include Gaussian, Laplace (Park and Casella, 2008), Horseshoe (Carvahlo et al., 2010), Normal-Gamma (Griffin and Brown, 2010), and Dirichlet-Laplace (Bhattachary et al., 2015). The regularization parameters $\boldsymbol{\gamma}_{aj}$ can be given hyperpriors or be estimated by empirical Bayes, which is described in detail for the spike-slab in Morris and Carroll (2006).

Model Fitting Approach—A fully Bayesian modeling approach is used to fit a Markov Chain Monte Carlo (MCMC) to the basis space model (4) for each k, and then transforming back to the data space, e.g. using $B_a(\mathbf{t}) = \sum_k B_{ak}^* \psi_k(\mathbf{t})$ to yield posterior samples for each parameter in the data-space FMM (1). This requires specification of priors for the variance components; a vague empirical Bayes approach is used that centers the prior on REML starting values of these parameters, minimally informative to be equivalent to two data points of information, as detailed in Section 5. BayesFMM fits a marginalized version of model (4) with all random effects $\mathbf{U}_*$ integrated out, which integrates over the interfunctional correlation induced by the random effect levels when updating the fixed effects, and speeds convergence of the chain and calculations since the random effect functions themselves need not be sampled. As detailed in the supplement, the fixed effect updates involve conjugate (spike-slab) Gibbs steps and variance components are updated via Metropolis-Hastings with proposal variances automatically computed based on the corresponding Fisher information. If desired, posterior samples for random effects can be obtained by sampling from conjugate Gaussians.

Bayesian Inference—Given these posterior samples, one can compute any desired posterior probabilities and pointwise or joint credible bands that can be used for Bayesian inference. Joint bands can be constructed using the approach described in Ruppert et al. (2003). These inferential summaries can be constructed in the data or basis space for any functional of model parameters, including contrasts (e.g. $B_1(\mathbf{t}) - B_2(\mathbf{t})$), nonlinear transformations (e.g. $\exp\{B_a(\mathbf{t})\}$), derivatives (e.g. $f(x_a, \mathbf{t})/\partial x_a$), or integrals (e.g. $\int_{\mathbf{t} \in \mathcal{T}_0} B_a(\mathbf{t}) d\mathbf{t}$ for some $\mathcal{T}_0 \subset \mathcal{T}$) aggregating information across regions of $\mathbf{t}$. We make use of these to produce inference on numerous scientifically interesting summaries for our scleral strain MPS data in Section 5.

Example FMM for Scleral Strain MPS Data—To illustrate this framework, suppose that we model MPS functions from both left and right eyes for each subject, but only for a specific IOP level, and suppose we are willing to assume a linear *age* effect. Let $Y_{i1}(\mathbf{t})$ and $Y_{i2}(\mathbf{t})$ be MPS functions for the left and right eyes respectively from the i-th subject, with $\mathbf{t} \in \mathcal{T}$ indexing the scleral domain, with $\mathbf{t} = (\theta, \phi)$ being spherical coordinates on the scleral surface. We could represent these data with the following FMM:

$$Y_{ij}(\mathbf{t}) = B_0(\mathbf{t}) + X_{\text{age}, i} B_{\text{age}}(\mathbf{t}) + U_i(\mathbf{t}) + E_{ij}(\mathbf{t}), \quad (7)$$

with $U_i(\mathbf{t}) \sim N\{\mathbf{0}, Q(\bullet)\}$ and $E_{ij}(\mathbf{t}) \sim N\{\mathbf{0}, S(\bullet)\}$. The $U_i(\mathbf{t})$ induce a generalized compound symmetry covariance structure between the functions from the left and right eyes from the same subject, with $\text{cov}\{Y_{ij}(\mathbf{t}), Y_{ij'}(\mathbf{t})\} = Q(\mathbf{t}, \mathbf{t}) + S(\mathbf{t}, \mathbf{t})/(j = j')$. To simultaneously model data for all IOP, we would need to add a random effects level to capture the serial correlation across MPS functions for different IOP for the same eye. We next describe how this can be done.

3.2 Accommodating Serial Interfunctional Correlation via Functional Growth Curves

Recall that in our scleral strain MPS data set, for each eye we have MPS functions from each of a series of IOP levels ranging from 7 mmHg to 45 mmHg, which induces a serial correlation across scleral strain functions for the same eye according to IOP. It is important to account for this serial interfunctional correlation in some appropriate fashion in order to obtain efficient estimates and accurate inference for any fixed effect functions in the model. Here we demonstrate how functional random effects in a functional mixed modeling framework can be used to capture this serial correlation using a type of functional growth curve model, and discuss the properties of this strategy in the context of the BayesFMM framework using a basis transform modeling approach.

Basic Growth Curve Model for Scleral Strain MPS Data—For illustration, here we only consider the serial effect of IOP and omit any age effect and consider only the left eye for each subject. Suppose $Y_{ip}(\mathbf{t})$ is the MPS function for the ith subject, $i = 1, \ldots, N$, after exposure to an IOP level of p. Note that $Y_{ip}(\mathbf{t})$ is a function of $\mathbf{t}$ that varies across serial variable p. To capture the serial effect of p, we consider the following *functional growth curve model*:

$$Y_{ip}(\mathbf{t}) = m(p, \mathbf{t}) + u_i(p, \mathbf{t}) + E_{ip}(\mathbf{t}), \quad (8)$$

where $m(p, \mathbf{t})$ is the mean MPS for IOP level p and scleral location $\mathbf{t}$, $u_i(p, \mathbf{t})$ is a mean zero random effect for subject i that represents a subject-specific growth curve in p that is allowed to vary across scleral location $\mathbf{t}$, and $E_{ip}(\mathbf{t})$ are residual error functions assumed to be independent and identically distributed mean zero Gaussians with covariance $S(\bullet)$. If we can find a suitable parametric form for the serial effect of p with basis functions $G_d(p)$, $d = 0, 1, \ldots, D$, we assume $m(p, \mathbf{t}) = \sum_{d=0}^{D} B_d(\mathbf{t})G_d(p)$ and $u_i(p, \mathbf{t}) = \sum_{d=0}^{D} U_{i,d}(\mathbf{t})G_d(p)$ with $\text{cov}\{U_{i,d}(\mathbf{t}_1), U_{i,d}(\mathbf{t}_2)\} = Q_d(\mathbf{t}_1, \mathbf{t}_2)$. In practice, we recommend using basis functions that are orthogonal across d, i.e. $\int G_d(p)G_{d'}(p)dp = 0$ for $d \neq d'$ to obviate the need to have cross-covariance terms between $U_{i,d}(\mathbf{t})$ and $U_{i,d'}(\mathbf{t})$ to avoid additional computational complexity in the model.

The introduction of this eye-level random growth curve induces serial covariance across functions for the same eye at different levels of IOP, with $\text{cov}\{Y_{ip}(\mathbf{t}), Y_{ip'}(\mathbf{t})\} = \sum_{d=0}^{D} G_d(p)G_d(p')Q_d(\mathbf{t}, \mathbf{t})$. Indexed by $\mathbf{t}$, the strength and shape of this serial covariance can vary across the scleral surface $\mathbf{t}$. Figure 3 panel (d) contains the induced serial correlation across IOP for the marked scleral location, and the file IOP_corr.mp4 in the supplementary materials is a movie file that demonstrates how this correlation varies over the scleral surface. Note also that $\text{cov}\{Y_{ip}(\mathbf{t}_1), Y_{ip'}(\mathbf{t}_2)\} = \sum_{d=0}^{D} G_d(p)G_d(p')Q_d(\mathbf{t}_1, \mathbf{t}_2)$, meaning that this structure enables "borrowing of strength" from nearby $\mathbf{t}$ in determining the strength and shape of the serial covariance according to the intrafunctional covariance indicated by the off-diagonal elements of $Q_d(\bullet)$. If model (8) is marginalized with respect to the random effect functions, the resulting error terms can be seen to contain the induced serial correlation structure,

which is subsequently accounted for in any estimation or inference of the fixed effect functions.

Incorporation into BayesFMM framework—It can be seen that model (8) can be written as a functional mixed model with $D+1$ fixed effect predictors and $D+1$ random effect levels, each with N subject-specific random effect functions. Thus, any FMM framework allowing multiple levels of random effect functions could be used to fit this model. In the BayesFMM framework, after transforming the observed functions $Y_{ip}(\mathbf{t})$ into the basis space through $Y_{ip}(\mathbf{t}) = \sum_{k=1}^{K} Y_{ipk}^* \psi_k(\mathbf{t})$ as described in Section 3.1, the model for coefficient k would be given by:

$$Y_{ipk}^* = \sum_{d=0}^{D} G_d(p) B_{dk}^* + \sum_{d=0}^{D} G_d(p) U_{idk}^* + E_{ipk}, \quad (9)$$

with $U_{idk}^* \sim N(0, q_{dk})$ and $E_{ipk}^* \sim N(0, s_k)$. This would induce serial covariance across p in the model for each basis coefficient with cov $(Y_{ipk}^*, Y_{ip'k}^*) = \sum_{d=0}^{D} G_d(p) G_d(p') q_{dk}$, and the marginalized model that is fit with the U_{idk}^* integrated out will contain this serial covariance in the error structure and explicitly account for it when updating the fixed effect coefficients. This basis space model induces the serial covariance structure described in the data space model (8), with $Q_d(\mathbf{t}_1, \mathbf{t}_2) = \sum_{k=1}^{K} \psi_k(\mathbf{t}_1) \psi_k(\mathbf{t}_2) q_{dk}$. The heteroscedasticity of the variance components across basis functions k allows the strength and shape of the serial covariance to vary across the scleral surface $\mathbf{t}$, and effectively borrows strength from nearby $\mathbf{t}$ through the chosen basis functions as determined by the induced off-diagonal elements of $Q_d(\mathbf{t}_1, \mathbf{t}_2)$. Figure 3 panel (e) plots the intrafunctional correlation structure corresponding the random intercept $Q_0(\mathbf{t}_1, \mathbf{t}_2)$ induced by the tensor wavelet basis chosen for the scleral strain MPS data set at the marked scleral location, and file `intrafunctional_correlation.mp4` in the supplement presents the induced intrafunctional correlation for each of the $Q_d(\mathbf{t}_1, \mathbf{t}_2)$, ($d = 0, 1, 2$) across all scleral locations.

This strategy can be used with any parametric model indicated by the $G_d(p)$, $d = 0, \ldots, D$, preferably orthogonalized. Section 5 demonstrates that the IOP effect in the scleral strain MPS data is hyperbolic, so we devise an orthogonalized hyperbolic model for these data. Also, note that while the serial variable IOP is sampled on a common grid across subjects for our data, this strategy can allow each the grid points for the serial variable to vary across subjects.

3.3 Smooth Nonparametric Covariate Functional Effects

One of the primary scientific goals in the scleral strain MPS data is to study the effect of age on MPS and assess how it varies around the scleral surface. Preliminary investigations of the data suggest that the age effect might not follow a simple parametric form, and a nonparametric representation might be appropriate. While the fixed effect functions in the BayesFMM framework are linear in the covariates, using the mixed model representation of

penalized splines shown by Wahba (1978), it is possible to fit a semiparametric functional mixed model with a smooth nonparametric age effect using this framework, as we will demonstrate in this section.

Smooth Nonparametric Age Effect for Scleral Strain MPS data—For ease of exposition, in this section we just consider a single smooth nonparametric term with no other covariates or random effects. Thus, suppose for each subject $i = 1, \ldots, N$ we only model a single scleral strain function $Y_i(\mathbf{t})$, say for left eye and IOP=45mmHg, with the model:

$$Y_i(\mathbf{t}) = B_0(\mathbf{t}) + f(X_{\text{age}_i}, \mathbf{t}) + E_i(\mathbf{t}), \quad (10)$$

where $B_0(\mathbf{t})$ is a functional intercept and $f(X_{\text{age}_i}, \mathbf{t})$ represents a nonparametric effect of age on MPS at scleral location $\mathbf{t}$, with $\int f(x, \mathbf{t})dx = 0 \,\forall \mathbf{t}$ and penalizing $\int \{f''(x, \mathbf{t})\}^2 dx$ to induce smoothness across x for each $\mathbf{t}$. As we will demonstrate, it is possible to represent this smooth nonparametric term as a sum of a linear fixed effect function for age and spline random effect functions,

$$f(X_{\text{age}, i}, \mathbf{t}) = X_{\text{age}_i} B_1(\mathbf{t}) + \sum_{m=1}^{M+2} Z_{\mathscr{B}, m}(X_{\text{age}_i}) U_{\mathscr{S}m}(\mathbf{t}) \quad (11)$$

for some suitably constructed random effect design matrix $\{Z_{\mathscr{B},m}(x), m = 1, \ldots, M+2\}$ based on Demmler-Reinsch basis functions (Demmler and Reinsch, 1975), with spline random effects $U_{\mathscr{S}m}(\mathbf{t})$ following a mean zero Gaussian with $\text{cov}\{U_{\mathscr{S}m}(\mathbf{t}_1), U_{\mathscr{S}m}(\mathbf{t}_2)\} = Q_{\mathscr{S}}(\mathbf{t}_1, \mathbf{t}_2)$. These model components can be incorporated within a FMM framework like BayesFMM, as we now describe.

Incorporation into BayesFMM framework—To fit model (10) using the BayesFMM framework, we fit separate penalized splines for each basis coefficient k, which induces correlated penalized spline fits for each scleral location $\mathbf{t}$. Specifically, after transforming the observed functions $Y_i(\mathbf{t})$ into the basis space according to the basis representation $Y_i(\mathbf{t}) = \sum_{k=1}^{K} Y_{ik}^* \psi_k(\mathbf{t})$ as described in Section 3.1, we specify the following model for each basis coefficient $k = 1, \ldots, K$:

$$Y_{ik}^* = B_{0k}^* + f_k^*(X_{\text{age}_i}) + E_{ik}^*, \quad (12)$$

with $E_{ik}^* \sim N(0, s_k)$ and $f_k^*(x)$ a smooth nonparametric function of x for basis function k. We pull out the intercept B_{0k}^* and constrain $\int f_k^*(x)dx = 0$ to ensure identifiability in additive models that contain multiple smooth nonparametric terms. We represent $B_{0k}^* + f_k^*(x)$ using B-spline basis functions,

$$B_{0k}^* + f_k^*(x) = \sum_{m=1}^{M+4} \mathcal{B}_m(x)\nu_{mk}^*, \quad (13)$$

where ν_{mk}^* are B-spline coefficients and $\mathcal{B}_m(x), m = 1, \ldots, M+4$ are the cubic B-spline basis functions defined by the knots $\eta_1, \ldots, \eta_{M+8}$ such that

$$a = \eta_1 = \eta_2 = \eta_3 = \eta_4 < \eta_5 < \cdots < \eta_{M+4} = \eta_{M+5} = \eta_{M+6} = \eta_{M+7} = \eta_{M+8} = b,$$

and a and b are two boundary knots (Hastie et al., 2009). We can write model (12) in matrix form,

$$\mathbf{y}_k^* = \mathcal{B}\nu_k^* + \mathbf{e}_k^*, \quad (14)$$

where $\mathbf{y}_k^* = (Y_{1k}^*, \ldots, Y_{Nk}^*)'$, $\mathcal{B}$ is the $N \times (M+4)$ B-spline design matrix with the (i,m)-th entry being $\mathcal{B}_m(X_{\text{age},i})$, $\nu_k^* = (\nu_{1k}^*, \ldots, \nu_{(M+4)k}^*)'$, and $\mathbf{e}_k^* = (E_{1k}^*, \ldots, E_{Nk}^*)' \sim N(\mathbf{0}, s_k I_N)$. Following Wand and Ormerod (2008), we assume the following prior distribution on the B-spline coefficients:

$$\nu_k^* \sim MVN(\mathbf{0}, q_{\mathcal{S}k}\mathbf{\Omega}) \quad (15)$$

where $\mathbf{\Omega}$ is a $(M+4) \times (M+4)$ matrix with $\mathbf{\Omega}_{mm'} = \int_a^b \mathcal{B}_m''(x)\mathcal{B}_{m'}''(x)dx$. The resulting posterior mean of the spline random effects is $\hat{\nu}_k^* = (\mathcal{B}'\mathcal{B} + \lambda_k^*\Omega)^{-1}\mathcal{B}'\mathbf{y}_k^*$, where $\lambda_k^* = s_k/q_{\mathcal{S}k}$. It can be shown that $\mathcal{B}\hat{\nu}_k^*$ corresponds to the O'Sullivan penalized spline estimator of $B_{0k}^* + f_k^*(x)$ with penalty term $\lambda_k^* \int_a^b \{f_k^{*\prime\prime}(x)\}^2 dx$ (Wand and Ormerod, 2008), and if the knots are placed at each observed $X_{\text{age},i}$, then this corresponds to the cubic smoothing spline estimator.

The spectral decomposition of Ω allows us to reformulate this prior specification as a mixed model with independent random effects as follows. It is known that $\text{rank}(\Omega) = M+2$. Therefore, the spectral decomposition of Ω has the form of $\Omega = PDP'$, where $D = \text{diag}(0, 0, d_1, \ldots, d_{M+2})$ and $P'P = I_{M+4}$. Let $P = (X_\Omega, Z_\Omega)$, where X_Ω is a $(M+4) \times 2$ sub-matrix of P corresponding to the first two columns of P and Z_Ω is a $(M+4) \times (M+2)$ sub-matrix of P corresponding to the other columns. Let β_k^* be a two-dimensional vector, and $\mathbf{u}_{\mathcal{S}k}^* = (U_{\mathcal{S}1k}^*, \ldots, U_{\mathcal{S}(M+2)k}^*)'$ be an $(M+2)$-dimensional random vector. It can be shown that

$\nu_k^* = X_\Omega \boldsymbol{\beta}_k^* + Z_\Omega \text{diag}(d_1^{-1/2}, ..., d_{M+2}^{-1/2})\mathbf{u}_{\mathscr{S}k}^*$ with $\boldsymbol{\beta}_k^*$ a fixed effect and $\mathbf{u}_{\mathscr{S}k}^* \sim MVN(\mathbf{0}, q_{\mathscr{S}k}I_{M+2})$

Therefore, we have the following mixed model representation of (12),

$$\begin{aligned}
\mathbf{y}_k^* &= \mathscr{B}\nu_k^* + \mathbf{e}_k^* \qquad\qquad\qquad (16)\\
&= \mathscr{B}\{X_\Omega\boldsymbol{\beta}_k^* + Z_\Omega\text{diag}(d_1^{-1/2}, ..., d_{M+2}^{-1/2})\mathbf{u}_{\mathscr{S}k}^*\} + \mathbf{e}_k^* \\
&= X_{\mathscr{B}}\boldsymbol{\beta}_k^* + Z_{\mathscr{B}}\mathbf{u}_{\mathscr{S}k}^* + \mathbf{e}_k^*,
\end{aligned}$$

where $X_{\mathscr{B}} = \mathscr{B}X_\Omega$ and $Z_{\mathscr{B}} = \mathscr{B}Z_\Omega\text{diag}(d_1^{-1/2}, ..., d_{M+2}^{-1/2})$. The $Z_{\mathscr{B}}$ are called the Demmler-Reinsch spline bases (Demmler and Reinsch, 1975). It can be shown that $X_{\mathscr{B}}$ is a basis for the space of the straight line, so model can equivalently be rewritten as

$$\mathbf{y}_k^* = \mathbf{1}_N B_{0k}^* + \mathbf{x}_{\text{age}}B_{1k}^* + Z_{\mathbf{B}}\mathbf{u}_{\mathscr{S}k}^* + \mathbf{e}_k^*, \quad (17)$$

where $\mathbf{1}_N$ is an N-dimensional vector consisting of 1's and $\mathbf{x}_{\text{age}} = (X_{\text{age},1}, ..., X_{\text{age},N})'$. We see that this is the form of a linear mixed model with a level of spline random effects with design matrix $Z_{\mathscr{B}}$ and random effects $\mathbf{u}_{\mathscr{S}k}^* \sim MVN(\mathbf{0}, q_{\mathscr{S}k}I_{M+2})$. With the intercept term pulled out as implied by the $\int f_k^*(x)dx = 0$ assumption, the term $f_k^*(X_{\text{age}_i})$ in (12) is given by $X_{\text{age}_i}B_{1k}^* + Z_{\mathscr{B}}(X_{\text{age}_i})\mathbf{u}_{\mathscr{S}k}^*$ and thus in the BayesFMM framework we can incorporate a nonparametric term $f_k^*(x)$ by simply including a linear fixed effect xB_{1k}^* plus a level of random effects with the corresponding Demmler-Reinsch design matrix $\sum_{m=1}^{M+2} Z_{\mathscr{B}m}(x)U_{\mathscr{S}mk}^*$. These penalized splines for each basis k, when projected back to the function space, induce a smooth nonparametric functional effect $f(X_{\text{age}_i}, \mathbf{t})$ given by (11), with $Q_{\mathscr{S}}(\mathbf{t}_1, \mathbf{t}_2) = \Sigma_k \psi_k(\mathbf{t}_1)\psi_k(\mathbf{t}_2)q_{\mathscr{S}k}$. Based on these derivations, we can add any additional smooth nonparametric term $f(z, \mathbf{t})$ to the FMM framework by simply adding a linear fixed effect function $zB_2(\mathbf{t})$ and an additional level of spline random effects $\sum_{m=1}^{M_z+2} Z_{\mathscr{B}(z)}U_{\mathscr{S}_z m}(\mathbf{t})$ with $U_{\mathscr{S}_z m}(\mathbf{t})$ a mean zero Gaussian with covariance $\text{cov}\{U_{\mathscr{S}_z m}(\mathbf{t}_1), U_{\mathscr{S}_z m}(\mathbf{t}_2)\} = Q_{\mathscr{S}_z}(\mathbf{t}_1, \mathbf{t}_2)$.

A similar procedure could be followed to utilize other spline modeling approaches within this framework, e.g. P-splines (Eilers and Marx, 1986) with differencing penalties or truncated polynomial splines (Ruppert et al., 2003), but we prefer the O'Sullivan splines (Wand and Ormerod, 2008) given their natural second derivative penalty and formal connection to smoothing splines.

Intrafunctional Correlation of $f(x, t)$ Across t—This framework allows the nonparametric smooth effect $f(x, \mathbf{t})$ of x to vary over $\mathbf{t}$, but is not the same as modeling

independent splines for each $\mathbf{t}$. Because the splines are fit in the basis space, the nonparametric fits are correlated intrafunctionally by:

$$\text{cov}\,\{f(x,\mathbf{t}_1), f(x,\mathbf{t}_2) \mid B_1(\mathbf{t}_1), B_1(\mathbf{t}_2), q_{\mathscr{S}m}\} = \sum_{k=1}^{K} \sum_{m=1}^{M+2} \psi_k(\mathbf{t}_1)\psi_k(\mathbf{t}_2)\{Z_{\mathscr{B}_m(x)}\}^2 q_{\mathscr{S}m}. \quad (18)$$

This means that the spline fit for $\mathbf{t}_1$ borrows strength from other functional locations $\mathbf{t}_2$ according to the effective intrafunctional covariance structure $Q_{\mathscr{S}}(\mathbf{t}_1, \mathbf{t}_2)$ that induces smoothing across $\mathbf{t}$ in the spline fits of $f(x, \mathbf{t})$.

Smoothing Parameter of x, $\lambda(\mathbf{t})$, is Nonstationary and Smooth Across $\mathbf{t}$—In our BayesFMM implementation of the smooth nonparametric term $f(x, \mathbf{t})$, we allow the penalized spline for each basis function k to have its own smoothing parameter $\lambda_k = s_k/q_{\mathscr{S}k}$. The basis space model induces a residual error covariance matrix $\text{cov}\{E_i(\mathbf{t}_1), E_i(\mathbf{t}_2)\} = S_{\mathbf{t}_1, \mathbf{t}_2}$ back in the data space, with diagonal elements $s(\mathbf{t})$, and a spline random effect covariance matrix $\text{cov}\{U_{\mathscr{S}m}(\mathbf{t}_1), U_{\mathscr{S}m}(\mathbf{t}_2)\} = Q_{\mathscr{S}}(\mathbf{t}_1, \mathbf{t}_2)$ back in the data space, with diagonal elements $q_{\mathscr{S}}(\mathbf{t})$. Thus, the effective smoothing parameter for the induced spline fit $f(x, \mathbf{t})$ at location $\mathbf{t}$ is given by

$$\lambda(\mathbf{t}) = s(\mathbf{t})/q_{\mathscr{S}}(\mathbf{t}), \quad (19)$$

meaning that the smoothness in x is allowed to vary across $\mathbf{t}$, enabling some parts of the function to be linear with large $\lambda(\mathbf{t})$ and others to be nonlinear with small $\lambda(\mathbf{t})$. Also, this smoothing parameter is not estimated independently for each $\mathbf{t}$, but the off-diagonal elements of $\mathbf{S}$ and $\mathbf{Q}_{\mathscr{S}}$ imply a dependency across $\mathbf{t}$ in $\lambda(\mathbf{t})$, meaning that the model "borrows strength" across $\mathbf{t}$ leading to smoothness in $\lambda(\mathbf{t})$ across $\mathbf{t}$.

We believe this to be the first presentation of a model with such flexibility in the literature, i.e. with $f(x, \mathbf{t})$ varying smoothly across $\mathbf{t}$ with the smoothing parameter in x, $\lambda(\mathbf{t})$, also varying smoothly across $\mathbf{t}$. The FAMM models of Scheipl et al. (2015) and Greven and Scheipl (2017a) estimate terms like $f(x, \mathbf{t})$ that are smooth across both x and $\mathbf{t}$, but utilize an additive penalty term involving marginal smoothing parameters in the x and $\mathbf{t}$ directions, λ_x and $\lambda_{\mathbf{t}}$. This structure does not allow the type of nonstationarities enabled here, which in Section 5 of the supplement we demonstrate are necessary to accurately model the scleral strain MPS data. It may be possible in the FAMM framework to accommodate this type of flexibility by putting a spline on λ_x that varies smoothly across $\mathbf{t}$, but this has not been done in any published paper to date, and it is not clear whether such an approach would be computationally feasible for large functional data sets.

Degrees of Freedom Function $DF(\mathbf{t})$—In the penalized spline literature with penalized spline estimator given by $\hat{f}(x) = \mathcal{B}(\mathcal{B}'\mathcal{B}+\lambda\Omega)'^{1}\mathcal{B}'\mathbf{y} = \mathbf{X}(\lambda)\mathbf{y}$, a standard summary of the nonlinearity of the fit is given by the dimensionality of the projection space given by $DF = \text{trace}\{\mathbf{X}(\lambda)\}$, called the *degrees of freedom* of the fit. A $DF = 2$ indicates a linear model and

$DF \gg 2$ indicates significant nonlinearity. To assess how the degree of nonlinearity of the spline fit $f(x, \mathbf{t})$ varies over $\mathbf{t}$, we can compute the *degrees of freedom function* $DF(\mathbf{t})$ marginally across $\mathbf{t}$ by

$$DF(\mathbf{t}) = \text{trace}[X\{\lambda(\mathbf{t})\}] = \text{trace}[\mathscr{B}\{\mathscr{B}'\mathscr{B} + \lambda(\mathbf{t})\Omega\}^{-1}\mathscr{B}'], \quad (20)$$

with $\lambda(\mathbf{t})$ defined as in (19). In general semiparametric functional mixed models with other levels of random effects to account for interfunctional covariance according to model (21) below, as necessary for modeling our scleral strain MPS data, the derivation for $DF(\mathbf{t})$ is more complex, and outlined in Section 2 of the supplementary materials. Panel (c) of Figure 3 presents $DF(\mathbf{t})$ for the MPS data.

3.4 General Bayesian Semiparametric Functional Mixed Model

In order to model the highly structured scleral strain MPS data set, we need include all of the modeling structures described in the preceding sections, including random effects to capture nested and serial interfunctional correlation and smooth nonparametric smooth covariate effect functions, together in a common BayesFMM model. To highlight its ability to model smooth nonparametric structures as described in Section 3.3, it is useful to adapt the notation of the core BayesFMM model to explicitly include these terms. We term this version of the FMM a *semiparametric functional mixed model* since it includes both linear and smooth covariate effects.

Given a sample of functions $Y_i(\mathbf{t})$; $i = 1, \ldots, N$; $\mathbf{t} \in \mathscr{T}$, with covariates for fixed linear effects X_{ia_l}, $a_l = 1, \ldots, A_l$, smooth nonparametric effects X_{ia_n}, $a_n = 1, \ldots, A_n$, and H levels of random effect covariates Z_{ihm}, $h = 1, \ldots, H; m = 1, \ldots, M_h$, we have the following semiparametric FMM:

$$Y_i(\mathbf{t}) = \sum_{a_l = 1}^{A_l} X_{ia_l} B_{a_l}(\mathbf{t}) + \sum_{a_n = 1}^{A_n} f(X_{ia_n}, \mathbf{t}) + \sum_{h = 1}^{H} \sum_{m = 1}^{M_h} Z_{ihm} U_{hm}(\mathbf{t}) + E_i(\mathbf{t}), \quad (21)$$

with $U_{hm}(\mathbf{t}) \sim GP(\mathbf{0}, Q_h)$ and $E_i(\mathbf{t}) \sim GP(\mathbf{0}, S)$ being mean zero Gaussian processes with covariance surfaces Q_h, $h = 1, \ldots, H$ and S defined on $\mathscr{T} \times \mathscr{T}$.

Using the structures defined in Section 3.3, this model can be directly fit by the BayesFMM software of Morris and Carroll (2006) using the following FMM:

$$Y_i(\mathbf{t}) = \sum_{a_l = 1}^{A_l} X_{ia_l} B_{a_l}(\mathbf{t}) + \sum_{a_n = 1}^{A_n} X_{ia_n} B_{a_n}(\mathbf{t}) + \sum_{a_n = 1}^{A_n} \sum_{m_{a_n} = 1}^{M_{a_n} + 2} Z_{\mathscr{B}m_{a_n}}(X_{ia_n}) U_{\mathscr{S}a_n m_{a_n}}(\mathbf{t}) \quad (22)$$

$$+ \sum_{h = 1}^{H} \sum_{m = 1}^{M_h} Z_{ihm} U_{hm}(\mathbf{t}) + E_i(\mathbf{t}),$$

with M_{a_n} being the number of interior knots for the spline for X_{ia_n}, $Z_{\mathscr{B}m_{a_n}}(X_{ia_n})$ the corresponding Demmler-Reinsch design matrix, $U_{\mathscr{S}a_n m_{a_n}}(\mathbf{t}) \sim GP(\mathbf{0}, Q_{\mathscr{S}a_n})$ the corresponding spline random effect functions, and $U_{hm}(\mathbf{t}) \sim GP(\mathbf{0}, Q_h)$ and $E_i(\mathbf{t}) \sim GP(\mathbf{0}, S)$ modeling the interfunction covariance structure. As described above, the model would be fit in the transformed basis space. We first fit the basis space model with all random effects integrated out, and then sample the spline random effects from their complete conditional distribution while integrating out the other H levels of random effects that capture any interfunctional covariance, and then project back to the function space in order to construct posterior samples of $f(x_{a_n}, \mathbf{t})$ on any desired grid of $\mathbf{t}$.

While omitted from (21) for ease of presentation, this model can also be easily made to include any desired parametric-nonparametric interaction terms, with interaction of parametrically modeled covariate X_{ia_l} and nonparametrically modeled covariate X_{ia_n} being represented by the term $X_{ia_l} f_a(X_{ia_n}, \mathbf{t})$. For X_{ia_l} that are categorical dummy variables, this allows separate nonparametric fits of $(X_{ia_n}, \mathbf{t})$ for different levels of the dummy variable. For continuous X_{ia_l}, this allows the corresponding slope to vary smoothly and nonparametrically with both X_{ia_n} and $\mathbf{t}$. For example, in our scleral strain data, one may wish to include an interaction term to allow the nonparametric age effect to vary across IOP levels. If dummy variables were specified for each IOP level, this would allow separate independent nonparametric age effects for each IOP level. If IOP is modeled continuously via a parametric model like the hyperbolic model described in Section 5, this would allow the hyperbolic coefficients to vary smoothly by age and scleral position, which would be equivalent to nonparametric age effects that vary across IOP but borrow strength from nearby IOP according to the structure induced by the hyperbolic model. In either case, the fixed effect and random spline design matrices corresponding to the $X_{ia_l} f_a(X_{ia_n}, \mathbf{t})$ would be given by $X_{ia_l} X_{ia_n}$ and $X_{ia_l} Z_{\mathscr{S}m}(X_{ia_n})$, respectively, which are straightforwardly included in the FMM. As described in Section 5, we considered these interaction structures, but found they did not appear necessary for representing the scleral strain MPS so were not included in the final model.

4. MODEL SELECTION HEURISTIC FOR SEMIPARAMETRIC BAYESFMM

Given the extensive flexibility of the semiparametric BayesFMM framework, there are a large number of modeling decisions to make. For example, in our sclera strain MPS data set, should the age effect be linear or nonparametric? If nonparametric, should the smoothing parameter in age be allowed to vary across the scleral surface $\mathbf{t}$, or is a common smoothing

parameter across all scleral locations sufficient? Should the fixed IOP effect be linear, hyperbolic, or nonparametric? Should there be an interaction of age and IOP? Should there be a fixed left vs. right eye effect? For the random effect levels, is the subject-specific random effect necessary to account for correlation between right and left eyes for the same subject? Is the correlation across functions from multiple IOP for the same eye sufficiently handled by a compound symmetry structure assuming equal correlations, or is a structure allowing serial correlation necessary? Should this serial correlation be based on a linear, parabolic, or hyperbolic model? These decisions are challenging to make in a simple generalized additive mixed model framework with scalar responses, and become even more challenging in the current setting with complex, high-dimensional functional responses.

There are a some papers in the frequentist literature for performing variable selection in functional regression contexts (Scheipl et al., 2013; Gertheiss, Maity and Staicu, 2013; Brockhaus et al., 2015). However, there is a lack of functional regression model selection methods for MCMC-based fully Bayesian models such as the semiparametric BayesFMM, which present special challenges. One could split the data into training and validation data sets, fit separate MCMC for each prospective model in the training data, and then compute ratios of predicted marginal densities, integrating over MCMC posterior samples, for the validation data, as done in Zhu et al. (2014), for example. These predictive Bayes Factors would provide a rigorous model selection measure, or alternatively parallel MCMC could be run for each prospective model, and a multinomial random variable with Dirichlet prior used to select and perform Bayesian model averaging across models as the chains progress. These strategies might work fine for simple, low dimensional data sets or settings with only a few prospective models, but for the current setting with complex, high-dimensional data, they are impractical.

In this section, we present a model selection heuristic that we have developed that can explore a number of potential model structures to find which seem to be most appropriate for the given data without running any MCMC, and also provides ML and REML estimates that can be used as starting values for the parameters in the BayesFMM. This heuristic is admittedly *ad hoc*, but is based on standard methods and appears to perform well in simulations, and so we believe can be a useful tool for modelers to assess which structures to included in their semiparametric FMM.

Our overall approach is to fit linear mixed models (LMM) to each basis coefficient k using the `lme` function in R (Pinheiro et al., 2017) for each prospective model, and then use a weighted voting scheme based on importance weights for each basis and an adapted Bayesian Information Criterion (aBIC) to obtain probability scores for each prospective model. Here we outline the steps in detail.

1. **Basis transform and importance weights:** Transform the raw functions $Y_i(\mathbf{t})$, $i = 1, \ldots, N$ to the basis space Y_{ik}^*, $k = 1, \ldots, K$, and compute a series of weights w_k that measure the relative importance of each basis for representing the data set. These weights can be computed by $w_k = \sum_i Y_{ik}^{*\,2} / \sum_i \sum_k Y_{ik}^{*\,2}$, with $\sum_k w_k = 1$.

For orthogonal $\psi_k(\mathbf{t})$, the w_k represent the relative percent energy captured by basis coefficient k.

2. **Fit basis-specific LMM and compute aBIC scores:** For each prospective model $\mathcal{M}_c$, $c = 1, \ldots, C$, use `lme` in R (Pinheiro et al., 2017) to fit the corresponding LMM to the data for each basis coefficient $k = 1, \ldots, K$, and compute an adapted version of the BIC ($aBIC_{ck}$), which we define as:

$$aBIC_{ck} = -2\log - \text{likelihood}_{ck} + n_{\text{par},c}\log(N),$$

where the log-likelihood is the marginal likelihood of the fixed effect and variance components of the model with the non-spline random effects integrated out conditional on the data for basis k, N is the total number of observations in the dataset for basis k, and $n_{\text{par},c}$ is the total number of parameters of model c. As discussed by Vaida and Blanchard (2005) and Spiegelhalter et al. (2002), selection of the effective number of parameters for LMM or Bayesian hierarchical models is tricky and context dependent. If inference is desired on the random effects themselves, then counting only fixed effects and variance components as parameters is not appropriate. In our setting, we are not interested in the random effects at levels capturing interfunctional correlation as we work with the marginalized model, but for nonparametric terms $f(x, \mathbf{t})$ we are clearly interested in the "random effects" corresponding to the spline coefficients. Thus, we count the number of parameters to be the sum of the number of fixed effects, the number of variance components and the estimated degrees of freedom for each nonparametric term. This last term adjusts appropriately for the extra parameters of the spline fits even thought they are captured as random effects in the LMM.

3. **Use weighted voting scheme to rank models:** We compute a probability weight P_c for each model $\mathcal{M}_c$; $c = 1, \ldots, C$:

$$P_c = \sum_{k=1}^{K} w_k I\{c = \arg\min_{c'} aBIC_{c'k}\}$$

This procedure is applied in two steps: first assessing different fixed effect models (including parametric and/or nonparametric effects), and second assessing various random effect structures for capturing interfunctional variability while conditioning on the best fixed effect model.

In principle, $I\{c = \text{argmin}_{c'} \ aBIC_{c'k}\}$ indicates whether the model $\mathcal{M}_c$ is the best in terms of $aBIC$ for the data set on the kth basis. Therefore, the P_c is computed via a weighted voting scheme, an aggregated measure of proportion of times $\mathcal{M}_c$ is the best model across the all basis coefficients, with basis coefficients weighted by w_k. In this way, the model fit for basis coefficients that account for a larger proportion of the total variability in the data count more towards the overall model selection. Empirically, we have found this weighted voting scheme seems to work well, as it is robust in the sense of not allowing any one basis

function, especially one explaining a relatively low proportion of total energy for the given data set, to dominate the model selection because of an extreme *aBIC* score. This can also be applied using alternative measures (e.g. *aAIC*). We acknowledge that this strategy is *ad hoc* and more rigorous model selection methods for settings like this are needed, but we believe it can provide useful guidance for modelers and performs well in simulations, as described below.

A word of caution—This heuristic is meant for selecting among various different modeling structures for specified covariates as done for our case study, or perhaps could be used to select among a few covariates, but *it is not intended for high-dimensional variable selection across many potential predictors*. In such settings, consideration of a large number of models and only fitting the "best" one can dramatically inflate type I error rates, and post-selection inference as described in Berk et al. (2013) would need to be considered.

Simulation Study on Model Selection—We conducted a simple simulation study to investigate the performance of this model selection heuristic. We considered four different models:

- Model 1 (null model): $Y_{ij}(\mathbf{t}) = B_0(\mathbf{t}) + E_{ij}(\mathbf{t})$,

- Model 2 (linear age effect): $Y_{ij}(\mathbf{t}) = B_0(\mathbf{t}) + X_{\text{age},i}B_{\text{age}}(\mathbf{t}) + E_{ij}(\mathbf{t})$,

- Model 3 (nonparametric age effect): $Y_i(\mathbf{t}) = f(X_{\text{age},i}, \mathbf{t}) + E_{ij}(\mathbf{t})$, and

- Model 4 (linear age effect, random effect): $Y_{ij}(\mathbf{t}) = B_0(\mathbf{t}) + X_{\text{age},i}B_{\text{age}}(\mathbf{t}) + U_j(\mathbf{t}) + E_{ij}(\mathbf{t})$.

We fit each of these models to the scleral strain data and used the fitted model as the truth, and simulated 100 replicate data sets for each model. For each simulated data set, we fit each of these four models and performed the model selection procedure using P_c to select the best model. In all four scenarios, this procedure selected the correct model 100/100 times. Average values of P_c for each model can be found in Section 4 of the supplementary materials, and Section 6 of the supplement investigates issues that can arise in variable selection of GAMMs when considering nonparametric smooth terms of subject-specific covariates in models including subject-level random effects.

5. GLAUCOMA SCLERAL STRAIN MPS CASE STUDY

5.1 Overview of Glaucoma Scleral Strain MPS Data

As described in Section 1, glaucoma is characterized by ONH damage related to IOP but its etiology is not fully known. Researchers have hypothesized that biomechanics of the scleral region close to the ONH may modulate the effect of IOP on the ONH, and thus may play an important role in glaucoma. In particular, the scleral surface is elastic so deforms under pressure, which can partially relieve IOP-induced forces on the eye, including the ONH. Thus, studies of these properties could reveal insights into the etiology of glaucoma.

Recently, novel custom instrumentation was developed that can precisely measure the mechanical strain in the posterior human sclera at a fixed level of IOP (Fazio, Bruno,

Reynaud, Poggialini and Downs, 2012; Fazio, Grytz, Bruno, Girard, Gardiner, Girkin and Downs, 2012). Briefly, the posterior 1/3 of the eye is clamped, sealed, and pressurized. Next, the eye is preconditioned, and then pressurized from 7 mmHg to 45 mmHg using an automated system with computer feedback control, while scleral surface displacements are measured by a laser speckle interferometer. This device measures a light interference distribution that is used to reconstruct the surface displacement field in three dimensions with nanometer-scale precision. These displacements were processed as described in Fazio, Bruno, Reynaud, Poggialini and Downs (2012) to compute the 3D strain tensor, a 3×3 matrix summarizing the displacement in the meridional, circumferential, and radial directions, continuously around the outer scleral surface. The leading eigenvalue of the strain tensor, called the maximum principal strain (MPS), was computed on a grid of scleral locations for 120 circumferential locations $\phi \in (0°, 360°)$ and 120 meridional locations $\theta \in (9°, 24°)$, where $\theta = 0°$ corresponds to the ONH. This yields MPS functions defined on a grid of 14,400 points on the scleral surface that comprises a partial spherical domain.

Using this custom instrumentation, Fazio et al. (2014) conducted a study to investigate age-related changes in the scleral surface strain. They obtained twenty pairs of eyes from normal human donors in the Lions Eye Bank of Oregon in Portland, OR and the Alabama Eye Bank in Birmingham, AL. For each subject, the MPS measurements were obtained as described above at nine different levels of IOP (7, 10, 15, 20, 25, 30, 35, 40, and 45 mmHg) for both left and right eyes. The data for both eyes from one subject failed a quality control check, so was excluded from analysis, as did one of the eyes from four other subjects. Thus, the data we analyzed consisted of 34 eyes from 19 subjects. With 14,400 measurements for each of 9 IOP levels $\times$ 34 eyes, this data set contained over 4.5 million measurements. Let $Y_{ijp}(\mathbf{t})$ be the MPS for eye j for subject i under IOP level p at scleral location indexed by $\mathbf{t} = (\theta, \phi)$, which on the sampling grid can be written as a vector $\mathbf{y}_{ijp}$ of length 14,400. The primary goals are to study MPS, assessing how it varies around the scleral surface, across IOP, and with age. The hypothesis is that MPS is greater near the ONH, which could confer a protective effect, and that MPS tends to decrease with age, which could contribute to increased stress on the ONH thus conferring increased glaucoma risk.

5.2 Model Specification

Basis Transform—Various criteria can be considered when choosing which basis to use within the BayesFMM framework, including sparse representation, fast calculation, richness for representing the functional parameters at the various levels of the models, ability to capture the key visual features of the observed functions, and flexibility for representing the intrafunctional correlation in the data. Multiresolution bases like wavelets have advantages for many of these considerations, so we constructed a custom rectangular wavelet basis defined on the cylindrical spherical projection of the partial scleral space $\mathbf{t} = (\theta, \phi)$, which is a tensor transform computed by successively applying 1D wavelet transforms to the meridional and circumferential directions.

Tensor Wavelets for Scleral Space—Specifically, we constructed $\psi_k(\mathbf{t}) = \psi_k(\theta, \phi)$ as a tensor wavelet, $\psi_k(\mathbf{t}) = \psi_{k1}^{\theta}(\theta) \otimes \psi_{k2}^{\phi}(\phi)$, with meriodonal wavelet $\psi_{k_1}^{\theta}(\theta)$ being a db3 wavelet

basis with three vanishing moments, reflection boundary condition, 5 levels of decomposition and circumferential wavelet $\psi_{k_2}^{\phi}(\phi)$ being a db3 wavelet with three vanishing moments, 5 levels of decomposition and periodic boundary conditions since its domain is circular, covering the entire circumferential space. This transform yielded a basis $\{\psi_k(\mathbf{t}); k = 1, \ldots, K = 17, 185\}$. While single-indexed here for simplicity of presentation, these basis coefficients can be written as multiindexed by circumferential scale $j_1 = 0, \ldots, 5$, meriodonal scale $j_2 = 0, \ldots, 5$, circumferential locations $k_1 = 1, \ldots, K_{1j_1}$ and meriodonal locations $k_2 = 1, \ldots, K_{2j_2}$ with $K = \Sigma_{j_1}\Sigma_{j_2} K_{1j_1} * K_{2j_2}$. The levels $j_1 = 0$ and $j_2 = 0$ correspond to the father wavelet coefficients at the lowest level of decomposition, and the other j_1 and j_2 index the corresponding mother wavelets at increasing levels of scale. With $\mathbf{y}_{ijp} = Y_{ijp}(\mathbf{t})$ being the observed function for subject i, eye j, and IOP p on the scleral surface sampling grid of size $T = 14, 400$ written in vector form, this basis representation can be written as $\mathbf{y}_{ijp} = \mathbf{y}_{ijp}^{*} \mathbf{\Psi}$, where $\mathbf{\Psi}$ is a $K \times T$ basis matrix with elements $\psi_k(\mathbf{t})$ and $\mathbf{y}_{ijp}^{*}$ is a vector of K corresponding basis coefficients. Because of the structure of the tensor transform, if we unstack $\mathbf{y}_{ijp}$ into a $(T_1 = 120) \times (T_2 = 120)$ matrix $\mathbf{Y}_{ijp}$ with rows indexed by equally spaced meriodonal locations $\theta_1 = 9°, \ldots, \theta_{120} = 24°$ and columns by equally spaced circumferential locations $\phi_1 = 0°, \ldots, \phi_{120} = 360°$, we could write $\mathbf{y}_{ijp}^{*} = \text{vec}(\mathbf{\Psi}_{\theta}\mathbf{Y}_{ijp}\mathbf{\Psi}_{\phi}')$, where $\text{vec}(\bullet)$ is the column-stacking vectorizing operator, $\mathbf{\Psi}_{\theta}$ is the $K_1 \times (T_1 = 120)$ basis matrix corresponding to the meriodonal wavelet $\psi_{k_1}^{\theta}(\theta)$ and $\mathbf{\Psi}_{\phi}$ the $K_2 \times (T_2 = 120)$ basis matrix corresponding to the circumferential wavelet $\psi_{k_2}^{\phi}(\phi)$. In principle, spherical wavelets could be used as the transforming basis, but currently available software does not handle transforms for part of the sphere, and since we only model θ over a limited range of $9° - 24°$, the distortion from using the basis on the projection and not the true spherical geodesic is not great.

Some eyes had technical processing artifacts that resulted in a spike of extremely high MPS at some local set of scleral locations, typically close to the boundary. Given the multiresolution nature of the wavelet transform, these artifacts were captured by wavelet coefficients at extremely high frequency scales. Given the relatively smooth nature of most of the MPS functions, these wavelet coefficients were essentially zero for all eyes except for those with the artifact which yielded very large coefficients. Thus, we removed these artifacts by filtering out any wavelet coefficients with *extremely* skewed distributions for which the mean across all samples was more than $100\times$ the median. As seen in the supplementary file *RawMPScurves.zip*, with illustration in Supplemental Figure 26, this strategy effectively removed the outlying spikes without substantively affecting MPS values for other scleral locations. We also applied the joint wavelet compression strategy described in Morris et al. (2011) to obtain a reduced dimension near-lossless basis function to use, and found a subset of 269 wavelet coefficients that jointly preserved $> 99.5\%$ of the total signal energy for each eye, and an average of $> 99.9\%$, leading to $> 50 : 1$ compression. As shown in Supplemental Figure 26 and *RawMPScurves.zip*, the data projected into the basis is essentially identical to the raw data, demonstrating its near-lossless nature. We considered this basis for our model.

We modeled these basis coefficients using the basis transform modeling approach described in Section 3.1. Besides providing a relatively sparse representation and enabling the adaptive removal of spiky artifacts, this transform being a location-scale decomposition allowed nonstationary intra-scleral correlations and adaptive borrowing of strength across scleral locations. File `intrafunctional_correlation.mp4` in the supplement contains a movie file demonstrating the form of the intrafunctional correlation structure induced by this choice, computed by constructing the basis and basis transform matrices Ψ and Ψ^-, respectively, and applying (5) to the basis space covariances at the various random effect and residual error levels of the model. For illustration, panels (e) and (f) of Figure 3 contains plots of this surface at a particular scleral location for two of the random levels.

We also considered using principal components computed on the wavelet-transformed (and compressed) data, similar to the strategy used in Meyer et al. (2015), which implies applying a singular value decomposition to the wavelet-space data matrix, and then using the resulting eigenvectors to construct the empirical basis functions $\psi_k(\mathbf{t})$ that are used for the BayesFMM modeling. In this case, we kept $K = 29$ basis functions that explained $> 99.5\%$ of the total variability in the data set according to the scree plot, which as estimated by four-fold cross validation retained a minimum of 96.7% of the total energy for each eye, so is somewhat near-lossless. We used the wavelets for our primary analysis given that it yielded a richer basis set for representing the various functional parameters at various levels of the models, but for sensitivity we also presented results using the BayesFMM using these wavelet-regularized principal components in Section 7 of the supplement, as well as other summaries including the induced intrafunctional correlation structures.

Model Selection—We applied the model selection heuristic described in Section 4 to help select the structures in the semiparametric FMM to include in the FMM. We first determined which fixed effect covariates to include and what their functional forms should be. Here we summarize the results, for which more details are provided in Section 3 of the supplementary materials. Three different fixed effects were considered: age, IOP, and eye (left vs. right). For the form of the age effect, we considered two possibilities: linear or nonparametric. For the form of the IOP effect, we considered three different possibilities: linear, hyperbola, or nonparametric. Models without the eye effect were also compared. As a result, we compared 12 different models for the fixed effect selection. It turned out that the model with the nonparametric age effect, the hyperbolic IOP effect, and no eye effect showed the highest P_c when using *aBIC*, and the model with nonparametric age effect, hyperbolic IOP effect, and a left vs. right eye effect had the highest P_c when using *aAIC*. For our primary analysis, we considered the model with no left vs. right eye effect, since there is no strong scientific rationale for such an effect, and present the other model as a sensitivity analysis in Section 7 of the supplementary materials. We also assessed whether the smoothing parameter for the nonparametric age effect should be constant or vary around the sclera, and found that the sclerally varying smoothing parameter was clearly necessary for good fit, as detailed in Section 5 of the supplement. Once we selected the main fixed effects, we assessed whether the interaction term between age and IOP was needed, and our model selection heuristic suggested the interaction was not necessary. Finally, with the selected fixed terms, we compared several different random effect distributions to capture the

interfunctional covariance structure. Two different levels of random effects were considered: the subject-level random effect and the serial eye-level random effect. For the form of the eye-level random effect in terms of IOP as illustrated in Section 3.2, we considered three different forms: constant (compound symmetry), linear, or hyperbola. Our model selection heuristic selected the eye-level random effect with the hyperbolic IOP effect, but not the subject-level random effect.

Model—Thus, the final fitted semiparametric FMM was:

$$Y_{ijp}(\mathbf{t}) = B_0(\mathbf{t}) + B_1(\mathbf{t})G_1(p) + B_2(\mathbf{t})G_2(p) + f(X_{\text{age},i}, \mathbf{t}) + U_{ij}(\mathbf{t}) + U_{ij1}(\mathbf{t})G_1(p) + U_{ij2}(\mathbf{t})G_2(p) + E_{ijp}(\mathbf{t}),$$

(23)

with $X_{\text{age},i}$ is the age for subject i, and $G_1(p)$ and $G_2(p)$ are the values of the orthogonalized hyperbolic basis corresponding to IOP $= p$ as described below. This is equivalent to the FMM:

$$Y_{ijp}(\mathbf{t}) = B_0(\mathbf{t}) + B_1(\mathbf{t})G_1(p) + B_2(\mathbf{t})G_2(p) + B_3(\mathbf{t})X_{\text{age},i} + \sum_{m=1}^{M+2} Z_{\mathcal{B}m}(X_{\text{age},i})U_{\mathcal{S}m}(\mathbf{t}) + U_{ij}(\mathbf{t})$$

$$+ U_{ij1}(\mathbf{t})G_1(p) + U_{ij2}(\mathbf{t})G_2(p) + E_{ijp}(\mathbf{t}),$$

(24)

with $Z_{\mathcal{B},m}(x)$ the Demmler-Reinsch basis functions corresponding to M interior knots on x, $U_{\mathcal{S}m}(\mathbf{t}) \sim GP\{\mathbf{0}, Q_{\mathcal{S}}\}$, $U_{ij}(\mathbf{t}) \sim GP\{\mathbf{0}, Q_0\}$, $U_{ij1}(\mathbf{t}) \sim GP\{\mathbf{0}, Q_1\}$, $U_{ij2}(\mathbf{t}) \sim GP\{\mathbf{0}, Q_2\}$, and $E_i(\mathbf{t}) \sim GP(\mathbf{0}, S)$, and with $Q_{\mathcal{S}}, Q_0, Q_1, Q_2$, and S being covariance surfaces defined on $\mathcal{T} \times \mathcal{T}$. Following the guidelines suggested by Ruppert et al. (2003), we chose $M = 5$ equally spaced knots over X_{age}.

Parameterization of IOP effect—From a preliminary investigation in which we fit separate models to each scleral location $\mathbf{t}$, we found that the serial IOP effects were well modeled by a hyperbola of special form $Y = b_0 + b_1 p + b_2 p^{-1}$, with an average R^2 of 0.98 across all eyes and scleral locations, and this form was also chosen by the model selection heuristic. To accommodate model fitting without having to include a covariance between b_1 and b_2, we utilized orthogonalized versions of these predictors: $G_1(p) = \sqrt{2}/2 X_{1,p} - \sqrt{2}/2 X_{2,p}$ and $G_2(p) = \sqrt{2}/2 X_{1,p} + \sqrt{2}/2 X_{2,p}$ where $X_{1,p}$ is a standardized version of p and $X_{2,p}$ is a standardized version of p^{-1}.

Basis Space Model—We used the fast rectangular 2D wavelet transform to compute the basis coefficients from the raw functions, equivalent to the matrix multiplication $\mathbf{y}_{ijp}^* = \mathbf{y}_{ijp}\mathbf{\Psi}^-$ with $\mathbf{\Psi}^- = \mathbf{\Psi}'(\mathbf{\Psi}\mathbf{\Psi}')^{-1}$ with $\mathbf{y}_{ijp}^*$ a vector of length $K = 269$ with elements Y_{ijpk}^*, $k = 1, \ldots, K$. We then fit the basis-space version of model (24):

$$Y_{ijpk}^* = B_{0k}^* + B_{1k}^*G_1(p) + B_{2k}^*G_2(p) + B_{3k}^*X_{\text{age},i} + \sum_{m=1}^{M+2} Z_{\mathscr{B}m}(X_{\text{age},i})U_{\mathscr{S}mk}^* + U_{ijk}^* + U_{ij1k}^*G_1$$

$$(p) + U_{ij2k}^*G_2(p) + E_{ijpk}^*, \text{ with}$$

$$U_{\mathscr{S}mk}^* \sim N(0, q_{\mathscr{S}k}), U_{ijk}^* \sim N(0, q_{0k}), U_{ij1k}^* \sim N(0, q_{1k}), U_{ij2k}^* \sim N(0, q_{2k}), \text{ and } E_{ijpk}^* \sim N(0, s_k).$$

(25)

Prior Specification—We specified vague conjugate inverse Gamma priors for each basis space variance component $\{q_{\mathscr{S}k}, q_{0k}, q_{1k}, q_{2k}, s_k\}$ in the model, with prior mode being the REML starting values and with effective sample size of 2, e.g. $s_k \sim \text{InverseGamma}(a_s, b_s)$ with $a_s = 2$ and $b_s = 3 * \hat{s}_k$ where $\hat{s}_k$ is the REML starting values for s_k. We used spike-slab priors for the basis-space fixed effects $\{B_{ak}^*, a = 0, \ldots, 3\}$, with regularization parameters $\{\pi_{aj}, \tau_{aj}\}$ varying over predictor $a = 0, \ldots, 3$, with *regularization sets* $j = 1, \ldots, J = 36$ determined by the tensor wavelet scale levels, with $j = 0$ for $j_1 = j_2 = 0$, $j = 1$ for $j_1 = 0$, $j_2 = 1, \ldots, j = J = 36$ for $j_1 = j_2 = 5$. We estimated the regularization parameters using the empirical Bayes algorithm specified in Morris and Carroll (2006). To assess sensitivity of results to these choices of regularization parameters, we also ran the model doing no additional regularization (beyond wavelet compression) by setting $\pi_{aj} \equiv 1$ and $\tau_{aj} \equiv 10^6$. Results are provided in Section 7 of the supplementary materials.

Model Fitting—We ran an MCMC to obtain posterior samples of the parameters of model (25) $\{B_{\bullet k}^*, q_{\bullet k}, s_k\}$ from the marginalized version of this model with $U_{\bullet k}^*$ all integrated out. We fit a total of 10,000 posterior samples after a burn-in of 5000, thinning by keeping every 10. We then sampled the spline random effects $U_{\mathscr{S}mk}^*$ from their complete conditional distributions with the other random effects still integrated out, which are conjugate multivariate normal Gibbs steps as detailed in Section 1 of the supplementary materials, from which posterior samples of $f_k^*(x) = xB_{3k}^* + \sum_m Z_{\mathscr{B}m}(x)U_{\mathscr{S}mk}^*$ were subsequently constructed for a grid of ages x of size 71 corresponding to ages 20–90. Let $\mathbf{F}_g^*$ be a $71 \times (K = 269)$ matrix representing posterior sample g of the basis space nonparametric age effect, $g = 1, \ldots, 1000$. We then transformed this back to the data space via $\mathbf{F}_g = \mathbf{F}_g^*\mathbf{\Psi}$ to obtain the $71 \times (T = 14, 400)$ matrix of posterior samples of the nonparametric age effect $f(X_{\text{age}}, \mathbf{t})$ in model (23), and similarly transforming the other fixed effects back to the data space to get

posterior samples of $\{B_0(\mathbf{t}), B_1(\mathbf{t}), B_2(\mathbf{t})\}$ on the sampling grid of $\mathbf{t}$ and used for posterior inference.

On a laptop computer, the entire analysis took 7hr39min on a single core, with the basis transform taking 1m37s, model selection 22m48s, each MCMC iteration 0.77s with 15,000 iterations taking 3hr15m, and the postprocessing including inverse basis transform of posterior samples and key inferential summary calculations 4hr. Many other summaries and plots were computed for the purposes of this paper for sensitivity and illustration of the deep properties of the modeling framework at more computational expense, but these additional analyses are not necessary for analysis of the data. The Metropolis-Hastings acceptance probabilities ($\approx 0.85 - 0.95$) were reasonable, and Geweke convergence statistics (median 0.01, $Q_{.025} = -1.98, Q_{0.975} = 1.94$) across the many parameters in the model showed that most were within roughly 2 standard deviations of zero, and only 4.4% of the corresponding p-values were less than 0.05, so suggest reasonable MCMC convergence (see Section 9 of the supplementary materials for more details). We simulated virtual MPS functions for hypothetical subjects with specified age and IOP from the posterior predictive distribution of the data (see Section 10 of the supplementary materials), and found that the simulated MPS data are visually similar to the MPS data from real eyes, suggesting the BayesFMM model with the tensor wavelet bases was sufficiently flexible to capture the salient features of these data, and lending support to its use for inference. We share these pseudo data on github (https://github.com/MorrisStatLab/SemiparametricFMM), as well as the real data and scripts to perform all analyses.

5.3 Scientific Results

Nonparametric MPS function of age—First, we computed the posterior mean MPS curve as a function of age over the entire meridional and circumferential domain at each level of IOP. As a thorough summary of the model fit, we generated a plot of these fits for each scleral location, and joined these together to make a digital movie file `MPSvAGE-wave.mp4` contained in the supplemental materials. Figure 2 shows a snapshot of the movie at the scleral location indicated by the white dot. Panel (a) depicts a polar azimuthal projection of the fitted MPS function for a left eye from a subject of age 90yr under 45 mmHg of IOP. Note that MPS is higher near the ONH, as expected as a protective effect. Panel (b) shows the posterior mean degrees of freedom (DF) of the nonparametric age effect. We see strong nonlinear age effects in scleral regions close to the ONH and towards the inferior and nasal regions of the sclera, while many other regions show linear or almost linear age effects. The right nine panels contain the fitted nonparametric age effect at nine different IOP levels at the scleral position indicated by the white dot. In each panel, the black dots are the raw data, the solid blue line is the estimated nonparametric mean MPS function of age, and the solid and dashed red lines correspond to joint and point-wise 95% credible bands of the mean MPS curve, with joint bands computed as described in Meyer et al. (2015). We see from this plot how the MPS increases with IOP, and that the hyperbolic model seems to capture the rate of increase very well. From this plot and the movie in the supplement showing results stepping across the scleral locations, we see this model fits the data for all IOP and scleral locations remarkably well in spite of the fact that independent splines were not fit to the data for each IOP and scleral location separately, but rather is the

result of the complex joint unified model (23) that borrows strength from other IOP according to the modeled hyperbolic serial effect and from other scleral locations according to the basis functions. The model also borrows strength from other nearby scleral locations according to the basis functions in estimating the nonlinearity of the age effect, as seen in the local smoothness of the $DF(\mathbf{t})$ plot.

Induced Intrafunctional and Interfunctional Covariance Structures—To assess the intrafunctional covariance structures induced by our tensor wavelet bases, we used equation (5) to estimate the scleral space intrafunctional covariance matrices $Q_d(\mathbf{t}_1, \mathbf{t}_2)$, $d = 0, 1, 2$ and $S(\mathbf{t}_1, \mathbf{t}_2)$. Supplementary Figure 22 plots the diagonals of these matrices representing the variances at the various hierarchical levels as a function of scleral location $\mathbf{t}$. Note that the variance for the eye intercept $Q_0(\mathbf{t}, \mathbf{t})$ is an order of magnitude greater than that of the eye-level IOP coefficients $Q_1(\mathbf{t}, \mathbf{t})$ and $Q_2(\mathbf{t}, \mathbf{t})$, which are in turn an order of magnitude greater than the residual error variance $S(\mathbf{t}, \mathbf{t})$. Also, note that these covariances vary around the scleral surface, with locations near the ONH having greater levels of variability. The supplement also contains a movie file `interfunctional_cor.mp4` that represents the corresponding intrafunctional correlation surfaces induced by our model by stepping around scleral locations and for each plotting a heatmap representing the correlation of the indicated location with all other scleral locations. For illustration, panels (e) and (f) of Figure 3 show the correlation of a specific scleral location $\mathbf{t}^*$ (indicated by the white dot) with all other scleral locations $\mathbf{t}$ at the eye intercept $Q_0(\mathbf{t}^*, \mathbf{t})$ and residual error $S(\mathbf{t}^*, \mathbf{t})$ levels, and Supplementary Figure 18 contains these plus those for the eye hyperbolic random effects $Q_1(\mathbf{t}^*, \mathbf{t})$ and $Q_2(\mathbf{t}^*, \mathbf{t})$ for this location, and `Intrafunctional_correlations.mp4` contains a movie file showing all scleral locations. Note how our model captures local intrafunctional correlation, and the strength and tails of this correlation are allowed to vary by scleral location and hierarchical level. Supplementary Figure 19 includes equivalent plots for the wavelet-regularized principal component basis functions, and `Intrafunctional_correlations-pc.mp4` contains a movie showing all scleral locations. Note that the PC basis functions, although global, induce intrafunctional correlation surfaces that are dominated by the local correlation among nearby scleral locations that is also captured by the tensor wavelet bases.

We also computed the induced interfunctional serial correlation across MPS curves for different IOP for the same eye using the formulas contained in Section 3.2. The supplement contains a movie file `Intra_IOP_corr.mp4` that plots the variance, var$\{Y_{ijp}(\mathbf{t})|\ IOP = p, \mathbf{t}\}$, as a function of IOP and scleral location and the serial correlation across IOP, corr$\{Y_{ijp}(\mathbf{t}),\ Y_{ijp'}(\mathbf{t})\}$, as a function of IOP as it varies around the scleral surface $\mathbf{t}$. Note the form of the serial correlation induced by the hyperbolic model, and how it is able to vary yet borrow strength across scleral locations. Panel (d) of Figure 3 portrays this serial correlation at a single scleral location.

Inference on functionals of the parameters—While Figure 2 and the accompanying movie file provide a thorough summary of the age effect on MPS estimated by our model, for interpretability it may be useful to aggregate results over IOP and/or scleral locations. One major advantage of our fully Bayesian approach is that we are able to compute posterior

samples and obtain estimates and inference for any functional of the model parameters. First, to aggregate information across all IOP, we considered the area under the MPS vs. IOP curve, defined as follows:

$$\text{AUC}(X_{\text{age}}, \mathbf{t}) = f(X_{\text{age}}, \mathbf{t}) + \int_{7}^{45} \{B_1(\mathbf{t})G_1(p) + B_2(\mathbf{t})G_2(p)\} dp.$$

This integral summarizes the total MPS behavior over the range of IOP in the study, which covers the practical range of IOP values in this context. The integral is estimated numerically for each MCMC sample to yield posterior inference on AUC. We plotted the posterior mean AUC vs. age curve and corresponding posterior pointwise and joint credible bands for each scleral location, and assembled into a digital movie file `AUCvAGE-wave.mp4` in the supplementary materials, and for illustration panel (b) of Figure 3 contains this plot for a single scleral location. Although our model was not fit to the AUC data, but rather the raw data for each IOP, note how the model provides a nonparametric smooth fit of AUC vs. age for each scleral location, and again these fits borrow strength across scleral locations. From these results, we see that that aggregating over IOP, the MPS tends to decrease with age at most scleral locations, especially near the ONH. With MPS decreasing with age, the eye seemingly becomes less elastic and less able to absorb IOP, potentially exposing the ONH to IOP-induced damage over time.

Given the nonlinearity of the age fit, it may be instructive to directly look at the rate of decline of MPS over age. We can do this by computing inference on the derivative of the AUC curve with respect to X_{age}. Given the lack of an IOP $\times$ age interaction in our final fitted model, it follows that $\partial \text{AUC}(X_{\text{age}}, \mathbf{t})/\partial X_{\text{age}} = \partial f(X_{\text{age}}, \mathbf{t})/\partial X_{\text{age}}$. Using the fact that $f(X_{\text{age}}, \mathbf{t}) = \beta_0(\mathbf{t}) + X_{\text{age}}\beta_3(\mathbf{t}) + Z_{\mathscr{B}}(X_{\text{age}})\mathbf{u}_{\mathscr{S}}(\mathbf{t})$ as described in Section 3.3, with $Z_{\mathscr{B}}(X_{\text{age}}) = \{\mathscr{B}_1(X_{\text{age}}), ..., \mathscr{B}_{M+4}(X_{\text{age}})\}'Z_{\Omega}\text{diag}(d_1^{-1/2}, ..., d_{M+2}^{-1/2})$, it can be easily seen that

$$\frac{\partial \text{AUC}(X_{\text{age}}, \mathbf{t})}{\partial X_{\text{age}}} = \beta_3(\mathbf{t}) + \frac{\partial Z_{\mathscr{B}}(X_{\text{age}})}{\partial X_{\text{age}}}\mathbf{u}_{\mathscr{S}}(\mathbf{t}). \quad (26)$$

Our R scripts contain calculations of $\dfrac{\partial Z_{\mathscr{B}}(X_{\text{age}})}{\partial X_{\text{age}}}$, which come from Wand and Ormerod

(2008). By applying (26) to the posterior samples of $\beta_3(\mathbf{t})$ and $\mathbf{u}_{\mathscr{S}}(\mathbf{t})$, we could obtain posterior samples of this derivative for each scleral location, although we do not present those results here. Since the nonparametric fits were done placing a smoothness penalty on the age effects themselves, the derivatives may appear slightly undersmoothed. If one had primary interest in estimating these derivatives smoothly, they could do so by simply choosing higher order penalties in the spline fits. A third order penalty would produce smoothness in the first derivative.

Aggregated summaries over functional regions—Although our model fits the entire data set over all scleral locations, for ease of interpretation researchers at times would like to

look at estimates and inference for aggregated scleral regions. Given the hypothesis that scleral strain is most important near the ONH, we averaged results over all circumferential regions to obtain results as a continuous function of meridional distance from the optic nerve head. Figure 4 depicts the posterior mean AUC as a function of age and distance from the ONH, aggregating over circumferential regions. We can see how MPS is higher near the ONH and decreases moving away from the ONH, potentially providing a protective effect to the ONH. Younger individuals have high MPS levels at scleral locations extending well out from the ONH, while for middle age individuals the regions of high MPS does not extend out far from the ONH, and for older individuals the MPS is quite low even close to the ONH. This coincides with the increased glaucoma risk in older individuals.

We also summarized the results aggregating over the innermost region closest to the optic nerve head, which is called peripapillary(PP) region and also within the adjacent region, called the mid-peripheral(MP) region. Here we present aggregated AUC results in both PP and MP regions. Figure 5 depicts the age effects aggregating AUC over the PP and MP regions, with the top two plots containing the posterior mean fits and the bottom containing the derivatives with respect to age. The blue line contains the posterior mean fit. The dotted and solid red lines indicating 95% pointwise and joint credible intervals, respectively. From this we can clearly see that the MPS is systematically higher in the PP region closest to the ONH than in the MP region further away, potentially conferring a protective effect. The MPS decreases with age in both regions, but the decrease is substantially steeper for the all-important PP region close to the ONH. This effect is nonlinear, with the rate of decrease accelerating throughout middle age (40–60 years old), an age at which glaucoma risk increases substantially.

These analyses confirm the hypothesis that MPS is higher in scleral regions closer to the ONH, decreases with age, and this decrease with age is more pronounced near the ONH. This agrees with the notion that biomechanical changes in the sclera may contribute to increased glaucoma risk.

Sensitivity Analyses—Section 7 of the Supplement and Supplementary Figures 2–25 present extensive results for the alternative model with the left vs. eye effect, with the alternative values for the prior shrinkage hyperparameters {τ_{aj}, π_{aj}}, and using the wavelet-regularized principal components as the projected basis. Substantive results do not change, so we see our conclusions are not driven by these choices.

6. DISCUSSION

In this paper, we demonstrated how to adapt the BayesFMM modeling framework to account for serial interfunctional correlation and smooth nonparametric covariate functional effects and applied it to an innovative glaucoma study investigating MPS of scleral strain tensors. We found that MPS is maximized near the ONH. We also found that MPS decreases with age, especially in regions closest to the ONH, and the decrease in MPS accelerates throughout middle-age. This could contribute to the increased glaucoma risk seen in elderly. The age effect on MPS tends to be non-linear near the ONH, especially towards inferior

and nasal sides, while the other scleral regions show the age effect close to linear. Focal glaucoma damage is most often observed in the inferior quadrant of the ONH.

While motivated by the glaucoma data application, the BayesFMM framework presented here is extremely general, with the ability to be used with any near-lossless basis tranform and applicable to many types of complex, high dimensional functional data of modern interest including wearable computing data, genome-wide data, proteomics data, geospatial time series data, neuroimaging data, and many others. Using this framework, one can not only accommodate nonparametric functional fixed effects, but also model serially correlated functions through functional growth curve effects. Our model allows the nonparametric fits, smoothness, and interfunctional correlations to potentially vary over the functional domain, which is necessary for good fit to these data and likely many other complex functional data sets. In addition, we introduced a model selection heuristic that can be used to select among fixed and random effects and decide whether they should be linear, parametric, or nonparametric, and whether the fits or smoothness should be constant or vary around the functional domain before running the MCMC.

On our github page (https://github.com/MorrisStatLab/SemiparametricFMM), we share the Matlab files required to fit these models, including scripts to apply the model selection heuristic, links to automated software to perform the MCMC, and scripts to compute posterior samples and posterior inference for the nonparametric functional effects and all summary plots included in this paper. The model is set up using *lmer* style model statements (Bates et al., 2015). The software can be used to fit models with any fixed or random effect function covariates, with different basis functions, and for functions on domains of any dimension or measure given suitable basis functions for the corresponding space. The method is efficient enough to feasibly apply to very large data sets like our glaucoma data set here with over 4.5 million observations. While it took a relatively long time to run the MCMC using a single core, we do not think this is inordinate considering the cost and time to collect these data and the extensive inferential summaries provided by the model fits. This run time was orders of magnitude less than a conceptually simpler approach of applying `lme` to each scleral location, applying a 2d smooth, and using a bootstrap for inference, which by our calculations would take 10 weeks to compute using 1000 bootstrap samples on the same computer we used for the BayesFMM model. If further speed is desired for our method, the model fitting is highly parallelizable and the MCMC code has cluster computing capabilities so can be sped up using GPU or cluster computing resources. While to code we share only utilizes a single core, future updates of the software will enable distributed computing for faster calculations for big data sets. We are in the process of extending the package to fit these models and including many other features of the BayesFMM framework in other publications but not included in this paper. We anticipate this package, which will be available in Matlab and R, will greatly enhance the usability of the method, and expect this package to be completed and freely available in the near future.

One significant benefit of our fully Bayesian approach to fitting these semiparametric functional mixed models is that we can produce inference on any parameters in the model, or any functional or aggregation of these parameters, and this inference integrates over the various sources of interfunctional variability in the model and over the uncertainty in

estimating the covariance parameters and level of nonlinearity of the spline fits. This allows us to perform a flexible, thorough analysis in the entire functional domain, yet produce inference and results for aggregated summaries that may be more interpretable to investigators. This raises the obvious question, "What is the advantage of fitting the entire functional model? Why not just compute the aggregated summaries and model those using standard tools?" The answer to this question is multi-faceted.

First, if one only looked at specific aggregated summaries, they may miss insights that could have been gleaned from their data but were not captured by these summaries. For example, if only looking at the PP and MP regions, one may miss out on different MPS behavior in the inferior and nasal regions of the sclera, which may be important. By modeling the scleral function in its entirety, we are able to examine the entire domain to ensure we are not missing out on any insights, and then we can still produce inference for any desired aggregated summaries for ease of interpretation, as well. This concept is also relevant in other application areas where summaries are used because of the complexity and high dimensionality of the raw functional data. This approach of flexible modeling followed by inferenct on summary measures can capture the best of both worlds, analyzing the entire function yet producing inference for summary measures interpretable for the scientific subject area.

Second, any aggregation draws arbitrary boundaries in the function space. MPS for scleral locations at the boundary of the PP and MP regions are highly correlated with each other, and yet PP and MP extracted summaries arbitrarily separate them from each other. The flexible functional modeling framework introduced in this paper models the entire functional space, yet captures and accounts for intrafunctional correlation through the chosen basis functions. This allows a smooth borrowing of strength from nearby locations, and if location-scale bases like wavelets are used, then this borrowing of strength can be adaptive, able to accommodate spatially heterogenous functions for which some functional regions are more correlated than others. In principle, accounting for this intrafunctional correlation leads to greater efficiency, as has been shown in various contexts.

While flexible, the BayesFMM framework presented here has some limitations and drawbacks, including the need to choose a common basis to use for transformation at all levels of the model, the independence in the basis space assumption that can limit certain types of intrafunctional covariance structure depending on the choice of basis, and the computational intensity from having to run a full MCMC to obtain estimates and inference for model quantities. It is designed for representing functional data sampled on a common fine grid, so is not suitable for sparsely sampled functional data or functional data for which the individual functions are sampled on wildly different sampling grids, a setting in which the Scheipl et al. (2015) and Greven and Scheipl (2017*a*) framework is well-suited. The model selection heuristic is *ad hoc*, and should not be used to select over a large number of variables.

In spite of these limitations and drawbacks, the BayesFMM modeling framework is very general, and this paper can serve as a template for how to utilize this modeling framework to model complex functions with various types of interfunctional correlation structures. It has

the potential to impact many areas of science yielding complex functional data, and the analysis presented here illustrates a rigorous, thorough workflow to analyze the entire data set, extract many types of information from them, yet provide interpretable graphical summaries and inferential results desired by investigators.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Aston JA, Chiou J-M, Evans JP. 2010; Linguistic pitch analysis using functional principal component mixed effect models. Journal of the Royal Statistical Society, Series C. 59(2):297–317.

Bates D, Mächler M, Bolker B, Walker S. 2015; Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software. 67(1):1–48.

Berhane K, Molitor N-T. 2008; A Bayesian approach to functional-based multilevel modeling of longitudinal data: applications to environmental epidemiology. Biostatistics. 9(4):686–699. [PubMed: 18349036]

Berk R, Brown L, Buja A, Zhang K, Zhao L. 2013; Valid post-selection inference. The Annals of Statistics. 41(2):802–837.

Bhattachary A, Pati D, Pillai NS, Dunson DB. 2015; Dirichlet-Laplace priors for optimal shrinkage. Journal of the American Statistical Association. 110(512):1479–1490. [PubMed: 27019543]

Brockhaus S, Scheipl F, Hothorn T, Greven S. 2015; The functional linear array model. Statistical Modeling. 15(3):279–300.

Brumback BA, Rice JA. 1998; Smoothing spline models for the analysis of nested and crossed samples of curves. Journal of the American Statistical Association. 93(443):961–976.

Burgoyne CF, Downs JC. 2008; Premise and Prediction How Optic Nerve Head Biomechanics Underlies the Susceptibility and Clinical Behavior of the Aged Optic Nerve Head. Journal of Glaucoma. 17(4):318–328. [PubMed: 18552618]

Carvahlo CM, Polson NG, Scott JG. 2010; The horseshoe estimator for sparse signals. Biometrika. 97(2):465–480.

Cederbaum J, Pouplier M, Hoole P, Greven S. 2015; Functional linear mixed models for irregularly or sparsely sampled data. Statistical Modeling. 16(1):67–88.

Chen K, Müller H-G. 2012; Modeling Repeated Functional Observations. Journal of the American Statistical Association. 107:1599–1609.

Demmler A, Reinsch C. 1975; Oscillation matrices with spline smoothing. Numerische Mathematik. 24:375–382.

Eilers PHC, Marx BD. 1986; Flexible smoothing with B-splines and penalties. Statistical Science. 11(2):89–121.

Fazio MA, Bruno L, Reynaud JF, Poggialini A, Downs JC. 2012; Compensation method for obtaining accurate, sub-micrometer displacement measurements of immersed specimens using electronic speckle interferometry. Biomedical Optics Express. 3:407–417. [PubMed: 22435090]

Fazio MA, Grytz R, Bruno L, Girard MJA, Gardiner S, Girkin CA, Downs JC. 2012; Regional variations in mechanical strain in the posterior human sclera. Investigative Ophthalmology & Visual Science. 53:5326–5333. [PubMed: 22700704]

Fazio MA, Grytz R, Morris JS, Bruno L, Gardiner SK, Girkin CA, Downs JC. 2014; Age-related changes in human peripapillary scleral strain. Biomechanics and Modeling in Mechanobiology. 13(3):551–563. [PubMed: 23896936]

George EI, McCulloch RE. 1993; Variable selection via Gibbs sampling. Journal of the American Statistical Association. 88(423):881–889.

Gertheiss J, Goldsmith J, Crainiceanu C, Greven S. 2013; Longitudinal scalar-on-functions regression with application to tractography data. Biostatistics. 14:447–461. [PubMed: 23292804]

Gertheiss J, Maity A, Staicu AM. 2013; Variable selection in generalized functional linear models. Stat. 2(1):86–101. [PubMed: 25132690]

Goldsmith J, Crainiceanu CM, Caffo B, Reich D. 2012; Longitudinal penalized functional regression for cognitive outcomes on neuronal tract measurements. JRSS-C. 61(3):453–469.

Goldsmith J, Kitago T. 2016; Assessing systematic effects of stroke on motor control by using hierarchical function-on-scalar regression. Journal of the Royal Statistical Society, Series C. 65(2): 215–236.

Greven S, Crainiceanu C, Caffo B, Reich D. 2010; Longitudinal functional principal component analysis. Electronic Journal of Statistics. 4:1022–1054. [PubMed: 21743825]

Greven S, Scheipl F. 2017a; A general framework for functional regression modeling. Statistical Modeling. 17(1–2):1–35.

Greven S, Scheipl F. 2017b; Rejoinder. Statistical Modeling. 17(1–2):100–113.

Griffin JE, Brown PJ. 2010; Inference with normal-gamma prior distributions in regression problems. Bayesian Analysis. 5:171–188.

Hasenstab K, Scheer A, Telesca D, Sugar CA, Jeste S, DiStefano C, Senturk D. 2017; A Multi-Dimensional Functional Principal Components Analysis of EEG Data. Biometrics. 73:999–1009. [PubMed: 28072468]

Hastie, T, Tibshirani, R, Friedman, J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2. New York Berlin Heidelberg: Springer-Verlag; 2009.

Islam MN, Staicu A-M, van Heugten E. 2016Longitudinal dynamic functional regression. :1–23.

Kundu MG, Harezlak J, Randolph TW. 2016; Longitudinal functional models with structured penalties. Statistical Modeling. 16(2):114–139.

Laird NM, Ware JH. 1982; Random-effects models for longitudinal data. Biometrics. 38(4):963–974. [PubMed: 7168798]

Li Y, Guan Y. 2014; Functional Principal Component Analysis of Spatio-Temporal Point Processes with Applications in Disease Surveillance. Journal of the American Statistical Association. 109:1205–1215. [PubMed: 25368436]

Meyer MJ, Coull BA, Versace F, Cinciripini P, Morris JS. 2015; Bayesian function-on-function regression for multilevel functional data. Biometrics. 71:563–574. [PubMed: 25787146]

Morris JS. 2015; Functional Regression. Annual Review of Statistics and Its Application. 2:321–359.

Morris JS. 2017; Comparison and contrast of two general functional regression modeling frameworks. Statistical Modeling. 17(1–2):59–86.

Morris JS, Baladandayuthapani V, Herrick RC, Sanna P, Gutstein H. 2011; Automated analysis of quantitative image data using isomorphic functional mixed models, with application to proteomics data. Annals of Applied Statistics. 5:605–1125. [PubMed: 21857889]

Morris JS, Carroll RJ. 2006; Wavelet-based functional mixed models. Journal of the Royal Statistical Society, Series B. 68:179–199.

Morris JS, Vannucci M, Brown PJ, Carroll RJ. 2003; Wavelet-based nonparametric modeling of hierarchical functions in colon carcinogenesis. Journal of the American Statistical Association. 98:573–583.

Park SY, Staicu A-M. 2015; Longitudinal functional data analysis. Stat. 4:212–226. [PubMed: 26594358]

Park T, Casella G. 2008; The Bayesian Lasso. Journal of the American Statistical Association. 103(482):681–686.

Pinheiro, J; Bates, D; DebRoy, S; Sarkar, D. R Core Team. nlme: Linear and Nonlinear Mixed Effects Models. 2017. R package version 3.1–131 URL:https://CRAN.R-project.org/package=nlme

Ruppert, D, Wand, MP, Carroll, RJ. Semiparametric Regression. New York: Cambridge University Press; 2003.

Scheipl F, Gertheiss J, Greven S. 2016; Generalized functional additive mixed models. Statistical Modeling. 10:1455–1492.

Scheipl F, Kneib T, Fahrmeir L. 2013; Penalized likelihood and Bayesian function selection in regression models. AStA Advances in Statistical Analysis. 97(4):349–385.

Scheipl F, Staicu A-M, Greven S. 2015; Functional additive mixed models. Journal of Computational and Graphical Statistics. 24(2):477–501. [PubMed: 26347592]

Shou H, Zipunnikov V, Crainiceanu CM, Greven S. 2015; Structured functional principal component analysis. Biometrics. 71(1):247–257. [PubMed: 25327216]

Sigal IA, Flanagan JG, Ethier CR. 2005; Factors Influencing Optic Nerve Head Biomechanics. Investigative Ophthalmology & Visual Science. 46(11):4189–4199. [PubMed: 16249498]

Spiegelhalter DJ, Best NG, Carlin BP, vd Linde A. 2002; Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society, Series B. 64(4):583–639.

Vaida F, Blanchard S. 2005; Conditional Akaike information for mixed-effect models. Biometrika. 92(2):351–370.

Wahba G. 1978; Improper priors, spline smoothing,a nd the problem of guarding against model errors in regression. Journal of the Royal Statistical Society, Series B. 40(3):364–372.

Wand MP, Ormerod JT. 2008; On semiparametric regression with OSullivan penalised splines. Australian & New Zealand Journal of Statistics. 50:179198.

Wood, SN. Generalized additive models: An introduction with R. CRC Press; 2006.

Zhang, L, Baladandayuthapani, V, Zhu, H, Baggerly, KA, Czerniak, BA, Morris, JS. Technical report. The University of Texas M.D. Anderson Cancer Center; 2014. Functional CAR models for large spatially correlated functional datasets.

Zhu H, Brown PJ, Morris JS. 2011; Robust, adaptive functional regression in functional mixed model framework. Journal of the American Statistical Association. 106:1167–1179. [PubMed: 22308015]

Zhu H, Versace F, Cinciripini F, Morris JS. 2014Robust and Gaussian spatial functional regression models for analysis of event-related potentials. Technical Report.

Zipunnikov V, Caffo B, Yousem DM, Davatzikos C, Schwartz BS, Crainiceanu C. 2011; Multilevel Functional Principal Component Analysis for High-Dimensional Data. JCGS. 20(4):852–873.

Zipunnikov V, Greven S, Shou H, Caffo B, Reich DS, Crainiceanu C. 2014; Longitudinal High-Dimensional Principal Components Analysis with Application to Diffusion Tensor Imaging of Multiple Sclerosis. The Annals of Applied Statistics. 8(4):2175–2202. [PubMed: 25663955]
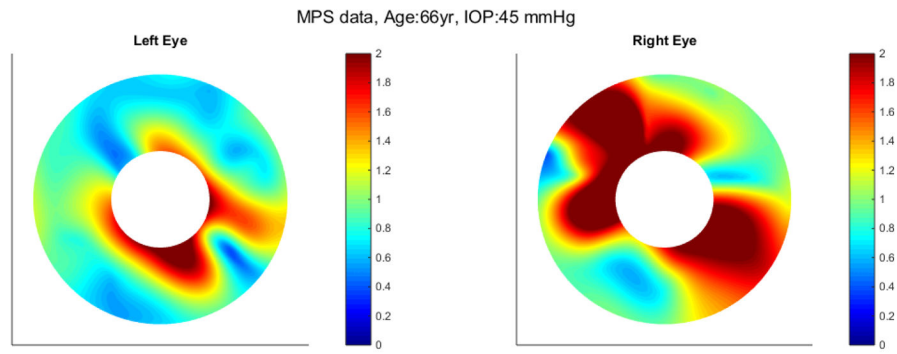
Figure 1. Data plot
Polar azimuthal projection of partial spherical MPS function for both left and right eyes from one subject of age 66yr under 45 mmHg of intraocular pressure.
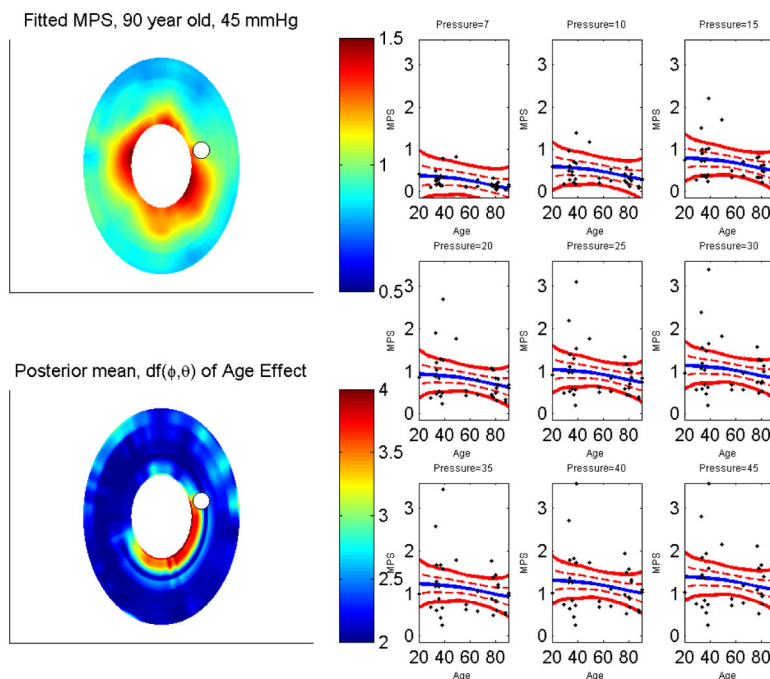
Figure 2.
(a) Polar azimuthal projection of fitted MPS function for left eye from one subject of age 90yr under 45 mmHG of IOP. (b) posterior mean of degrees of freedom of nonparametric MPS fit of age. The right nine panels depict estimated nonparametric MPS fit of age for all nine IOP levels at the scleral location indicated by the white dot in (a) and (b), along with the raw data indicated by the black dots.
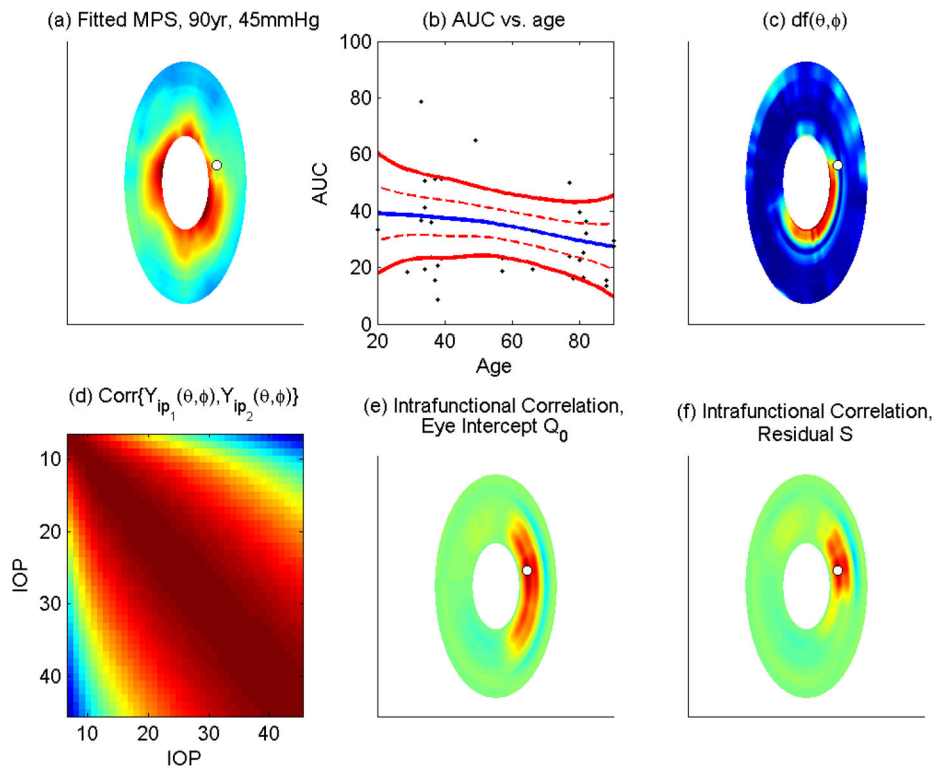
Figure 3. Key Summaries of Fitted Model

Shows key summaries at the scleral position marked by the open circle, including (a) the fitted MPS for a 90yr old with IOP=45mmHg, (b) the nonparametric MPS vs. age curve using the AUC to integrate over IOP, with blue line being posterior mean, dotted and solid red lines pointwise and joint credible bands, and the raw data (computing AUC for this scleral location for each eye) indicated by dots, (c) the degrees of freedom of the nonparametric age fit as a function of the scleral location, (d) the serial correlation across IOP induced by the model at this scleral position, and (e) and (f) being the intrafunctional correlation surface induced by our model and choice of tensor basis for the eye-to-eye random intercept and residual error levels, respectively, at this scleral position. The file `combo_plot.mp4` in the supplement is a movie file showing how these summaries vary across scleral locations.
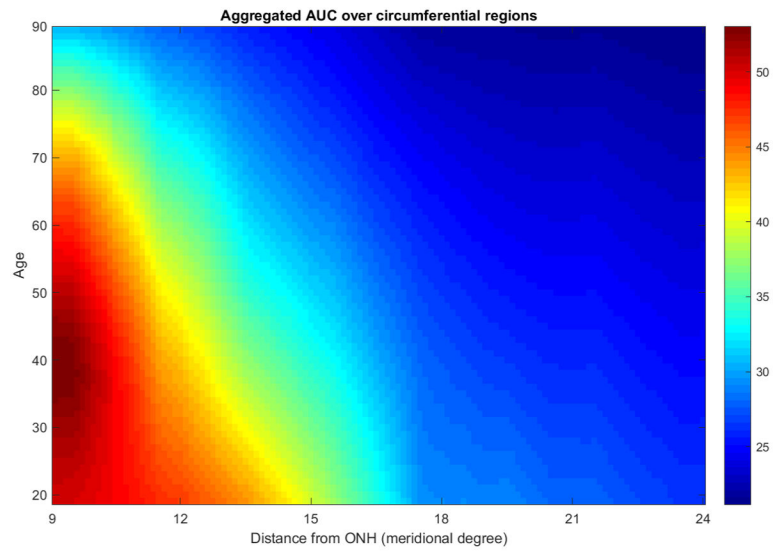
Figure 4. Posterior mean AUC (of MPS) as a function of age and distance from ONH
These results are obtained by aggregating posterior samples over all circumferential regions.
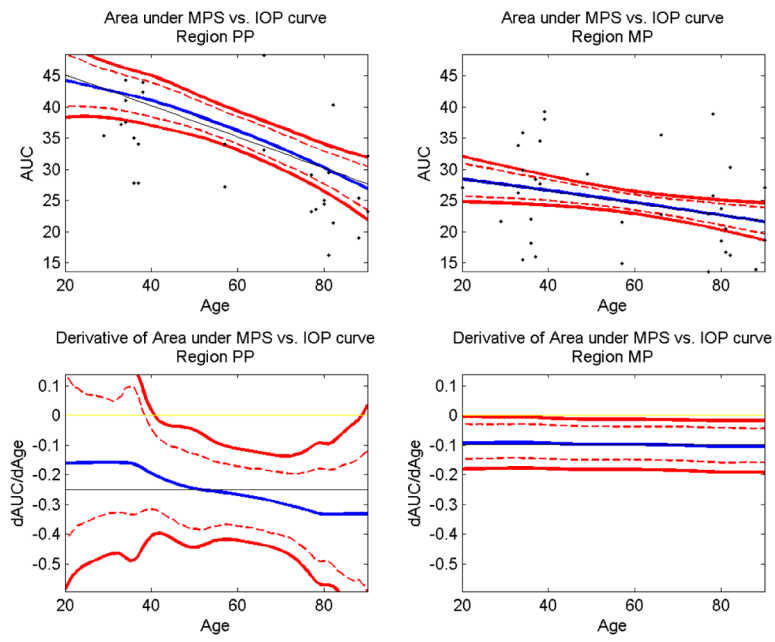
Figure 5. Aggregated AUC summaries

AUC aggregated over peripapillary (PP) region adjacent to the ONH and the mid-peripheral(MP) region just beyond PP. (a) and (b) show AUC summaries over the two regions. (c) and (d) show derivatives of aggregated AUC summaries presented in (a) and (b).