SPATIAL FUNCTIONAL DATA MODELING OF PLANT REFLECTANCES

BY PHILIP A. WHITE,*†, HENRY FRYE‡, MICHAEL F. CHRISTENSEN§, ALAN E. GELFAND§, AND JOHN A. SILANDER, JR.¶

Brigham Young University †, University of Connecticut‡, and Duke University§

Plant reflectance spectra – the profile of light reflected by leaves across different wavelengths - supply the spectral signature for a species at a spatial location to enable estimation of functional and taxonomic diversity for plants. We consider leaf spectra as “responses” to be explained spatially. These reflectance spectra are also functions over wavelength that respond to the environment.

Our motivating data are gathered for several plant families from the Greater Cape Floristic Region (GCFR) in South Africa and lead us to develop rich novel spatial models that can explain spectra for genera within families. Wavelength responses for an individual leaf are viewed as a function of wavelength, leading to functional data modeling. Local environmental features become covariates. We introduce a wavelength - covariate interaction since the response to environmental regressors may vary with wavelength, as may variance. Formal spatial modeling enables prediction of reflectances for genera at unobserved locations with known environmental features. We incorporate spatial dependence, wavelength dependence, and space-wavelength interaction (in the spirit of space-time interaction). We implement out-of-sample validation for model selection, finding that the model features above are informative for the functional data analysis. We supply ecological interpretation of the results under the selected model.

1. Introduction. The reflectance of the surface of a material is the fraction of incident electromagnetic radiation reflected at the surface. It is a function of the wavelength (or frequency) of the light, its polarization, and angle of incidence. The reflectance as a function of wavelength is called a reflectance spectrum. The literature on reflectances is substantial, with a large portion focused on the interaction of electromagnetic energy with the atmosphere and terrestrial objects, e.g., reflectances associated with different land cover/vegetation types. Typically, such reflectances are gathered by satellites, aircraft, and ground-level sensors. The focus of this manuscript is on plant reflectances gathered for plants at leaf level.

*Corresponding Author

Keywords and phrases: environmental regressors, functional data analysis, heterogeneity, hierarchical model, interaction, kernel weighting, reflectance, spatial confounding
The importance of leaf level reflectance modeling arises because the scales at which remote sensing devices detect reflectance spectra often do not match those relevant to ecological scales (Gamon et al., 2020). For example, a satellite imager can measure the reflectance signal for an entire 30m² pixel, but this signal is a composite of all the different spectral signatures of the plant species within that area. To disentangle which plants are on the ground, remote sensing scientists use spectral unmixing techniques which rely on spectral libraries (Quintano et al., 2012; Shi and Wang, 2014). These libraries are collections of pure endmembers, i.e., the pure reflectance spectra of leaf surfaces, which serve as representative spectra for different plant functional types, taxonomic groups, and/or individual species. The use of such leaf spectral libraries in ecology or biodiversity science has been termed as a “spectranomic” approach (Asner and Martin, 2016).

Being able to statistically predict leaf reflectance spectra across environmental gradients is a major advancement because it enables prediction-based spectral libraries that could be used in the validation and inference of remote sensing data at large spatial extents. Large hyperspectral remote sensing efforts are already under way, e.g., NASA’s Surface Biology Geology satellite mission (Cawse-Nicholson, 2021), making pressing the need to predict spectral signals of plants.

Furthermore, leaf-level spectra have become an invaluable tool to capture the diversity in leaf traits that have accumulated over the course of seed plant evolution (Reich et al., 2003; Cornwell et al., 2014), enabling estimation of functional diversity (Kokaly et al., 2009; Schneider et al., 2017) and taxonomic diversity (Clark, Roberts and Clark, 2005; Cavender-Bares et al., 2016a). These spectra provide drivers for ecosystem processes (Schweiger et al., 2018) and guide conservation (Asner et al., 2017).

Statistical analyses of plant reflectance spectra have been limited to treating spectra as functional predictors of scalar variables such as plant traits. That is, we are in the realm of functional linear regression modeling. Modeling approaches rely on dimension reduction, e.g., spline basis representation (Ordoñez et al., 2010), partial least squares regression (Doughty, 2017), partial least squares-discriminant analysis (Cavender-Bares et al., 2016b), or some form of machine learning (Féret, 2019). Some approaches generate hypothetical leaf reflectance spectra from physical first principles, i.e., radiative transfer models (Jacquemoud and Baret, 1990; Jacquemoud and Ustin, 2019a). However, these are not functional response models driven by environment. That is, our contribution is to model plant reflectance curves as a functional response variable (as a function of wavelength), in particular, at genus level within family. We incorporate local spatial environmental covariates as regressors.

We introduce the following innovations, motivated by careful exploratory data analysis. We specify a functional data model incorporating spatial random effects to predict reflectance curves at genera level, at locations without collected plant
samples. We also capture wavelength dependence through random effects. For further enrichment, we add space-wavelength interaction (in the spirit of space-time interaction) by constructing a space-wavelength random effect through wavelength kernel convolutions of spatial Gaussian processes. In general, this random effect has nonseparable covariance and is wavelength nonstationary. Additionally, we model the variance to be heterogeneous across wavelengths. Also, expecting that the reflectance response to environmental regressors may vary with wavelength, we include wavelength-covariate interactions. Lastly, since the rich space-wavelength modeling essentially annihilates significance of covariate/wavelength effects, we present a novel orthogonalization to remove spatial and functional confounding between random effects and environmental regressors.

Functional data analysis (FDA) is well established for analyzing data representing curves/surfaces varying over a continuum. The physical continuum over which these functions are defined is often time but here, it is wavelength. Pioneering work for FDA is attributed to Ramsey and Silverman (e.g., Ramsay, 2005; Ramsay and Silverman, 2007). The field has undergone rapid growth, and numerous applications have been found in areas such as imaging (Locantore et al., 1999) (including MRI brain imaging (Tian et al., 2010)), finance (Laukaitis, 2008), climatic variation (Besse, Cardot and Stephenson, 2000), spectrometry data (Reiss and Ogden, 2007), and time-course gene expression data (Leng and Müller, 2006). For a more comprehensive overview of applications, see Ullah and Finch (2013).

Explicit modeling of functional data is usually carried out by specifying functions in one of two ways: (i) as finite linear combinations of some set of basis functions or (ii) as realizations of some stochastic process. A key feature of functional data analysis implementation is some version of dimension reduction to specify functions. Here, we have random functions over a wavelength span as well as over a spatial region. We combine both approaches, using basis functions over wavelength with process realizations over space to build space by wavelength regressions over environment.

We work with plant reflectances gathered from the Greater Cape Floristic Region (GCFR) in South Africa. We present an extensive cross-validation study for model selection across a rich collection of models, showing the ability of our space-wavelength modeling to predict reflectances well for genera within a family at unobserved locations. We discuss our findings for three plant families found within the GCFR, providing the most detail for the most prevalent family (Asteraceae).

The format of the paper is as follows. Section 2 describes the collected data. Section 3 offers a broad exploratory data analysis to motivate the features we incorporate in our modeling. Section 4 explains our modeling, model comparison, and presents a novel orthogonalization for functional regression coefficients. Section 5 discusses the GCFR data analysis while a brief Section 6 offers a summary
and suggestion for future work. Substantial detail of our exploratory analysis, as well as model sensitivity analysis, has been placed in the Supplemental Material. However, we comment on relevant details throughout the main manuscript.

2. The Dataset. We work with plant reflectances gathered from the Greater Cape Floristic Region (GCFR) in South Africa, see Figure 1. Reflectances were measured with a USB-4000 Spectrometer (manufactured by Ocean Optics) using a leaf clip attachment. Sun leaves from the top of each selected canopy were measured. The spectrometer has a range of 450-950 nanometers (nm) with a total of 500 reflectance measurements. Plant reflectance is viewed as a function of wavelength $t$, across the window $t \in [450, 950]$, typically referred to as a spectral signature.

With interest in a spatial model for plant reflectance that enables prediction of reflectance for genera within a family at unobserved locations, we work with adjacent subregions of the GCFR characterized by a fynbos landscape, known as the Hantam-Tanqua-Rogeveld (HTR) and Cederberg. Three prevalent families that often characterize landscapes in this area are Aizoaceae, Asteraceae, and Restionaceae (Slingsby and Wistow, 2014). These families have broad overlap in their reflectances (Figure 2). However, a linear discriminant analysis (LDA) to predict these families based on their reflectances yields clear separation of the groups, demonstrating that reflectances can be used to effectively predict taxonomic differences. More precisely, a structured classification by family using plant samples was conducted employing LDA and reveals the separation between families (See Supplemental Material for full details).

Much of the observed variation across the three families is likely due to differences in composite leaf traits though isolating the relative impact of each trait on the reflectances is beyond our intentions here. However, it is established that reflectance variation is a signal of leaf trait variation (e.g., anatomical, physiological, and structural traits (Jacquemoud and Ustin, 2019b)) and can be influenced by the environmental factors (e.g., climate and soil) that the plants inhabit.

We present brief exploratory data analysis for all plant families in Section 3. However, we focus our model development, comparison, and results on the Asteraceae family, the most prevalent plant family in this dataset. For the Aizoaceae and Restionaceae families, we limit our results to comparing the estimated effect of environmental variables over the reflectance spectrum (See Section 5.1).

3. Exploratory Data Analysis and Modeling. We explore the characteristics of plant reflectances for the three families given above (Aizoaceae, Asteraceae, Restionaceae) in the HTR and Cederberg areas. We retain the entire dataset because it is somewhat small from a spatial perspective. Note that the spatial domains for the three families do not overlap well (Figure 1), so we fit each family separately when implementing our spatial modeling. Moreover, the between-family correlations in the reflectance function are generally weak (See
Supplemental Material).

The number of genera with observed reflectances within each family is: Aizoaceae - 16, Asteraceae - 38, and Restionaceae - 10. The Supplemental Material provides: (i) the proportion of sites where each family is present, (ii) the number of sites with one, two, or three families, and (iii) a more detailed breakdown of family co-occurrence. To summarize, Aizoaceae and Restionaceae rarely co-occur; in fact, Restionaceae is mostly limited to the Cederberg region apart from a few HTR observations. Replication at the genus level is uncommon and even more uncommon at the species level.

3.1. Data Locations. In Figure 1, we show all locations, where reflectances are observed with sites coded by region (shape) and family (color). We also plot locations coded by the number of families observed at that site. In the HTR and Cederberg regions, only 22 of the 133 sites have more than one reflectance spectrum for a given species; only 27 of the 183 total species (across all families) are observed at more than one site. This suggests that species-level modeling is infeasible. The Supplemental Material offers more commentary on data locations and duplication.

3.2. Reflectance spectra and Environmental Variables. To visualize the form and variability in reflectance spectra, we plot all of the curves by family in Figure 2 along with plots of the genus-specific means. We can see that the family-specific means do not capture the spread of the variability seen in all the curves while the genus-specific means show nearly the same variability for all of the curves.

To assess within reflectance function variability as well as between-function variability, we calculate binned standard deviations for every curve. For these binned standard deviations, we estimate a smooth family-specific average standard deviation. Additionally, we calculate the family-specific between-curve standard deviation. These are plotted in Figure 3 and show that variability within reflectance spectrum changes with wavelength and, perhaps, with family. In addition, the variability between curves changes as a function of wavelength and differs by family. These findings lead us to impose heterogeneity in variance across wavelength, adopting wavelength varying variance curve models on the log scale.
Given these plots, we are led to four modeling needs: (i) to allow for family and genus differences, (ii) to model heterogeneity for the reflectance across wavelength (iii) to capture between-curve variability through spatial modeling and/or environmental variables, and (iv) to adopt heteroscedastic errors since reflectances at lower wavelengths (< 500 nm) appear to be more volatile.

For each family, we calculate the correlations between the environmental variables (see Supplemental Material) and the observed log-reflectances, using wavelength bins (See Figure 4), to assess whether this relationship changes with wavelength. We find consequential changes in correlation as a function of wavelength. The strongest correlations are of magnitude 0.3 to 0.4.

**4. Spatial Wavelength Modeling.** Functional data modeling for our spatial reflectance spectra data is motivated by the foregoing exploratory analyses. We model families separately, at genus level, treating species within genus as replicates. We utilize elevation, annual precipitation, rainfall concentration, and minimum January temperature as environmental predictors. We introduce wavelength dependent variances to account for heterogeneity. Model choice focuses on four issues: (i) do we need wavelength dependent regression coefficients? (ii) do we need genus specific wavelength random effects? (iii) do we need genus specific spatial random effects? (iv) how do we specify space-wavelength interaction?
In Sections 4.1 and 4.2, we elaborate the modeling, while Section 4.3 takes up model comparison, yielding the model for which results are presented. For the model selected as best in Section 4.3, we discuss prior distributions and prediction in Section 4.4.

4.1. Model Development. For a given family, let \( i \) denote genera within the particular family, let \( j \) denote replicates within genus. Let \( s \) denote spatial location and \( t \) denote wavelength. There is severe imbalance in the data. The genera observed vary across locations and the number of replicates observed within a genus varies considerably across the locations.

Altogether, our most general model for \( \log \) reflectance takes the form:

\[
Y_{ij}(s, t) = \mu_i(s, t) + \gamma_i(t) + \alpha_i(s) + \eta(s, t) + \epsilon_{ij}(s, t)
\]

Specifically, with regard to the site level covariates, \( X(s) \), we write the mean \( \mu_i(s, t) = \delta_i + X(s)^T \beta(t) \), where, hierarchically, \( \delta_i | \mu_\delta \sim N(\mu_\delta, \sigma_\delta^2) \). We have family level regression coefficients, \( \beta(t) \), which vary with wavelength. (In a setting with richer data at genus-level, introducing \( \beta_i(t) \) would be of ecological interest.)

So, a first model choice clarification is whether constant coefficients are adequate or whether wavelength varying coefficients are needed. Our EDA (Figure 4) suggests the latter, and there is also supporting evidence/suggestion in the literature (Jacquemoud and Ustin, 2019c). We do not consider these coefficients at genus level; with the very irregular observation (including absence) of genera across locations, we cannot learn about coefficients at genus scale. However, we can learn about genus specific intercepts, the \( \delta_i \).

Further, we introduce genus level spatial (\( \alpha_i(s) \)) and wavelength (\( \gamma_i(t) \)) random effects but family level space-wavelength interaction effects, \( \eta(s, t) \). In the Sup-
plemental Material, we provide empirical semivariograms over wavelength bins to motivate the need for spatial random effects. Although our models are explicitly functional, we note that the empirical semivariograms show different spatial patterns for different wavelength bins, as well as residual dependence by genus and wavelength. An additive model (removing $\eta(s, t)$), therefore, seems inadequate; the $\eta$’s allow the functional model for the reflectances to vary more adaptively over space. However, $\eta(s, t)$ is not genus specific. While we have enough data to examine additivity in wavelength and spatial random effects at genus scale, we are unable to find genus level explanation for the interaction given the limited genus duplication. For all three families, most genera are represented three or fewer times in the dataset. Thus, despite potential ecological interest, genus-level modeling will yield limited benefit only for some model components. So, two model choice comparisons are whether the $\gamma$’s and or $\alpha$’s should be genus specific?

Heterogeneity in the variance arises through the $\epsilon_{ij}(s, t)$ terms where we would have $\text{var}(\epsilon_{ij}(s, t)) = \sigma^2(t)$. We could accommodate this heteroscedasticity using a log GP for $\sigma^2(t)$, or perhaps simply binned variances over suitable wavelength bins. As a compromise, for simplicity and flexibility, we specify $\log(\sigma^2(t))$ to be a piecewise linear spline. For all model terms, we present explicit specification in Section 4.2.

4.2. Explicit Specifications. To examine sensitivity of model fit to change in specifications, we provide a sensitivity analysis for the Asteraceae family in the Supplemental Material using using average deviance, the deviance information criterion (DIC), and estimated model complexity (Spiegelhalter et al., 2002). In Section 4.3, we use cross-validation to justify the inclusion/exclusion of various model terms.

We specify each $\alpha_i(s)$ as a genus-level mean 0 Gaussian process (GP) with exponential covariance function. The GPs are conditionally independent across genera given a shared decay and shared scale parameter. Specifically,

$$\alpha_i(s) \sim \mathcal{GP}(0, \sigma^2 \alpha e^{-\phi \alpha d(s, s')})$$

Because our goal is prediction, the exponential covariance function is expected to yield interpolation very similar to that for any choice of Matérn covariance function (Zhang, 2004).

For wavelength terms in the model ($\gamma_i(t)$ and $\eta(s, t)$), we adopt process convolutions because of their simple connection to GPs; the kernels of the process convolution connect the low-rank process to the GP covariance (Higdon, 2002). Our goal in using process convolution is solely dimension reduction; alternatively, spline or wavelet basis function representations could be adopted. Because of the close connections between basis function approaches and Gaussian processes (see
Kimeldorf and Wahba, 1970; Higdon, 2002; Guo, 2002; Morris and Carroll, 2006), the choice of dimension reduction approach is not expected to be consequential. Moreover, there are connections between M-splines (scaled B-splines) and kernel convolutions (Ramsay, 1988; Higdon, 1998; White, Keeler and Rupper, 2021).

We specify $\gamma_i(t)$ using process convolution of normal random variables (Higdon, 1998, 2002). We adopt wavelength knots $t_{\gamma}^1, ..., t_{\gamma}^J$, spaced every 25 nm from 437.5-962.5 nm (22 in total). In our sensitivity analysis, we find that including fewer knots in $\gamma_i(t)$, e.g., spaced every 50 nm or more, led to worse model fit, while including more knots did not improve model fit. In addition, we found that the Gaussian kernel was preferred to a double exponential kernel. Precisely, we let $\gamma_i(t) = \sum_{j=1}^{J_{\gamma}} k_{t_{\gamma}^j}(t - t_{\gamma}^j; \theta_{t_{\gamma}^j}) \gamma^*_i(t_{\gamma}^j)$, where $\gamma^*_i(t_{\gamma}^j)$ are independent, normally distributed, and centered on a common $\gamma^*(t_{\gamma}^j)$. We use Gaussian kernels for $k_{t_{\gamma}^j}(\cdot; \theta_{t_{\gamma}^j})$ with bandwidths $\theta_{t_{\gamma}^j}$ (standard deviation of the Gaussian pdf) varying over wavelength. We assume that the log-bandwidths follow a multivariate normal distribution centered at a global log-bandwidth and $\text{Cov}[\log(\theta_{t_{\gamma}^j}), \log(\theta_{t_{\gamma}^j}')] = \nu^2 \exp(-|t_{\gamma}^j - t_{\gamma}^j'|/\phi_{\gamma})$, yielding a non-stationary process because of the heterogeneous bandwidth. (Due to identifiability challenges, we fix $\phi_{\gamma} = 1/50$ and define the correlation matrix for the log bandwidths as $R_{\gamma,\gamma}$.) We found that this nonstationary specification outperformed a full-rank stationary GP with squared-exponential covariance (See Supplemental Material).

We also specify $\beta(t)$ using kernel convolutions, where $\beta(t) = BK_{\beta}(t)$. With $p$ covariates, $B$ supplies a $p \times q$ matrix representation of the $p$ regression coefficient functions $\beta(t)$. Here, the kernel convolution has knots every 25 nm from 437.5 - 962.5 nm. As with $\gamma_i(t)$, we use Gaussian kernels to specify $K_{\beta}(t)$; however, unlike $\gamma_i(t)$, we assume common bandwidths for all kernels, for all wavelengths, and for each coefficient function. In our sensitivity analysis, we found that including fewer (every 50 nm) or more knots (every 20 nm) for $\beta(t)$ led to a bigger DIC.

Turning to $\eta(s, t)$, we use wavelength kernel convolutions of spatially-varying variables. That is, we consider low-rank but heterogeneous and nonstationary (in the wavelength domain) specifications. We select a set of wavelength knots $t_{\eta}^1, ..., t_{\eta}^J$, spaced every 25 nm from 437.5-962.5 nm (22, in total). In our sensitivity analysis, we found that including fewer knots led to a significantly worse fit, while including more knots did not lead to significantly better fit. We define the space-wavelength function as

$$\eta(s, t) = K(t)^T z(s) = \sum_{j=1}^{J_{\eta}} k_{t_{\eta}^j}(t - t_{\eta}^j; \theta_{t_{\eta}^j}) z_{t_{\eta}^j}(s), \tag{2}$$

where $z_{t_{\eta}^j}(s)$ are spatially-varying random variables associated with Gaussian wavelength kernels $k_{t_{\eta}^j}(\cdot; \theta_{t_{\eta}^j})$. Unlike the kernel structure for $\gamma_i(t)$, we use a com-
mon bandwidth \( \theta^{(\eta)} \) for all knots. We use the Gaussian kernel and common bandwidth because alternative kernel and wavelength-dependent bandwidth choices did not improve DIC. The construction in (2) allows heterogeneity and nonstationarity in wavelength space, where the heterogeneity is introduced through \( z_{y_j}(s) \) (see White, Keeler and Rupper, 2021, for a similar construction for spatial monotone regression). Nonstationarity can be obtained using through other basis functions, including splines or wavelets (Ray and Mallick, 2006; Yu et al., 2020).

We briefly remark on choosing the form \( \eta(s,t) = K(t)^T z(s) \) vs. \( \eta(s,t) = K(s)^T z(t) \). With \( n \) sites, the former introduces \( J_\eta n \) random effects, the latter \( 500n \) random effects. With \( J_\eta \) relatively small, the former is preferred computationally. Importantly, it yields better fits to the data (see Supplemental Material).

While we may want dependence between components in \( z(s) \) at \( s \), that dependence should have nothing to do with the \( t^{(\eta)} \)'s. We are capturing association with regard to the distances between wavelength knots through the \( K \)’s and our objective for the \( z \)’s is to obtain perhaps nonseparable and nonstationary covariance structure for \( \eta(s,t) \). So, we write \( z(s) = A w(s) \) where \( A \) is \( J_\eta \times r \) and the components of \( w(s) \) are independent mean 0 GP’s with variance 1 and correlation functions, \( \rho_r(s-s') \).

When \( r = J_\eta \), we have the familiar linear model of coregionalization (Wackernagel, 1998). We consider using \( A = I \) and \( A_{p \times r} \), for various \( r \), as well as a separable specification for \( z(s) \), where, with \( V \) a positive definite matrix, \( \text{Cov}(z(s), z(s')) = \exp(-\phi_w \| s - s' \|) V \). With \( A_{p \times r} \), we fix and order the decay parameters \( \phi_w \), so that each latent GP has a different spatial range (see White and Gelfand, 2020), providing great flexibility to the latent factor representation of \( z(s) \). We compare the various specifications through out-of-sample prediction in Section 4.3.

Under the general form \( \eta(s,t) = K(t)^T A w(s) \),

\[
\text{cov}(\eta(s,t), \eta(s',t')) = K(t)^T A \Sigma_{w(s), w(s')} A^T K(t').
\]

If \( A = I \), we have \( \Sigma_{w(s), w(s')} = D(s-s') \), a \( J_\eta \times J_\eta \) diagonal matrix with entry \( d_{jj} = \rho_j(s-s') \). Thus, \( \text{cov}(\eta(s,t), \eta(s',t')) = K(t)^T D(s-s') K(t') = \sum_j k_j(t-t_j; \theta^{(\eta)}) k_j(t'-t_j; \theta^{(\eta)}) \rho_j(s-s') \). The covariance is always nonseparable and, if \( A \) is unconstrained, it is nonstationary.

As an illustration, we have \( \Sigma_{w(s), w(s')} = \begin{pmatrix} \rho_1(s-s') & 0 \\ 0 & \rho_2(s-s') \end{pmatrix} \) if we take \( A \) to be \( J_\eta \times 2 \). Now, with \( a_1 \) and \( a_2 \) the two columns of \( A \), \( \text{cov}(\eta(s,t), \eta(s',t')) = (K(t)^T a_1)(K(t')^T a_1) \rho_1(s-s') + (K(t)^T a_2)(K(t')^T a_2) \rho_2(s-s') \). We achieve both dimension reduction and space-wavelength interaction. Further, we have nonseparability and nonstationarity (in the wavelengths) if there are different bandwidths for the different \( t_j \). If we set \( r = 1 \), we have separability but still nonstationarity in the wavelengths.
Turning to the variance, we specify $\log(\sigma^2(t))$ as a piecewise linear spline with knots every 20 nm from 440 - 960 nm, such that $\log(\sigma^2(t)) = K_\sigma(t)^T \mathbf{\beta}_\sigma$. In our sensitivity analysis presented in the supplemental material, we find that including fewer knots (spaced every 25 nm or more) yields a worse fit, while including more knots (spaced every 10 nm) provides no additional improvement.

4.3. Model Comparison. We focus our model comparison on Asteraceae, the most abundant family, using 10-fold cross-validation but briefly discuss results for Aizoaceae and Asteraceae. In the Supplemental Material, we present cross-validation results examining various specifications of the spatial process in $\eta(s,t)$. When comparing models with different specifications of $\eta(s,t)$, all models include spatially-varying genus-specific intercepts $\delta_i + \alpha_i(s)$, a global (not genus-specific) wavelength random effect $\gamma(t)$, and functional regression coefficients $\mathbf{\beta}(t)$. For $\eta(s,t)$, we compare separable, independent, and latent factor models. We find that the latent factor specification of $\eta(s,t)$ with $r = 10$ has the best out-of-sample predictive performance and use this for $\eta(s,t)$ in the remainder of the manuscript. For this specification of $\eta(s,t)$, we focus our model comparison on eight special cases of (1) arising by (i) including or excluding $\alpha_i(s)$, (ii) using $\gamma_i(t)$ or only $\gamma(t)$, and (iii) having functional coefficients $\mathbf{\beta}(t)$ or scalar coefficients $\beta$.

We hold out reflectances imagining the setting where researchers visited a site but failed to measure reflectances for some genus at that site. So, at random, we leave out spectra that have (i) at least one other observed reflectance spectrum at the same site and (ii) at least one other observed spectrum of the same genus located elsewhere. For Asteraceae, this yields 117 candidates out of the 185 in total. Holding out a subset, we fit the model using Markov chain Monte Carlo, and, with each posterior sample, we predict the hold-out reflectance spectra. We compare models by averaging across the wavelengths to obtain the predicted mean squared error (MSE), mean absolute error (MAE), and the mean continuous ranked probability score (MCRPS), see Gneiting and Raftery (2007). The results are summarized in Tables 1 and in the Supplement.

Following the results in Table 1 and the Supplemental Material, we adopt a model with (1) a global wavelength random effect, (2) a spatially-varying genus-specific intercept, (3) functional regression coefficients, and (4) a space-wavelength random effect specified through the wavelength kernel convolution of a multivariate spatial process with 10 latent spatial GPs having different decay parameters. We use this model to analyze the GCFR data for the Asteraceae family. As discussed in Section 4.2, several of the model specifications (e.g., knot spacing) are discussed in more detail in the sensitivity analysis in the Supplemental Material.
β using As with the Asteraceae family, there is little separation between the models 
and the random effect specification. We focus our comparison between families on how the estimated relationships 
we assume the following prior distributions:

\[
\alpha_i(s) \sim \mathcal{GP} \left(0, \sigma^2_{\alpha(s)}e^{-\phi_{\alpha}d(s,s')} \right) \\
\gamma_i(t) \sim \mathcal{N} \left(0, \sigma^2_{\gamma(t)} \right) \\
w_j(s) \sim \mathcal{GP} \left(0, e^{-\phi_{w_j}d(s,s')} \right) \\
\sigma^2_{\delta} \sim \mathcal{IG} \left(3, 0.2 \right) \\
\sigma^2_{\alpha(s)} \sim \mathcal{IG} \left(3, 2 \right) \\
\sigma^2_{\beta(t)} \sim \mathcal{IG} \left(3, 2 \right) \\
\phi_{\alpha} \sim \text{Unif} \left(1/100, 1 \right) \\
\theta^{(\beta)} \sim \text{Gamma} \left(5, 1/10 \right) \\
A_{ik} \sim \mathcal{N} \left(0, \sigma^2_{\gamma} \right) \\
\phi_{\gamma} \sim \text{Unif} \left(1/100, 1 \right) \\
\theta^{(\gamma)} \sim \text{Gamma} \left(5, 1/10 \right) \\
\beta_{\sigma} \sim \mathcal{N} \left(0, 10^2 \right) \\
\sigma^2_{\gamma} \sim \mathcal{IG} \left(3, 3^2 \right) \\
\sigma^2_{\alpha(s)} \sim \mathcal{IG} \left(3, 2 \right) \\
\sigma^2_{\beta(t)} \sim \mathcal{IG} \left(5, 2 \right)
\]

None $\gamma(t)$ $\beta(t)$ $\alpha_i(s)$ $\gamma_i(t)$ $\beta(t)$ $\alpha_i(s)$ $\gamma_i(t)$ $\beta(t)$ $\alpha_i(s)$ $\gamma_i(t)$ $\beta(t)$

<table>
<thead>
<tr>
<th>$\alpha_i(s)$</th>
<th>$\gamma_i(t)$</th>
<th>$\beta(t)$</th>
<th>MSE</th>
<th>MAE</th>
<th>MCRPS</th>
<th>Relative MCRPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>$\gamma(t)$</td>
<td>$\beta(t)$</td>
<td>0.148</td>
<td>0.301</td>
<td>0.241</td>
<td>1.229</td>
</tr>
<tr>
<td>None</td>
<td>$\gamma(t)$</td>
<td>$\beta(t)$</td>
<td>0.141</td>
<td>0.293</td>
<td>0.234</td>
<td>1.196</td>
</tr>
<tr>
<td>None</td>
<td>$\gamma_i(t)$</td>
<td>$\beta(t)$</td>
<td>0.175</td>
<td>0.320</td>
<td>0.265</td>
<td>1.355</td>
</tr>
<tr>
<td>None</td>
<td>$\gamma_i(t)$</td>
<td>$\beta(t)$</td>
<td>0.169</td>
<td>0.318</td>
<td>0.262</td>
<td>1.340</td>
</tr>
<tr>
<td>$\alpha_i(s)$</td>
<td>$\gamma_i(t)$</td>
<td>$\beta(t)$</td>
<td>0.102</td>
<td>0.244</td>
<td>0.200</td>
<td>1.023</td>
</tr>
<tr>
<td>$\alpha_i(s)$</td>
<td>$\gamma_i(t)$</td>
<td>$\beta(t)$</td>
<td>0.097</td>
<td>0.237</td>
<td>0.196</td>
<td>1.000</td>
</tr>
<tr>
<td>$\alpha_i(s)$</td>
<td>$\gamma_i(t)$</td>
<td>$\beta(t)$</td>
<td>0.348</td>
<td>0.435</td>
<td>0.393</td>
<td>2.009</td>
</tr>
<tr>
<td>$\alpha_i(s)$</td>
<td>$\gamma_i(t)$</td>
<td>$\beta(t)$</td>
<td>0.290</td>
<td>0.420</td>
<td>0.380</td>
<td>1.940</td>
</tr>
</tbody>
</table>

Table 1 Out-of-sample predictive performance model comparison for Asteraceae family. Models vary by including or excluding genus-specific terms, as well as comparing scalar and functional coefficients. All models use $r = 10$ spatial factors to construct $\eta(s,t)$.

We also carry out cross-validation on the Aizoaceae and Restionaceae families. As with the Asteraceae family, there is little separation between the models using $\beta$ and $\beta(t)$. However, because of its inferential value, we include $\beta(t)$ for all families. The Restionaceae data prefers the same specification as Asteraceae with $\alpha_i(s)$ and $\gamma_i(t)$, while the Aizoaceae data is best including $\gamma_i(t)$ and excluding $\alpha_i(s)$. In Section 5.1, where we compare the results for the three plant families, we focus our comparison between families on how the estimated relationships between environmental variables and reflectances captured through $\beta(t)$ differ between families. As we discuss in Section 4.5, the results of $\beta(t)$ are robust to the random effect specification.

4.4. Prior Distributions, Model Fitting, and Prediction. We present the relevant prior distribution for the model with the best out-of-sample predictive performance: $Y_{ij}(s,t) = \delta_i + \alpha_i(s) + X(s)\beta(t) + \gamma(t) + \eta(s,t) + \epsilon_{ij}(s,t)$. For this model, we assume the following prior distributions:

\[
\delta_i \sim \mathcal{N} \left(\mu_\delta, \sigma^2_\delta \right) \\
\mu_\delta \sim \mathcal{N} \left(0, 10^2 \right) \\
\alpha_i(s) \sim \mathcal{GP} \left(0, \sigma^2_{\alpha(s)}e^{-\phi_\alpha d(s,s')} \right) \\
\gamma_i(t) \sim \mathcal{N} \left(0, \sigma^2_{\gamma(t)} \right) \\
w_j(s) \sim \mathcal{GP} \left(0, e^{-\phi_{w_j}d(s,s')} \right) \\
\sigma^2_{\delta} \sim \mathcal{IG} \left(3, 0.2 \right) \\
\sigma^2_{\alpha(s)} \sim \mathcal{IG} \left(3, 2 \right) \\
\sigma^2_{\beta(t)} \sim \mathcal{IG} \left(3, 2 \right) \\
\phi_{\alpha} \sim \text{Unif} \left(1/100, 1 \right) \\
\theta^{(\beta)} \sim \text{Gamma} \left(5, 1/10 \right) \\
A_{ik} \sim \mathcal{N} \left(0, \sigma^2_{\gamma} \right) \\
\phi_{\gamma} \sim \text{Unif} \left(1/100, 1 \right) \\
\theta^{(\gamma)} \sim \text{Gamma} \left(5, 1/10 \right) \\
\beta_{\sigma} \sim \mathcal{N} \left(0, 10^2 \right) \\
\sigma^2_{\gamma} \sim \mathcal{IG} \left(3, 3^2 \right) \\
\sigma^2_{\alpha(s)} \sim \mathcal{IG} \left(3, 2 \right) \\
\sigma^2_{\beta(t)} \sim \mathcal{IG} \left(11, 10 \right) \\
\sigma^2_{\phi_\gamma} \sim \mathcal{IG} \left(5, 2 \right)
\]
In general, we use weakly informative prior distributions. For variance parameters, we use finite variance inverse gamma prior distributions that, given the scale of our data, are weakly informative. We assume weakly informative prior distributions for $\text{vec}(B)$, $\gamma^*$, and $\beta_s$ that provide flexibility to the low-rank representations of the regression coefficients, the global wavelength random effect, and piecewise linear log-variance function, respectively. For the decay parameter $\phi_\alpha$, we use a uniform prior distributions that allows effective spatial ranges between 3 and 300 km. Given that fewer than 0.3% of locations are separated by 300 or more km, this is quite flexible. The prior distributions of bandwidths $\theta^{(B)}$ and $\theta^{(n)}$ have prior means of 50 nm but have high variance to allow very flexible functional forms.

We fit our model using Markov chain Monte Carlo (MCMC) through a Gibbs sampler when closed-form posterior conditional distribution are available and using Metropolis-Hastings within Gibbs when these conditional distributions are not available in closed form. These distributions are presented in the Supplemental Material. We sample from the posterior distribution 200,000 times. We discard the first 150,000 samples and thin the remaining 50,000 samples to 10,000 posterior samples. We base our inference on these 10,000 posterior samples.

After model fitting, when of interest, we sample from the posterior predictive distribution,

$$\int_{\psi} f(y_{\text{new}} | \psi) \pi(\psi | Y) d\psi,$$

using composition sampling (see Tanner, 1996, for early reference), where $\psi$ denotes all model parameters and $Y$ represents all data. When interpolating reflectances to a new site, this involves sampling spatially-varying intercepts $\alpha_i(s)$ and latent GP parameters $w_j(s)$ from the appropriate conditional normal distribution conditional on sampled values.

4.5. **Confounding and Orthogonalization.** The flexibility of the residual specification in our best performing model results in annihilation of the significance of the spatial regressors. This is a well-documented problem in the literature (see, e.g., Hodges and Reich, 2010; Khan and Calder, 2020). A solution in the literature is orthogonalization; that is, projection of the random effects (the spatial residuals) onto the orthogonal complement of the manifold spanned by the spatial covariates. This yields revised regression coefficients with direct interpretation in the presence of the random effects. The coefficients are more aligned with those that arise from model fitting ignoring spatial random effects.

We propose a similar orthogonalization approach here but our setting is more demanding because we have both space and wavelengths in our residuals. We have to introduce orthogonalization with regard to the manifold spanned by the spatial covariates as well as with regard to the manifold spanned through the use of kernel functions with knots. We present the details below for the simpler case where we
have no replicates at locations. However, in our application, we have replicates associated with the spatial locations and also with different genera. So, formally, the orthogonalization requires us to introduce a location by genus matrix to align the number of observed sites with the number of observed reflectances. We present the more detailed argument in the Supplemental Material.

With \( n \) sites and 500 wavelengths, we can express (1) in matrix form as

\[
Y = \mu_\delta \mathbf{1} + XBK_\beta^T + \eta^* + \epsilon
\]

where \( Y \) is the \( n \times 500 \) matrix of log-reflectance spectra data by sites, \( \mathbf{1} \) is an \( n \times 500 \) matrix of ones, \( \mu_\delta \) is the global mean, \( X \) is the \( n \times p \) spatial design matrix (with \( p \) covariates), \( B \) is \( p \times J_\beta \) with \( J_\beta \) knots, \( K_\beta \) is the \( 500 \times J_\beta \) kernel design matrix with \( J_\beta \) knots. \( \eta^* \) is also \( n \times 500 \) summing the corresponding matrix forms for the mean-zero random effects \( (\gamma(t), \alpha_i(s), \delta_i - \mu_\delta, \text{and } \eta(s,t)) \). Then, using standard results, we can vectorize (4) to

\[
\text{vec}(Y) = \mu_\delta \mathbf{1} + \underbrace{(X \otimes K_\beta)\text{vec}(B)} + \underbrace{\text{vec}(\eta^*)} + \underbrace{\text{vec}(\epsilon)}
\]

where \( \text{vec}(Y) \) is an \( n \times 500 \) vector with \( X \otimes K_\beta \) an \( n \times 500 \times pJ \) matrix.

Now, define the joint projection matrix,

\[
P_{XK_\beta} \equiv (X \otimes K_\beta)((X \otimes K_\beta)^T(X \otimes K_\beta))^{-1}(X \otimes K_\beta)^T = (X(X^TX)^{-1}X^T) \otimes (K_\beta(K_\beta^TK_\beta)^{-1}K_\beta^T) = PX \otimes PK_\beta,
\]

and write \( \text{vec}(\eta^*) = P_{XK_\beta} \text{vec}(\eta^*) + (I - P_{XK_\beta}) \text{vec}(\eta^*) \). Then, we can write

\[
\text{vec}(Y) = (X \otimes K_\beta)\text{vec}(B^*) + (I - P_{XK_\beta}) \text{vec}(\eta^*) + \text{vec}(\epsilon),
\]

where the updated unconfounded coefficients are (in vec and block form)

\[
\text{vec}(B^*) = \text{vec}(B) + ((X \otimes K_\beta)^T(X \otimes K_\beta))^{-1}(X \otimes K_\beta)^T \text{vec}(\eta^*),
B^* = B + (X^TX)^{-1}X^T\eta^* K_\beta (K_\beta^TK_\beta)^{-1}.
\]

Here, \( \text{vec}(B^*) \) and \( B^* \) provide the vector and matrix of regression coefficients, respectively, under the orthogonalization. The model is fitted using (1). Then, with the posterior samples of the \( \beta \)'s, \( \gamma \)'s, \( \eta \)'s, and intercept terms \( (\alpha_i(s), \delta_i, \mu_\delta) \), along with the \( X(s) \) and \( K_\beta(t) \), the unconfounded \( B^* \)'s can be obtained using (7).

5. Analysis of the GCFR Reflectance Spectra Data. We focus discussion on a comparison between families but give specific attention to the results on Asteraceae, the most abundant family. We compare and discuss results from the orthogonalized coefficients using the approach in Section 4.5. In addition, we summarize covariate importance on log-reflectance. Again using the orthogonalized random effects and unconfounded regression functions, we discuss the proportion of variance explained by each model term.
The confounding between random effects (genus, wavelength, and spatial) and covariates pushes $\beta(t)$ to zero, obliterating any significant inference with regard to the effect of environmental variables on log-reflectance. For each MCMC posterior sample, we calculate the proportion of the variance in each random effect ($\alpha_i(s), \gamma(t), \eta(s, t)$) explained by $X$ and $K_\beta$. We orthogonalize our random effects with respect to $X$ and $K_\beta$ as described in Section 4.5 to remove the diminishing of the effect of the regressors.

For the Asteraceae family, we explore the proportion of the variance explained by each of the mean-zero model terms. For every posterior sample, we calculate the empirical variance of all nonorthogonalized and orthogonalized terms (See Figure 5 to the 95% credible regions): $\epsilon_{ij}(s, t)$, $X(s)^T \beta(t)$, $\alpha_i(s) + \delta_i - \mu_\delta$, $\gamma(t)$, and $\eta(s, t)$. We take $\alpha_i(s) + \delta_i$ to capture both genus-specific terms and subtract $\mu_\delta$ to make $\alpha_i(s) + \delta_i - \mu_\delta$ a mean-zero random effect. For orthogonalized terms, $\gamma(t)$ explains slightly under 25% of the variability of the data, while both $\eta(s, t)$ and $X(s)^T \beta(t)$ explain over 30% of the total variance. Without orthogonalization of the random effects, the environmental regression explains almost no variance. The genus-specific spatially-varying intercept $\alpha_i(s) + \delta_i - \mu_\delta$ explains over 10% of the total variance while $\epsilon_{ij}(s, t)$ accounts for about 5% of variance in the data.

In Figure 5, we plot the proportion of between-spectrum variability explained by all orthogonalized mean-zero terms as a function of wavelength (posterior mean and 95% credible interval). Even though $\gamma(t)$ is common to all spectra, after orthogonalization, it is no longer a constant term for all spectra. For wavelengths less than 700 nm, we find that unconfounded environmental regression and space-wavelength random effects are most important in explaining between-spectrum variance. For higher wavelengths (> 750 nm), where there is little variation in the wavelength functions; the orthogonalized global wavelength random effects $\gamma(t)$ and the unconfounded environmental regression explain the most between-spectrum variance. The spatially-varying genus-specific offset, $\alpha_i(s) + \delta_i - \mu_\delta$ explains between 10-20% of between-spectrum variance for most wavelengths but appears particularly influential for wavelengths between (675-725 nm). The $\epsilon_{ij}(s, t)$ account for the 0 to 10% of unexplained between-spectrum variance in log-reflectance, depending on wavelength.

After updating $\beta(t)$ in the presence of orthogonalization, we present inference on covariates for all families. Figure 6 shows the posterior mean, 95% credible interval for each element of $\beta(t)$. All coefficient functions are significantly non-zero for most wavelengths (around 99% for all wavelengths). With covariates centered and scaled, (i) we can interpret effects as the expected change in log-reflectance for a one standard deviation change in the covariate, holding the other covariates constant, and (ii) we can compare the scales of the coefficient functions among covariates.

The four covariates have positive effects for some wavelengths, negative effects
for others, with a transition around 700 nm, a threshold/boundary between visible (450-700nm) and near-infrared regions (NIR, 700-1400nm) of the spectrum. The visible region is most strongly affected by differences in plant pigment composition/concentration while the NIR is most affected by structural properties related to the cell wall, to air interface within the leaf (Jacquemoud and Ustin, 2019c). Traits can exhibit uniform effects across multiple parts of the spectrum (e.g., often in water content) or can cause increased reflectance in parts of the spectrum and decreased reflectance in others (Feng et al., 2008; Jacquemoud and Ustin, 2019b). Different sets of traits acting in concert in response to environment likely drive the positive and negative shifts across the 700 nm threshold in Figure 6.

For Asteraceae, we estimate that higher elevations are associated with lower reflectance levels at wavelengths less than 700 nm but higher reflectance at wavelengths above 700 nm. The relationships of precipitation and temperature with reflectance are similar. On the other hand, rainfall concentration is positively correlated with reflectance at low wavelengths and becomes negatively correlated with reflectance as wavelength increases. We note that rainfall concentration, the environmental feature that reflectance responds differently to, is largely longitudinally driven in comparison to the other features. Specifically, the extreme western and to some extent the extreme eastern sample sites have significantly higher rainfall concentrations than more central locations. Because there is between-covariate correlation, the coefficient functions must be interpreted as partial slopes, i.e., holding all other covariates constant.

To compare covariate importance, we calculate the mean integrated absolute coefficient over the wavelength domain, $|\beta_j| = \frac{1}{500} \int_{450}^{950} |\beta_j(t)|dt \approx \frac{1}{500} \sum_{i=1}^{500} |\beta_j(t_i)|$, for each covariate. This metric weights the contribution of the coefficient equally regardless of sign or wavelength. We calculate $|\beta_j|$ for every posterior sample and plot these in Figure 7. In terms of $|\beta_j|$, elevation and temperature are more influential on reflectance than precipitation and rainfall concentration.

Lastly, to illustrate the utility of our model, we examine visualize out-of-sample
predictive reflectance. Specifically, we hold out six reflectance curves over the spatial domain. The locations of the hold out curves are shown in Figure 12 of the Supplementary Material. In Figure 13, the hold-out data, the posterior mean log-reflectances, and 95% simultaneous credible intervals are shown. Discussion is presented in the Supplement but, to summarize, the posterior predictions at holdout sites show good agreement with the held-out curves.

5.1. Comparison across families. We compare the regression coefficient functions for the three families in this study (posterior mean and 95% credible interval): Aizoaceae, Asteraceae, and Restionaceae (See Figure 6). The regression coefficient functions are clearly distinct across the families. However, between-covariate correlation or different spatial sites covered by each family may account for some of these differences.

The estimated effects of elevation, annual precipitation, and temperature are opposite in direction for all wavelengths between Asteraceae and Aizoaceae. For these covariates, we see positive effects on Aizoaceae log-Reflectance for wavelengths < 700 nm and negative effects for wavelengths > 700 nm, with opposite patterns for Asteraceae. For Asteraceae, the estimated effects of rainfall concentration are positive for lower wavelengths and negative for higher wavelengths, while they are nearly zero for Aizoaceae. Restionaceae has very small estimated temperature effects. For elevation and rainfall concentration, Restionaceae shows significant effects on log-reflectance for wavelengths < 700 nm, but essentially no effect for higher wavelengths. The estimated effect of precipitation for Restionaceae is similar to Aizoaceae in pattern but is smaller in magnitude.

In Figure 6, we also plot the variance function for $\epsilon_{ij}(s, t)$ for each family (posterior mean and 95% credible interval). Asteraceae has the highest estimated variance for most low wavelengths (450 - 700 nm), a trend that matches the between spectrum variance patterns in Figure 3. Restionaceae has the lowest estimated variance for (450 - 700 nm). All families have very low estimated variance for most high wavelengths (700 - 950 nm).

The differing responses in visible and near-infrared reflectance to environment between Aizoaceae and Asteraceae likely indicate that genera within the two families employ different adaptive strategies in response to their local environments across the landscape. The Aizoaceae family consists of small succulent stemmed and leaved plants while the Asteraceae family largely consists of non-succulent leaved herbs and shrubs. Both plant families adapt via other traits tied to aridity tolerance (e.g., water storage for periods of drought) and avoidance (e.g., leaf hairs, wax, and anthocyanin pigmentation that block UV radiation). The adaptive traits in the respective “evolutionary toolboxes” of Aizoaceae and Asteraceae are constrained by their phylogenetic ancestry, resulting in differing strategic responses to environment in their traits and thus, reflectances. In contrast, the Restionaceae consist of grass-like plants with tough fibrous pho-
tosynthetic stems that vary less than the other two families in adaptation to drought.

![Graphs showing the effect of elevation, rainfall concentration, and annual precipitation](image)

**Fig 6.** Between family comparison of (Top-Left to Middle-Right) environmental regression coefficient functions (Elevation, Annual Precipitation, Rainfall Concentration, and Average Minimum January Temperature) and (Bottom) wavelength-varying variance $\sigma^2(t)$.

We show the posterior distribution (box plots) for $|\beta_j|$ across all covariates and families (See Figure 7). Since $|\beta_j|$ represents the relative importance of covariates for log-reflectance, we see that the covariates are more important in describing log-reflectances for Asteraceae than Aizoaceae and more important for Aizoaceae than for Restionaceae, except for rainfall concentration. Perhaps the relative importance $|\beta_j|$ may be higher for Aizoaceae and Asteraceae because these have more expansive spatial distributions and thus experience higher variability in environmental variables.

Despite the differences in spatial ranges, the families differ in terms of which environmental variables have the highest relative importance to their reflectance signals. The most important variable for Asteraceae is elevation, likely a proxy
for several environmental factors; prominent among them is the biome shift from the higher elevation Fynbos biome within the Cederberg mountains to the lower elevated Succulent Karoo biome. These biomes differ widely in their environments, the Fynbos biome having nutrient-poor soils and a regular fire cycle while the Succulent Karoo is largely arid with low levels of rainfall. Asteraceae is the only family to fully span both biomes in large numbers and these biomes feature a wide difference in environments. The most important variable for Aizoaceae is the minimum average temperate in January (the peak austral summer month), a strong indicator of the maximum temperature a plant can tolerate. This suggests that the major driver of Aizoaceae reflectances are underlying adaptations related to heat tolerance/avoidance. While more limited in its spatial extent, the Restionaceae reflectance spectra responded most to rainfall concentration. Under the notion that higher concentrations of rainfall in fewer months out of the year would lead to more dramatic periods without water, much of the differences in Restionaceae reflectance may be in response to underlying traits managing water during times of drought.

![Figure 7](image_url)

**Figure 7.** Covariate importance $|\beta|$ for all families.

6. **Summary and Future Work.** We have offered plant reflectance modeling to capture variation over space between reflectance across genera within a family. We incorporate wavelength heterogeneity, spatial dependence, and also wavelength - covariate interaction as well as space - wavelength interaction. We have fitted these models to reflectances from the Greater Cape Floristic Region in South Africa, demonstrating successful model performance and revealing a range of novel inference as well as successful spatial prediction.

This work has several future applications and opportunities for further development. Our current data only included the visible and near-infrared reflectance spectra of leaves. These data could be expanded to include the reflectance of plant canopies across a broader spectral range to make predictions relevant to the re-
flectance spectra collected by broader band sensors aboard aerial and satellite remote sensing platforms. Our spatially explicit predictions of plant reflectance would be highly relevant for spectral unmixing analyses which seek to predict the abundances of spectral end members, i.e., individual species, in a canopy of vegetation. Future modeling efforts include exploring reflectance signatures following evolutionary history, explicitly taking into account phylogeny among different groups of plants.

Our space-wavelength model could also be adapted for space-time applications. For suitable spatiotemporal settings, it may be useful to construct spatial kernel convolutions of wavelength/temporal GPs. Also, our approach to spatial orthogonalization for functional regression coefficients could be applied to dynamic regression in spatiotemporal settings.

Acknowledgement. We thank Matthew Aiello-Lammens, Douglas Euston-Brown, Hayley Kilroy Mollmann, Cory Merow, Jasper Slingsby, Helga van der Merwe, and Adam Wilson for their contributions in the data collection and curation. Special thanks to Cape Nature and the Northern Cape Department of Environment and Nature Conservation for permission to collection leaf spectra and traits. Data collection efforts were made possible by funding from National Science Foundation grant DEB-1046328 to J.A. Silander. Additional support was provided by NASA FINESST grant award 19-EARTH20-0266 to H.A. Frye and J.A. Silander.

Supplementary Material. Extended data analysis, residual analysis, orthogonalization, and results. (LINK ADDED LATER)

References.


Hodges, J. S. and Reich, B. J. (2010). Adding spatially-correlated errors can mess up the fixed effect you love. The American Statistician 64 325–334.


