Joint multivariate and functional modeling for plant traits and reflectances

Philip A. White1,2 · Michael F. Christensen3 · Henry Frye4 · Alan E. Gelfand3 · John A. Silander Jr.4

Received: 29 September 2022 / Revised: 30 June 2023 / Accepted: 2 July 2023 / Published online: 20 July 2023
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract
The investigation of leaf-level traits in response to varying environmental conditions has immense importance for understanding plant ecology. Remote sensing technology enables measurement of the reflectance of plants to make inferences about underlying traits along environmental gradients. While much focus has been placed on understanding how reflectance and traits are related at the leaf-level, the challenge of modelling the dependence of this relationship while accounting for environmental gradients has limited this line of inquiry. Here, we take up the problem of jointly modeling traits and reflectance given environment. Our objective is to assess not only response to environmental regressors but also dependence between trait levels and the reflectance spectrum in the context of this regression. We jointly model the response vector of traits with reflectance, which is a function of wavelength. To conduct this investigation, we employ a dataset from a global biodiversity hotspot, the Greater Cape Floristic Region in South Africa.

Keywords Conditional model validation · Dimension reduction · Functional data · Gaussian process convolution · Markov chain Monte Carlo · Multivariate data

1 Introduction

Plant species are comprised of traits that mediate their interactions with their surroundings and influence their evolutionary fitness. These traits encompass a variety of structural, anatomical, physiological, phenological, and biochemical aspects of a plant. Depending on the ecological scale at which they are measured, plant traits can provide insight into organismal function, plant–environment interactions, species coexistence and community dynamics, ecosystem structure and function, and
biogeography and diversification. Information on the diversity and abundance of plant traits can help us understand and predict complex ecological processes and responses to global change (Reich et al. 1997; Diaz and Cabido 2001; Cadotte et al. 2011). Ecologists are interested in how plant traits vary along spatial-environmental gradients since this can yield insights into the underlying principles of how communities originate and the convergence of survival strategies that plants evolved in response to their environment (Reich et al. 1999; McGill et al. 2006; Anderegg 2023). These relationships provide a way to infer how ecosystems may change under novel environments, an important need as the rate human-driven environmental change increases (Schleuning et al. 2020).

Despite their potential for ecological insight, plant traits are time consuming to collect and global trait databases still only include a portion of the total diversity of plant species in the world (Kattge et al. 2020). One proposed solution to this challenge is to use remote sensing to make inferences about the traits of plants and the communities they comprise (Jetz et al. 2016; Cavender-Bares et al. 2022). Hyperspectral spectroscopy, which measures reflectance values contiguously at small intervals across wavelengths of the electromagnetic spectrum, has been a prominent data source in predicting plant traits from remotely sensed imagery (Asner et al. 2011; Singh et al. 2015; Shiklomanov et al. 2016; Yang et al. 2016; Wang et al. 2020). At the leaf level, various wavelengths can be useful in predicting a suite of chemical, structural, and physiological traits (see Jacquemoud and Ustin 2019a; Kothari et al. 2023, and references therein) and the diversity of leaf-level reflectance within a community has been shown to be correlated with community plant diversity (Schweiger et al. 2018; Frye et al. 2021). Further, leaf-level spectra provide the basis for understanding mechanisms occurring at larger scales observed by remote sensing instruments flown aerially and in space.

There has been a wide set of literature separately exploring the relationships of plant traits versus reflectance spectra and plant traits versus environment. Ecologists are interested in comparing trait and environment relationships between lineages, i.e., groups of species derived from a common ancestor, to infer about the adaptive significance of traits for various plant lineages (Felsenstein 1985). However, the perspective of how plant reflectance signatures vary across lineages and along environmental gradients has only been recently developed (Meireles et al. 2020; White et al. 2022a). Given the viewpoint that plant reflectance spectra are integrative measures of plant traits (Kothari and Schweiger 2022), these efforts are an important step in linking our understanding of reflectance spectra to ecology and evolutionary biology. However, the high dimensionality of spectral data alongside the co-correlation of individual wavelengths and leaf traits poses a methodological hurdle. Little work has sought to understand the relationship between reflectance and traits after the effects of the environment have been considered. Modelling reflectance and traits jointly in response to environmental factors not only yields lineage-specific comparisons of trait and reflectance variation but also provides a correlation of the responses that can be used to infer lineage-specific differences in how traits affect reflectance signals. Most approaches to predict traits from reflectance spectra create separate models based on a plant growth forms, e.g., vines, grass, or shrubs (Wang et al. 2020; Kothari et al. 2023). While a lineage-based modelling approach has less
immediate use for application, it can provide inferences into the mechanisms of reflectance and trait relationships and ultimately help explain unaccounted model error in trait prediction efforts.

We offer novel joint modeling of traits and reflectances given environmental/habitat features. We consider joint modeling of two data types: multivariate continuous traits and a reflectance function obtained at high wavelength resolution. We focus on understanding the effects of environmental regressors on plant traits and the reflectance spectrum at leaf-level, as well as the relationships between traits and reflectance, captured through correlations, under this regression. Modeling such relationships requires a multivariate and functional response to regressors, as well as a model that relates these responses. Traits may, in fact, be ordinal or categorical, e.g., the degree or state of leaf pubescence or waxiness, but consideration of such traits is beyond our scope here.

Conceptually, we can build trait/reflectance models over different taxonomic scales, e.g., family, genus, species. Here, we work at family scale to obtain the largest sample size of trait/reflectance data. This is needed in order to best understand the very large number of correlations of interest, i.e., four traits by 500 wavelength bands, across hundreds of sites, each with individual environmental features. Thus, replicates become all of the observations available for the family in our database.

Multivariate modeling in ecology is well established (see, e.g., Schliep and Hoetting 2013; Clark et al. 2017). In particular, modeling the joint patterns of plant traits improves prediction (see Schliep et al. 2018). Functional data analysis (FDA) is widely used to represent curves/functions with continuous domains (e.g., time or wavelength) (see Ramsay 2005; Ramsay and Silverman 2007, for pioneering work in the field). In general, FDA relies on representing the function through a low-rank representation (e.g., splines, wavelets, or kernels). Our challenge is relating multivariate data response specifications for traits to functional data response specifications for reflectance to allow relational inference between the responses.

Our primary contribution is jointly modeling a trait vector, $T$, and a functional reflectance spectrum, $R$, given a vector of environment/habitat features, $E$. Specifically, our model uses a joint multivariate and function-on-scalar regression, after which we can extract the residual association between $T$ and $R$ given $E$. We prefer a joint specification in the form $[T, R|E]$ to a conditional times marginal specification, $[T|R, E][R|E]$ since the former directly reveals how we capture association in the residuals between traits and reflectances at the replicate level. Specifically, we directly model the correlation between $R$ and $T$ through a joint model for the coefficients of functional bases and trait residuals. Beyond the primary goal of inferring relationships between environment, traits, and reflectance, a further use for this model is prediction of traits or reflectance in the frequent scenario where, at a given site, measurements of either plant traits or reflectances were made but not both. Our model also provides functional heterogeneity and heteroscedasticity.

Turning to model assessment, under our specifications, we are assuming dependence among the traits, dependence across the reflectance spectrum, and dependence between traits and reflectances. So, model comparison should be based upon conditional prediction. We demonstrate improved out-of-sample prediction under the
dependence model vs. an independence model given partial information at a site, i.e., when seeking to predict traits or reflectances which we didn’t collect at the site.

We use our modelling framework to elaborate upon the joint $T$, $R$, and $E$ relationships for four family lineages within the Greater Cape Floristic Region (GCFR) of South Africa. We model data for each plant family separately. The GCFR is of special importance to global biodiversity as it contains two adjacent global biodiversity hotspots, the Fynbos and Succulent Karoo biomes. Such biodiversity hotspots are important to global biodiversity conservation because they are regions that not only contain large numbers of species, but also have many species that are not found anywhere else on Earth (Myers et al. 2000; Latimer et al. 2005; Born et al. 2006). The four families we focus our analysis on- the Aizoaceae, Asteraceae, Proteaceae, and Restionaceae- are iconic families found throughout the GCFR and are comprised of large species radiations (Mucina and Rutherford 2006; Manning and Goldblatt 2012; Manning 2013). We focus on four leaf traits- leaf water content, leaf mass per area, percent nitrogen, and succulence (water content/leaf area)- as these are both commonly used traits in the trait ecology and remote sensing literature and represent major evolutionary strategies among plants (see Wright et al. 2004; Jacquemoud and Ustin 2019b, and references therein). While individual correlations between environmental covariates, traits, and reflectances may be weak (Anderegg 2023), one of the primary benefits of our joint modeling is its ability to account for and synthesize many relationships simultaneously. The methodological advances we present in this paper pave the way for new ecological insights into lineage-based comparisons of leaf trait and reflectance relationships.

We continue the paper with a presentation of the dataset and exploratory data analysis that motivates our analysis in Sect. 2. Based on data characteristics demonstrated in our exploratory data analysis, we present a joint model for multivariate traits and functional reflectance data in 3. We then offer interpretation of our results for each of the four families in Sect. 4 and conclude with a brief summary and potential future work.

2 The dataset and exploratory analysis

Our dataset contains leaf trait and leaf reflectance spectra data collected from species across the Greater Cape Floristic Region. These data were collected in sites located nearby previous floristic inventories with efforts made to sample traits and reflectance for species found within these nearby floristic inventories (see Frye et al. (2021) and references therein).

We model a subset of this plant trait and reflectance data at the family level, fitting each of the families Aizoaceae, Asteraceae, Proteaceae, and Restionaceae individually. These four families are generally speciose in the Greater Cape Floristic Region (GCFR) in South Africa, though the Proteaceae and Restionaceae have less prevalence in the arid regions of the GCFR (Mucina and Rutherford 2006; Manning and Goldblatt 2012; Manning 2013). The Aizoaceae are a family comprised of succulent-leaved shrubs with their diversity centered in the arid Succulent Karoo biome (Mucina and Rutherford 2006). The Asteraceae predominate throughout both
the Fynbos and Succulent Karoo biomes with large variation in form and function across the region. Our data largely focus on the shrub elements within Asteraceae which have leaf adaptions that range from succulent to sclerophyllous, i.e., stiff or leathery, leaves. The Proteaceae are a family of shrubs with broad sclerophyllous leaves and comprise an iconic portion of the Fynbos biome both in terms of biomass and diversity. Lastly, the Restionaceae is an herbaceous graminoid, i.e., grass-like, family that also comprises a dominant component of the Fynbos biome (Mucina and Rutherford 2006).

In Table 1, we provide the number of times each family is observed in the data-set as well as the number of genera and species within each family. We see that the families differ dramatically in how often they appear, as well as how many genera and species appear for that family. Although some plant families have similar spatial ranges, they are not generally co-located (See Fig. 1).

We analyze four continuous leaf traits - leaf fresh water content (FWC), leaf mass per area (LMA, g/cm²), percent Nitrogen (pN), and leaf succulence (LS, g/cm²) along with leaf reflectance as a function over wavelength $\lambda \in [450, 950]$ nanometers, observed as a 500 dimensional reflectance vector, by nanometer. Fresh water content of the leaf is an expression of how much water is in the leaf and is calculated as: $$\frac{\text{leaf wetmass} - \text{leaf drymass}}{\text{leaf drymass}}$$ (Pérez-Harguindeguy et al. 2016). Leaf succulence is a related measure and is calculated as: $$\frac{\text{leaf wetmass} - \text{leaf drymass}}{\text{leaf area}}.$$ Leaf mass per area is a prominent structural leaf trait and is measured as the ratio leaf dry mass to leaf area (Poorter et al. 2009). Lastly, percent nitrogen is the ratio of nitrogen in a leaf compared to the rest of the leaf mass. Nitrogen has a multiple impacts on leaf optical properties since it is related to pigments and proteins. Leaf reflectance was measured from sun leaves collected from the top of the canopy using a USB-4000 Spectrometer (Ocean Optics, Largo, Florida, USA) with a leaf clip attachment. For

Table 1

| Family      | Observations (N) | Genera | Species | Sites |
|-------------|------------------|--------|---------|-------|
| Aizoaceae   | 64               | 17     | 45      | 48    |
| Asteraceae  | 310              | 58     | 197     | 174   |
| Proteaceae  | 110              | 10     | 58      | 71    |
| Restionaceae| 152              | 16     | 89      | 79    |
further details on spectra and trait data collection see Frye et al. (2021) and Aiello-Lammens et al. (2017), respectively. In some cases, at a site, for a given species within a family, we have more than one trait observation and/or reflectance observation. Again, viewing the sites as replicates, such observations are averaged to obtain a single $T$ value and a single $R$ value for the site. However, because we analyze the data at family level, there may be duplication at a site because more than one species is sampled at many sites.

We analyze traits and and reflectance on the log-scale as a standard transformation that improves the assumptions of our Gaussian model (discussed in Sect. 3 below). We define the response as $y_j = \left( T_j', R_j(w) \right)'$, where $T_j'$ is a vector of four plant traits (on the log scale) and $R_j(w)$ is a log-reflectance function. We model $R_j(w)$ as a random function where, again, $R_j(w)$ is observed at 500 wavelengths at one nanometer (nm) spacing between 450-949 nm. For both traits and reflectance, we use $j$ to index replication within family.

![Fig. 2](image)

**Fig. 2** (Top) All log-reflectance curves for each family. Dashed line represents the family-specific mean. Wavelength is given in units of nm. (Bottom) Boxplots for the four plant traits (on the log scale), paneled by each family. The trait are abbreviated as leaf mass per area (LMA) in units of g cm$^{-2}$, fresh water content (FWC) which is unitless, succulence (succulence) in units of g cm$^{-2}$, and percent nitrogen (percent_N) in units of percent.
To visualize the overall patterns present in the data, we plot all log reflectances by family in Fig. 2, including the family-specific mean. All families show some similarities with relatively low reflectance for blue (450 - 500 nm) and some red wavelengths (600 - 675 nm), with a local maximum around 550 nm (green). Reflectance increases in the red (around 700 nm) and remains uniformly high in the near-infrared (740-949 nm). In Fig. 2, we also include box plots of the observed log plant traits for each family. Although traits and reflectances are similar across families, each family shows different amounts of heterogeneity.

To jointly explain plant traits and reflectance, we include four environmental covariates in $E_j$: (i) elevation, (ii) annual precipitation (Precip), (iii) rainfall concentration (RF CONC), and (iv) minimum average temperature in January, the peak of the austral summer (Min Temp). Elevation data was derived from 30 m resolution digital elevation maps (JPL 2020) while the other climate variables were taken from Schulze (1997) which provides data at a resolution of 1 arc minute, or roughly 1.5 km at -33 degrees latitude. We also introduce the family-level plant abundance at the location of the $j$th replicate as an explanatory variable. Though not a customary environmental predictor, we include it in $E_j$ to play the role of a proxy for site level environmental suitability for the family. That is, the other environmental regressors above operate at larger spatial scales and we seek to supply a more local regressor.

As a measure of suitability, we define abundance for a family at a site as the aggregated percent cover of all species in that family at the site. These percent cover values are based on previous species inventories, see Frye et al. (2021). To not disturb the vegetation at the permanent species inventory plots, leaf trait and reflectance spectra were sampled on individuals located in near the original species inventory plots. Due to this misalignment between the sites where trait and spectra data were collected and sites with available percent cover, we estimate percent cover using ordinary kriging on the scale of $\log(x + 1)$ to yield predictions on $[0, \infty)$ as well as to deal with many zeros. Specifically, we use an exponential covariance function with parameters estimated from empirical semivariograms.

To motivate our analysis, we calculate the empirical correlation between the environmental variables and each log trait and log reflectance for each family. We plot these correlations in Fig. 3.

Apart from a few exceptions, most correlations between $E_j$ and $T_j$ are weak, not surprising given only modest correlations between traits and environment within the region (Mitchell et al. 2015; Aiello-Lammens et al. 2017) and at global and other local scales (Wright et al. 2004; Wright and Sutton-Grier 2012). Importantly, there is very little common correlation pattern shared across families, expected given the large differences in growth form and likely ecological strategies that each lineage has evolved. As with traits, examining the empirical correlations between the environmental variables and the log reflectance curves reveals little common pattern between the families. However, importantly, there are evident differences in the correlations between reflectance and environment over the wavelength spectrum, suggesting the need for wavelength-varying coefficients in functional modeling of reflectance as a response.
Because plant families prefer specific environmental conditions, we explore the variability of environmental variables across the sites in Fig. 4. For all families, there is overlap in the environmental predictors; however, there are differences. For example, the Aizoaceae in this dataset occur at sites with lower precipitation than other plant families, while the Restionaceae occur at higher precipitation areas. Interestingly, the Restionaceae in the dataset occur at sites with high abundance of Restionaceae, while Aizoaceae are at sites with lower abundance.

We present the empirical correlations between log traits and log reflectance in Fig. 5.

There are few similarities in the trait-reflectance correlations among the families. In general, the families with more data have weaker correlations between traits and reflectances. Families with more genera and species likely represent speciation
“hot-beds”, e.g., see Verboom et al. (2009), Pirie et al. (2016), and Mitchell et al. (2017), where we would expect that the number of species would result in greater variability of traits and reflectance within lineages. It has been shown that higher within-group trait variation dilutes trait associations (Laughlin et al. 2017; Anderegg et al. 2018) and thus we may anticipate a similar effect for trait and reflectance relationships. Again, our goal is to assess the strength of these trait/reflectance relationships while accounting for environment.

Our exploratory analysis shows that the relationships between traits, reflectance, and environmental predictors (See Figs. 3 and 5) differ greatly across families, both

---

**Fig. 4** (Top) Map of the environmental covariates at observed locations. (Bottom) Box plots for the each covariate by all plant families Aizoaceae, Asteraceae, Proteaceae, and Restionaceae, respectively. To allow simple visualization, we center and scale all covariates (across all families), and the y-axis represents the number of standard deviations that the environmental covariate deviates from the mean. The labels “AI”, “AS”, “P”, and “R” represent Aizoaceae, Asteraceae, Proteaceae, and Restionaceae, respectively.
in shape and magnitude. In preliminary modeling efforts, we found no benefit to modeling the families jointly. In different words, joint modeling of families, e.g., with family level random effects, is not suitable because this implies a “global centering” of the families and borrowing strength across families. In this dataset, trait behavior across families is sufficiently different so that such shrinkage is not appropriate and that global parameters over families are not meaningful. Additionally, different families have very different numbers of genera and species, further complicating joint interpretation. While analysis at a higher taxonomic scale is not appropriate with our data, our approach is applicable at higher scale and would enable potentially richer dependence stories.

3 Model and methods

3.1 The joint specification

Since we model the families individually, we need only subscript the replicates within a family. So, consider vector $T_j$, an $s \times 1$ vector of trait responses for replicate $j$. We assume, after the log-transformation, that this vector can be modeled to follow a multivariate normal distribution. With a binary or categorical response, we would view the corresponding entry in $T_j$ as latent, driving the observed response. We model $T_j$ as

$$T_j = \alpha^{(T)} + B^{(T)}E_j + U_j^{(T)}. \tag{1}$$

Here, $\alpha^{(T)}$ are trait-specific intercepts, $E_j$ is a vector of environmental predictors, say $p \times 1$, for replicate $j$ and $B^{(T)}$ is a $s \times p$ matrix of trait-specific regression coefficients. The $U_j^{(T)}$ are pure errors, i.i.d. $\sim \text{MVN}(0, \Omega^{(T)})$.

We model the reflectance as a functional response variable, observed at 500 wavelengths, with wavelengths denoted by w’s. Specifically,
Here, \( \alpha^{(R)}(w) = K^\prime_w(w)\alpha^{s(R)} \) is a wave-length varying intercept, where dimension reduction is given through \( l \) basis functions in \( K^\prime_w(w) \). \( E_j \) is, as above with the \( p \times 1 \) wavelength specific coefficient vector \( \beta^{(R)}(w) \equiv B^{(R)}K_{\beta}(w) \). That is, we imagine \( K_{\beta}(w) \) as an \( m \times 1 \) vector of basis functions or convolution functions with \( B^{(R)} \) an \( p \times m \) matrix of coefficients providing dimension reduction. Aggregating, \( \beta^{(R)}_{p \times 500} = B^{(R)}K_{\beta}^\prime \) where \( K_{\beta} \) is 500 \( \times \) \( m \).

Further, again using dimension reduction, the term \( K^\prime_U(w)U_j^{(R)} \) introduces the \( q \times 1 \) replicate level vector \( U_j^{(R)} \) of reflectance random effects to supplement the fixed effects contribution. These random effects vectors adopt \( q << 500 \) to provide a dimension reduction for the reflectances. More will be said about the choice of \( q \) below. The \( K_U(w) \) can be collected into a 500 \( \times \) \( q \) matrix, \( K_U \). Finally, the \( \psi_j(w) \) provide independent wavelength specific pure error terms with variances \( \sigma^2(w) \). We model \( \log(\sigma^2(w)) = K_\sigma \gamma_\sigma \) as a linear spline as in White et al. (2022a). The entire reflectance response vector for replicate \( j \) becomes

\[
R_j(w) = \alpha^{(R)}(w) + E_j^\prime \beta^{(R)}(w) + K_U^\prime U_j^{(R)} + \psi_j(w).
\]  

We introduce dependence between the traits and reflectances, the \( T \)'s and the \( R \)'s, through the \( U \)'s. That is, the dependence is at the replicate level. Specifically, we assume

\[
\begin{pmatrix}
U_j^{(T)}
U_j^{(R)}
\end{pmatrix}
\]

is distributed as a mean \( 0 \) multivariate normal with covariance matrix

\[
\Omega = \begin{pmatrix}
\Omega^{(T)}
\Omega^{(TR)}
\Omega^{(R)}
\end{pmatrix}.
\]

As a result the induced covariance matrix for

\[
\begin{pmatrix}
T_j
R_j
\end{pmatrix}
\]

becomes

\[
\Sigma = \begin{pmatrix}
\Omega^{(T)}
K_U \Omega^{(TR)} K_U^\prime + \Psi_j
\end{pmatrix}
\]

where \( \Psi_j \) is the diagonal matrix of pure error reflectance variances. From this matrix we can extract all covariances and correlations. If \( \Omega^{(TR)} \) is a matrix of zeros, then we have an independence model, which we denote as \([T|E]|R|E]\).

We provide details of the dimension reduction used in Sect. 3.2, justify the use of the joint model through cross validation in Sect. 3.4 and Appendix B, and then present the results using the joint model in Sect. 4.

### 3.2 Dimension reduction details

Following White et al. (2022a), we specify the low-rank functional terms wave-length-varying intercept \( \alpha^{(R)}(w) \), random effects \( K^\prime_U(w)U_j^{(R)} \), and \( \beta^{(R)}(w) \) through process convolutions enabling simple connection to Gaussian processes (GPs). That is, the kernels of the process convolution connect the low-rank process to the GP covariance (Higdon 2002). For every basis, we include an intercept so that the wave-length-varying intercepts, regression coefficients, and variances have an overall centering.
We use a rich specification for the wavelength-varying intercept $\alpha^{(R)}(w)$, employing Gaussian kernels with wavelength knots spaced every 10 nm from 450-950 nm ($N_w = 52$, in total, including the intercept) to obtain $K_\alpha(w)$. To specify the wavelength-varying random effects through $K_{F_i}(w)$, we use Gaussian kernels with wavelength knots spaced every 25 nm from 450-950 nm ($N_{F_i} = 22$, in total, including an intercept). To specify wavelength-varying coefficient functions, we use Gaussian kernels with wavelength knots spaced every 100 nm from 450-950 nm ($N_w = 7$, including the intercept) for $K_\beta(w)$. We allow the scale parameters for the dimension-reduced coefficient to be unknown. However, we fix the bandwidth of the Gaussian kernels to be 1.5 times the kernel spacing to alleviate well-known lack of identifiability with scale and range parameters of Gaussian process models (see, e.g., Zhang 2004). Lastly, we use a linear spline with interior knots every 50 nm from 475-925 nm to specify $K_\phi(w)$.

For the remaining unknown scale parameters that shrink the low-dimensional functional bases toward zero, we use proper prior distribution for wavelength-varying intercept and regression coefficient functions, we have

$$\alpha^{(R)}_1 \sim N(0, 10^3),$$

$$\alpha^{(R)}_j \sim N(0, \sigma^2_{\alpha}); \quad j = 2, ..., N_w,$$

$$B^{(R)}_{k1} \sim N(0, 10^3); \quad k = 1, ..., p;$$

$$B^{(R)}_{kj} \sim N\left(0, \sigma^2_{\beta_k}\right); \quad k = 1, ..., p; j = 2, ..., N_{\beta},$$

$$\alpha_{(T)}_k \sim N(0, 10^4); \quad k = 1, ..., p,$$

$$B_{jk}^{(T)} \sim N(0, 10^4); \quad j = 1, ..., s; \quad k = 1, ..., p.$$
burn-in period of the model fitting. Specific details of the posterior conditional distributions are provided in Appendix A.

As discussed in Sect. 1, this model can be used to predict traits or reflectance in the frequently occurring scenario where, at a given site, measurements of either plant traits or reflectances were made but not both. In such settings, the residuals/random effects \( U \) would be attached to only partially observed samples. Therefore, predictions for reflectance or traits are made by conditionally predicting \( U, U^{(R)} | U^{(T)} \) or \( U^{(T)} | U^{(R)} \), respectively. Under our model, these predictions rely on conditional normal theory. To illustrate this, we drop the \( j \) subscript and consider prediction of traits or reflectance would, respectively be

\[
\hat{T}_m = \alpha^{(T)}_m + B^{(T)}_m \mathbf{E} + U^{(T)}_m,
\]

\[
\hat{R}(w) = \alpha^{(R)}_m(w) + E' \Omega^{(W)}_m(w) + K'_U(w)U^{(R)}_m + \psi_m(w),
\]

where

\[
U^{(T)}_m \sim N(\mu_{T|R}, \Sigma_{T|R}),
\]

\[
\Sigma_{T|R} = \Omega^{(T)}_m - \Omega^{(TR)}_m \Omega^{(R)^{-1}} \Omega^{(TR)'},
\]

\[
\mu_{T|R} = \Omega^{(R)} \Omega^{(T)^{-1}} U^{(T)}_m,
\]

\[
\Sigma_{R|T} = \Omega^{(R)} - \Omega^{(TR)} \Omega^{(T)^{-1}} \Omega^{(TR)'}.
\]

This process is repeated for all posterior samples \( \eta_1, ..., \eta_M \), yielding a set of \( M \) predictions.

### 3.4 Model comparison

We justify the joint model specification \( [T, R | E] \) by comparing conditional predictive performance using 10-fold cross-validation. Specifically, we compare the joint model to a model where traits and reflectance are independent \( [T | E] [R | E] \). For each fold of the cross validation, we hold out 10% of all traits (jointly), as well as 10% of reflectance spectra (the entire spectrum). The traits and reflectance spectrum are held out exactly one time in the 10-fold cross validation. We adopt this comparison approach to mirror the scenario above where either plant traits or reflectances are measured but not both. Of course, other holdout schemes could be investigated.

We compare models by using predicted root mean squared error (RMSE), mean absolute error (MAE), and the mean energy score (ES),

\[
ES(F, \mathbf{x}) = \frac{1}{2} \mathbb{E}_F || \mathbf{X} - \mathbf{X}' || - \mathbb{E}_F || \mathbf{X} - \mathbf{x} ||,
\]

where \( \mathbf{X} \) and \( \mathbf{X}' \) follow the same distribution \( F \) and \( \mathbf{x} \) is a vector of hold-out values (see Gneiting and Raftery 2007). To estimate this empirically, from a set of posterior predictions \( \mathbf{X}_1, ..., \mathbf{X}_M \), forming \( \hat{F} \) for a vector \( \mathbf{x} \), the energy score is calculated as

\[
ES(\hat{F}, \mathbf{x}) = \frac{1}{2M^2} \sum_{m=1}^{M} \sum_{m'}^{M} || \mathbf{X}_m - \mathbf{X}_{m'} || - \frac{1}{M} \sum_{m=1}^{M} || \mathbf{X}_m - \mathbf{x} ||,
\]

and we average this for all hold-out vectors. The ES quantifies how well a distribution of multivariate predictions matches observed quantities, and smaller values of
the ES indicate better predictions. Moreover, the ES is a proper scoring rule (Gneiting and Raftery 2007), meaning that its expected value is minimized under the correct model. In this study, we use ES to assess our predictions for all traits and all reflectances. We use the ES as our primary model selection criterion when predicting all traits or all reflectances. We present MAE and RMSE for each trait individually but, for simplicity, choose to average these quantities for reflectances over all wavelengths. On the other hand, because either traits or reflectances are predicted jointly conditioning on the other, a single mean ES is given for all traits and for all reflectances, respectively.

4 Results and discussion

4.1 Model comparison results

We focus our model comparison here on the Asteraceae family because it has the most data to help in estimating the proposed correlation structure. The model comparison results are summarized in Table 2, while the comparison results for the other families are in Appendix B.

For the Asteraceae family, the joint model improves out-of-sample prediction performance for reflectances and all traits, relative to the independent model. Overall, the benefit of the joint model is much larger for reflectance than traits, and this benefit is substantial. Based on these findings, we use the joint model to present an interpretation of the results.

Table 2 Model comparison between joint and independent models for Asteraceae. $[T|E][R|E]$ represents the independent model, while $[T,R|E]$ is the joint model. The ranges of values for log leaf mass per area (LMA) in units of g cm$^{-2}$, log fresh water content (FWC) which is unitless, log leaf succulence (LS) in units of g cm$^{-2}$, and log percent nitrogen (percent_N) in units of percent are (-6.01,-2.95), (-1.15,3.51), (-5.33,-0.78), and (-1.24,1.44), respectively. Log reflectances are in units of log percent and range between (0.56,4.42).

| Quantity | Model       | MAE  | RMSE | ES   |
|----------|-------------|------|------|------|
| log LMA  | $[T|E][R|E]$ | 0.449| 0.557| 0.826|
| log FWC  |             | 0.646| 0.814|      |
| log LS   |             | 0.533| 0.706|      |
| log pN   |             | 0.395| 0.495|      |
| log Reflectance | $[T|E][R|E]$ | 0.543| 0.981| 16.396|
| log LMA  | $[T,R|E]$   | 0.398| 0.517| 0.658|
| log FWC  |             | 0.465| 0.612|      |
| log LS   |             | 0.401| 0.546|      |
| log pN   |             | 0.343| 0.457|      |
| log Reflectance | $[T,R|E]$ | 0.166| 0.244| 3.703|
4.2 Correlation between traits and reflectance

We focus our discussion on the estimated correlations between the plant traits (log leaf water content, log leaf mass area, log percent Nitrogen, and log succulence) and the log reflectances. In Fig. 6, we plot the estimated correlations between log traits and reflectance for the Asteraceae, Proteaceae, and Restionaceae families. Specifically, we plot the posterior mean and 90% credible intervals for between trait and reflectance correlation.

Given the environmental variables, we observe relatively weak relationships (typically between 0.2 to −0.2) between traits and reflectance. This finding is not unexpected and matches expectations from previous studies (Laughlin et al. 2017; Pau et al. 2022; Wang et al. 2022), despite the amount of literature devoted to the strength of relationships and prediction potential between traits and reflectance within species and communities (see Jacquemoud and Ustin (2019b) for a review). Most studies focus on differences in these relationships by growth form, e.g., Kothari et al. (2023), with less research examining how these relationships vary across species within broader lineages such as families and genera. The fact that we observe differing $T$ and $R$ relationships between families aligns with our expectation that the species within these families are comprised of various suites of leaf traits that in turn represent different adaptive strategies and ancestral constraints. In other words, spectra represent underlying biology, a point underscored...
in recent literature (Meireles et al. 2020; Cavender-Bares et al. 2022; Kothari and Schweiger 2022).

Further, we again highlight the fact that we are examining $T$ and $R$ relationships jointly given environment. This type of inquiry has often been done indirectly through controlled experimental manipulations (Thenot et al. 2002; Inoue and Peñuelas 2006; Ripullone et al. 2011; Caturegli et al. 2020) or along environmental gradients (Coops et al. 2002; Asner et al. 2009). What our modelling shows is the joint effects that environment has in terms of traits and reflectance. In cases of 0 or insignificant correlation we interpret that either the chosen environmental parameters have little effect on that lineage’s traits (which in turn affect reflectance) or that the species within the lineage differ in their responses to the same environmental parameter resulting in low signal, i.e., the ecological fallacy. In the case of significant correlations, we observe shifts in $T$ and/or $R$ that suggest lineage-wide signals indicative of ecological and evolutionary processes such as adaptation or ancestral constraints.

We observe relatively weak relationships between log reflectance and log leaf water content with one exception. For wavelengths greater than 700 nm (red and near-infrared [NIR]), we estimate that Asteraceae’s reflectance is positively correlated with leaf water content. The near infrared portion of the spectrum is well known in several instances to be a signal for various forms of water content in leaves (Pu et al. 2003; Rodríguez-Pérez et al. 2007; Seelig et al. 2008) and canopies (Peñuelas et al. 1993; Penuelas et al. 1997). On the other hand, Aizoaceae, Proteaceae, and Restionaceae show no or few wavelengths where the 90% credible interval excludes 0. Note that this does not indicate that leaf water content and reflectance are unrelated in these lineages, but rather, after accounting for environment, these correlations are negligible. We suspect that the leaf water signal observed in the Asteraceae is attributable to the fact that the lineage has one of the broadest distributions extending from the arid Succulent Karoo to more mesic Fynbos.

For log leaf mass per area, we generally expected to observe the strongest correlation within the near-infrared region (NIR), i.e., wavelengths greater than 700 nm (Asner et al. 2011; Jacquemoud and Ustin 2019c; Serbin et al. 2019). Within the NIR, some families exhibit positive correlations with reflectance (Restionaceae), while others have negative correlations with reflectance (Asteraceae). Interestingly, Proteaceae has negative correlations for wavelengths between 500 and 725 nm for leaf mass per area. This could be the result of other co-correlated traits such as photosynthetic pigments that more often affect the visible region. There is some limited evidence within species for the negative correlation in the visible region for leaf mass per area (Ourcival et al. 1999).

Nitrogen within leaves is typically linked to the visible region of the spectrum (450-700 nm) given its strong links to the pigment chlorophyll, though there are nitrogen signals found at longer wavelengths (Jacquemoud and Ustin 2019d). For Aizoaceae and Asteraceae, the relationship between log pN and reflectance appears to be weak for all wavelengths, suggesting no lineage wide signals across the environmental range present in the study. Proteaceae shows negative correlations between log pN and reflectance for wavelengths between 500 and 725 nm. Although
weak, the estimated relationship between pN and reflectance is negative for most wavelengths greater than 550 nm for Restionaceae.

Leaf succulence is calculated by dividing the leaf area by leaf water content such that succulence represents the amount of water distributed throughout the leaf. Thus, we expected results like that for leaf water content. Overall, our results matched these expectations, but the results did have different significance compared to water content. Except for the Restionaceae, the other families had significant relationships within the visible range (wavelengths below 700 nm) which was surprising given the typically stronger signal of water within near infrared range. However, previous studies have found relationships between water content and the visible range attributed to the link between plant water status and photosynthetic machinery (Thenot et al. 2002; Inoue and Peñuelas 2006; Ripullone et al. 2011; Hmimina et al. 2014). The Aizoaceae, a lineage dominated by succulent plants, now displayed significant, albeit weak, positive correlations for wavelengths below 700 nm. Asteraceae has weak negative correlations between succulence and log reflectance for wavelengths less than 700 nm but weak positive correlations for wavelengths greater than 700 nm. Proteaceae shows negative correlations between succulence and log reflectance for wavelengths less than 725 nm, but the correlations are essentially 0 otherwise. As with LMA, succulence and reflectance have very weak relationships for Restionaceae.

![Fig. 7](image-url) Posterior mean and 90% credible intervals for regression coefficient functions $\beta_t(w)$ the reflectance spectrum for all families and environmental predictors
We report the estimated effect of the environmental predictors and abundance on reflectance and traits. As discussed in Sect. 2, we have centered and scaled the covariates to aid in the interpretation of the results. In Figs. 7 and 8, we plot the estimated coefficient functions and regression coefficients for each trait, for each family, respectively. Again, abundance is unlike the other environmental covariates in that it is not a direct driver of leaf traits and subsequent reflectances. It is viewed as a proxy for unmeasured local environmental contributing to the “success” of species based on their biomass. In terms of interpretation, we re-emphasize the fact that these models treat traits and reflectance jointly. As shown in Fig. 3, the reflectance spectra within each lineage have clear correlations with environmental parameters. Our results in Fig. 7 are much weaker and less varied for all families, which was expected given the joint nature of the model where more variation is likely to be attributable to leaf traits (Laughlin et al. 2017; Pau et al. 2022; Wang et al. 2022). We interpret Fig. 7 as capturing the signal of family-wide spectral responses to environment that are being
driven by traits that are not currently measured, e.g., pigments, leaf surface features, or other measures of leaf anatomy. We would expect the trait and environment relationships in Fig. 8 to roughly match the initial correlations in Fig. 3 given that it is traits that respond to environment and are subsequently manifested in the reflectance spectra. All covariate effects should be interpreted as the estimated effect of the particular environmental covariate holding all other environmental covariates constant.

### 4.3.1 Aizoaceae

For the Aizoaceae, a family dominated by succulent leaved shrubs, reflectances have a negative relationship with abundance, elevation, annual precipitation and temperature for all wavelengths. In contrast, Aizoaceae’s reflectances have a very weak positive relationship with rainfall concentration. Although the relationships are generally weak, it is the only family whose entire reflectance spectrum is related to all covariates. We suspect that this could be a signal of overall leaf succulence given that precipitation and temperature have similar relationships and the expectation for leaves to be driven towards succulence as an adaptation to drier and hotter environments. Alternatively these family wide responses could be the result of other unmeasured traits such as those in leaf epidermal surfaces (Heim et al. 2015).

In terms of traits, elevation has a significant negative relationship with leaf water content (LWC) and percent nitrogen (pN). Temperature has a positive relationship with leaf mass per area (LMA) but negative relationships with LWC and pN. The LMA and temperature relationship roughly matches expectations found by the Leaf Economic Spectrum (LES), a study that examined trait and environment relationships at a global scale for a large sample of species (Wright et al. 2004). The LES also finds that nitrogen and leaf mass per area tend to be negatively related, fitting in with our results as well.

### 4.3.2 Asteraceae

The Asteraceae are a family of shrubs with a wide regional distribution and have a corresponding diversity of leaf forms. The effects of all environmental covariates are very weak for the entire reflectance spectrum even though there are many significant estimated effects. Specifically, abundance and rainfall concentration are slightly negatively associated with log reflectance less than 700 nm, while annual precipitation has a weak positive relationship with reflectance for all wavelengths over 500 nm. Out of the families chosen, the Asteraceae appear to have some of the weakest environmental signals for reflectance. This may be attributable to the fact that the Asteraceae is one of the most widely distributed groups across the Greater Cape Floristic Region, with a high diversity of leaf adaptations, tolerating a high number of environmental conditions (Mucina and Rutherford 2006; Manning and Goldblatt 2012; Manning 2013). This diversity would likely result in a high variation of reflectance signals that could weaken relationships.

LWC is negatively associated with elevation and annual precipitation, holding other covariates constant, while LMA is positively associated with elevation,
precipitation, temperature, and abundance. The latter results have mixed correspondence to previous global analyses, with global LMA having negative to insignificant relationships with precipitation (Wright et al. 2004). Percent nitrogen is negatively associated with elevation, annual precipitation, and January’s minimum temperature, while it is positively associated with rainfall concentration. Leaf succulence has a weak positive relationship with temperature, while all other covariates have 90% credible intervals that include 0. This partially matches the expectation that leaves would become succulent in more arid areas.

4.3.3 Proteaceae

The Proteaceae are a family of broad-leafed sclerophyllous shrubs. Of the four families, the log reflectance of Proteaceae has the strongest relationship with elevation and annual precipitation. For elevation, reflectance has a strong positive relationship for shorter wavelengths and a weak positive relationship for longer wavelengths. As iterated previously, we interpret the environment and reflectance relationships in Fig. 7 as indicating trends in leaf traits that are unmeasured but varying across the environment. In the case of elevation and the visible region, this may be a trend in traits such as photosynthetic pigments which are strongly associated in the visible region of spectra (Jacquemoud and Ustin 2019c). The estimated relationships between reflectance and annual precipitation is positive and significant for all wavelengths, likely representing changes in traits co-correlated with the amount of water in leaves. Reflectance has a weak negative relationship with temperature for wavelengths greater than 700 nm. Similarly, the relationship between abundance and reflectance is negative for wavelengths less than 750 nm.

In general, Proteaceae’s traits have weak relationships with environmental covariates and abundance. However, LWC reveals a slightly positive relationship with temperature and negative relationship with abundance while LMA has a slightly positive relationship with rainfall concentration. In a study of the genus Protea, a prominent genus of the Proteaceae family within the Greater Cape Floristic Region, Mitchell et al. (2015) found similar results for the LWC and temperature relationship (though theirs were non-significant) and LMA and rainfall seasonality (a related measure to rainfall concentration).

4.3.4 Restionaceae

The Restionaceae was the only grass-like family that we examined. Although Restionaceae showed very weak correlations between traits and reflectance, reflectance is strongly connected with environmental covariates for shorter wavelengths, suggesting a shift in underlying traits associated with the visible region, e.g., photosynthetic pigments, along environmental gradients. Abundance and rainfall concentration are positively related to reflectance at short wavelengths, while both annual precipitation and temperature are negatively related to reflectance. We estimate that increases in temperature are related to decreases in reflectance for all wavelengths. For wavelengths above 525 nm, we find a negative relationship between elevation and reflectance.
Turning to traits, Restionaceae’s LWC is slightly positively related to temperature and annual precipitation but appears to be negatively related to abundance and rainfall concentration. Leaf mass area has slight positive relationships with rainfall concentration and abundance. For log pN and LS, all 90% credible intervals include 0. We suspect that the lack of clear trait and environmental trends could be a result of other known environmental drivers, e.g., fire and soil fertility, that drive differing adaptive strategies within the Restionaceae (Wüst et al. 2016).

5 Summary and future work

For four plant families within the Greater Cape Floristic Region, we have presented modeling to enable assessment of the importance of environmental/habitat predictors in predicting traits and reflectance. This approach allows us to address the novel question of how traits and reflectance vary along environmental gradients. For remote sensing efforts aimed at regional and global extents, this question should be of immediate interest since it is the shifting nature of these relationships across different sets of plant functional types that reduces the generalizability of empirical models for trait prediction (Schimel et al. 2015; Kothari and Schweiger 2022; Wang et al. 2022). Our current model presents an initial step in exploring an area that we feel has been under-utilized in ecology due to a lack of available statistical tools. Lastly, we have shown that joint modeling of traits and reflectances provides better conditional predictive performance than modeling them independently.

In future work, our approaches could be adapted to include discrete or categorical traits, as in Schliep and Hoeting (2013) or Clark et al. (2017). We acknowledge that spatial dependence across sites is anticipated in joint modeling (White et al. 2022a), where, with different intentions, White et al. (2022a) focused on a marginal functional data model in a spatial setting. However, in attempting to obtain the needed large sample sizes for each of the families, we span regions that are disjoint and too spatially distant to employ sensible spatial modeling specifications. Extending the framework in White et al. (2022a) to spatially model the dependence between traits and reflectance would also be of interest, possibly including shape constraints White et al. (2021) or physical models (White et al. 2022b). In addition, with richer datasets, we could explore how reflectance/trait relationships vary along environmental gradients. Overall, the functional data approach in ecology is underutilized despite a plethora of ecological data that would be suitable for such analysis, e.g., spectral reflectance, organismal movement, and time series. We envision models incorporating joint responses of both scalar and functional data will be of high value to ecological problems beyond those in the present study.

Appendix A: Markov Chain Monte Carlo details

In this section, we provide the full conditional distributions used for the Gibb’s sampler. We use \( \theta \mid \cdots \) to denote the full conditional distribution of the parameter \( \theta \). For simplicity, we let \( T \) be an \( n \times s \) matrix of all observed traits and \( R \) be an \( n \times 500 \) log
Table 3 Model comparison between joint and independent models for Restionaceae. 
[|T|E][|R|E] represents the independent model, while [|T, R|E] is the joint model

| Quantity          | Model | MAE | RMSE | ES  |
|-------------------|-------|-----|------|-----|
| log Lma           | [T|E][|R|E] | 0.354 | 0.431 | 0.585 |
| log fwc           | [T|E][|R|E] | 0.297 | 0.384 |       |
| log succulence    | [T|E][|R|E] | 0.443 | 0.580 |       |
| log percent_N     | [T|E][|R|E] | 0.345 | 0.420 |       |
| log Reflectance   | [T|E][|R|E] | 0.587 | 1.079 | 17.925 |
| log Lma           | [T, R|E]  | 0.294 | 0.424 | 0.488 |
| log fwc           | [T, R|E]  | 0.239 | 0.315 |       |
| log succulence    | [T, R|E]  | 0.333 | 0.439 |       |
| log percent_N     | [T, R|E]  | 0.306 | 0.532 |       |
| log Reflectance   | [T, R|E]  | 0.162 | 0.260 | 3.900 |

reflectances. For traits and reflectances, respectively, we use $r_{\theta}^{(T)}$ and $r_{\theta}^{(R)}$ to be the residuals when excluding a the parameter $\theta$ ($\theta$ is used as a placeholder). For example, $r_{\theta}^{(R)}$ is the residuals when removing the environmental regression from the model. In the case of wavelength-varying parameters, we use a similar notation to indicate the exclusion of the first term in a vector (e.g., $\theta_{-1}$). In addition, we let $D_\sigma$ be diagonal matrix with elements of $\sigma^2(w)$ and $D_{\theta_{-1}}$ is a diagonal matrix with the prior variances given in (3). In the case of updates for $U_j$, we refer to terms defined in Sect. 3.3. The posterior conditional distributions are as follows:

\[
\begin{align*}
\mu_{\beta_k} &= (K'_\beta \otimes E') \text{vec}(r_{-\beta,R}) D_\sigma^{-1} \\
\Sigma_{\beta_k} &= \left(K'_\beta D_\sigma^{-1} K_\beta \otimes E'E + D^{-1}_{\beta,R}\right)^{-1} \\
\mu_{\beta_L} &= (I_{s \times s} \otimes E') \text{vec}(r_{-\beta,R}) \Omega(T)^{-1} \\
\Sigma_{\beta_L} &= \left(\Omega(T)^{-1} \otimes E'E + D^{-1}_{\beta,T}\right)^{-1} \\
\mu_{\sigma_k} &= K'_\alpha 1_{-\sigma,R} D_\sigma^{-1} \\
\Sigma_{\sigma_k} &= (nK'_\alpha D_\sigma^{-1} K_\alpha + D^{-1}_{\sigma,R})^{-1} \\
\mu_{\sigma_L} &= K'_\alpha 1_{-\sigma,T} \Omega(T)^{-1} \\
\Sigma_{\sigma_L} &= \left(n\Omega(T)^{-1} + D^{-1}_{\sigma,T}\right)^{-1} \\
\mu_{U_j} &= K'_U D_\sigma^{-1} F_{j,-U} + \Sigma_{R_j|-T}^{-1} \\
\Sigma_{U_j} &= \left(K'_U D_\sigma^{-1} K_U + \Sigma_{R_j|-T}\right)^{-1} \\
a_a &= 1 + (N_a - 1)/2 \\
b_a &= 1 + (\alpha_a^{(R)} \alpha_a^{(R)} )/2 \\
a_{\beta_k} &= 1 + (N_{\beta} - 1)/2 \\
b_{\beta_k} &= 1 + (B_{-1,k}^{(R)} B_{-1,k}^{(R)} )/2 \\
v_\Omega &= N_U + s + 1 + n \\
S_\Omega &= (10^{-3} I + U'U)^{-1}
\end{align*}
\]
Appendix B: Model comparison

In Tables 3, 4 and 5, we present the cross-validation results (in order) for Restionaceae, Proteaceae, and Aizoaceae. The joint model improves out-of-sample prediction performance for reflectances for all families, and this benefit is significant. On the other hand, the conditional out-of-sample prediction performance for traits depends on the family. For Asteraceae and Restionaceae, the families with the most data, prediction of plant traits benefits from joint modeling of traits and reflectance. For Proteaceae, plant trait predictions are slightly better under the independent model using the ES. However, we emphasize that the improved prediction is minimal for the independent model. For Aizoaceae, the family with the fewest data, prediction of plant traits suffers under the joint model. In summary, we find that reflectance predictions are uniformly and significantly better under the joint model for all plant families. Trait predictions are better under the joint model for two of the four families (in terms of ES) and only marginally worse for Proteaceae. We speculate that the benefit of the joint model appears when there is enough data to adequately

Table 4  Model comparison between joint and independent models for Proteaceae. 

| Quantity       | Model    | MAE     | RMSE   | ES    |
|----------------|----------|---------|--------|-------|
| Log lma        | [T|E]|R|E|   | 0.246   | 0.322   | 0.397   |
| Log fwc        | [T|E]|R|E|   | 0.186   | 0.274   |
| Log succulence | [T|E]|R|E|   | 0.283   | 0.367   |
| Log percent_N  | [T|E]|R|E|   | 0.217   | 0.287   |
| Log reflectance| [T|E]|R|E|   | 0.449   | 0.773   | 13.107  |
| Log lma        | [T,R]|E|   | 0.250   | 0.325   | 0.405   |
| Log fwc        | [T,R]|E|   | 0.215   | 0.303   |
| Log succulence | [T,R]|E|   | 0.279   | 0.361   |
| Log percent_N  | [T,R]|E|   | 0.238   | 0.303   |
| Log reflectance| [T,R]|E|   | 0.139   | 0.224   | 3.432   |

Table 5  Model comparison between joint and independent models for Aizoaceae. 

| Quantity       | Model    | MAE     | RMSE   | ES    |
|----------------|----------|---------|--------|-------|
| Log lma        | [T|E]|R|E|   | 0.496   | 0.322   | 0.728   |
| Log fwc        | [T|E]|R|E|   | 0.409   | 0.274   |
| Log succulence | [T|E]|R|E|   | 0.490   | 0.367   |
| Log percent_N  | [T|E]|R|E|   | 0.394   | 0.287   |
| Log reflectance| [T|E]|R|E|   | 0.444   | 0.773   | 12.705  |
| Log lma        | [T,R]|E|   | 0.793   | 0.325   | 1.063   |
| Log fwc        | [T,R]|E|   | 0.601   | 0.303   |
| Log succulence | [T,R]|E|   | 0.445   | 0.361   |
| Log percent_N  | [T,R]|E|   | 0.621   | 0.303   |
| Log reflectance| [T,R]|E|   | 0.140   | 0.224   | 3.158   |

[T|E]|R|E| represents the independent model, while [T, R|E] is the joint model.
estimate the relationship between traits and reflectance. Based on these findings, we use the joint model to present interpretation of the results.

Acknowledgements 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 for the collection of 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 with a Future Investigators in NASA Earth and Space Science and Technology (FINESST) grant award (80NSSC20K1659) to H.A. Frye and J.A. Silander.

Author contributions All authors contributed to early model development. PW, AG, HF, and JS wrote the main manuscript text. PW prepared all figures and tables. PW implemented models. All authors reviewed the manuscript.

Declarations

Competing interest The authors declare no competing interests.

References

Aiello-Lammens ME, Slingsby JA, Merow C, Mollmann HK, Euston-Brown D, Jones CS, Silander JA Jr (2017) Processes of community assembly in an environmentally heterogeneous, high biodiversity region. Ecography 40(4):561–576

Anderegg LDL (2023) Why can’t we predict traits from the environment? New Phytol 237(6):1998–2004

Anderegg LDL, Berner LT, Bagdley G, Sethi ML, Law BE, HilleRisLambers J (2018) Within-species patterns challenge our understanding of the leaf economics spectrum. Ecol Lett 21(5):734–744

Asner GP, Martin RE, Ford AJ, Metcalfe DJ, Liddell MJ (2009) Leaf chemical and spectral diversity in Australian tropical forests. Ecol Appl 19(1):236–253

Asner GP, Martin RE, Tupayachi R, Emerson R, Martinez P, Sinca F, Powell GVN, Wright SJ, Lugo AE (2011) Taxonomy and remote sensing of leaf mass per area (LMA) in humid tropical forests. Ecol Appl 21(1):85–98

Born J, Linder HP, Desmet P (2006) The greater cape floristic region: greater cape floristic region. J Biogeogr 34(1):147–162

Cadotte MW, Carscadden K, Mirotechnick N (2011) Beyond species: functional diversity and the maintenance of ecological processes and services: Functional diversity in ecology and conservation. J Appl Ecol 48(5):1079–1087

Caturegli L, Matteoli S, Gaetani M, Grossi N, Magni S, Minelli A, Corsini G, Remorini D, Volterrani M (2020) Effects of water stress on spectral reflectance of bermudagrass. Sci Rep 10(1):15055

Cavender-Bares J, Schneider FD, Santos MJ, Armstrong A, Carnaval A, Dahlin KM, Fatoyinbo L, Hurt GC, Schimel D, Townsend PA, Ustin SL, Wang Z, Wilson AM (2022) Integrating remote sensing with ecology and evolution to advance biodiversity conservation. Nat Ecol Evol 6(5):506–519

Clark JS, Nemergut D, Seyednasrollah B, Turner PJ, Zhang S (2017) Generalized joint attribute modeling for biodiversity analysis: median-zero, multivariate, multifarious data. Ecol Monogr 87(1):34–56

Coops N, Dury S, Smith M-L, Martin M, Ollinger S (2002) Comparison of green leaf eucalypt spectra using spectral decomposition. Aust J Bot 50(5):567

Díaz S, Cabido M (2001) Vive la différence: plant functional diversity matters to ecosystem processes. Trends Ecol Evol 16(11):646–655

Felsenstein J (1985) Phylogenies and the comparative method. Am Nat 125(1):1–15

Frye HA, Aiello-Lammens ME, Euston-Brown D, Jones CS, Kilroy Mollmann H, Merow C, Slingsby JA, Merwe H, Wilson AM, Silander JA (2021) Plant spectral diversity as a surrogate for species, functional and phylogenetic diversity across a hyper-diverse biogeographic region. Glob Ecol Biogeogr 30(7):1403–1417
Gneiting T, Raftery AE (2007) Strictly proper scoring rules, prediction, and estimation. J Am Stat Assoc 102(477):359–378

Heim R, Jürgens N, Große-Stoltenberg A, Oldeland J (2015) The effect of epidermal structures on leaf spectral signatures of ice plants (Aizoaceae). Remote Sens 7(12):16901–16914

Higdon D (2002) Space and space-time modeling using process convolutions. In: Quantitative methods for current environmental issues. Springer, pp 37–56

Hnimina G, Dufrêne E, Soudani K (2014) Relationship between photochemical reflectance index and leaf ecophysiological and biochemical parameters under two different water statuses: towards a rapid and efficient correction method using real-time measurements: Disentangling PRI variability. Plant Cell Environ 37(2):473–487

Inoue Y, Peñuelas J (2006) Relationship between light use efficiency and photochemical reflectance index in soybean leaves as affected by soil water content. Int J Remote Sens 27(22):5109–5114

Jacquemoud S, Ustin S (2019a) Leaf optical properties. 1st edn. Cambridge University Press, Cambridge

Jacquemoud S, Ustin S (2019b) Leaf optical properties. Cambridge University Press, Cambridge

Jacquemoud S, Ustin S (2019c) Variation due to leaf structural, chemical, and physiological traits. In: Leaf optical properties, 1st edn. Cambridge University Press, Cambridge

Jacquemoud S, Ustin S (2019d) Variations due to leaf abiotic and biotic factors. In: Leaf optical properties, 1 edn. Cambridge University Press, Cambridge

Jetz W, Cavender-Bares J, Pavlick R, Schimmel D, Davis FW, Asner GP, Guralnick R, Kattge J, Latimer AM, Moorcroft P, Schaeppman ME, Schildhauer MP, Schneider FD, Schrodt F, Stahl U, Ustin SL (2016) Monitoring plant functional diversity from space. Nat Plants 2(3):16024

JPL N (2020) NASADEM Merged DEM Global 1 arc second V001. Type: dataset

Kattge J, Bönisch G, Díaz S, Lavorel S, Prentice IC, Leadley P, Tautenhahn S, Werner GD, Aakala T, Abedi M et al (2020) Try plant trait database-enhanced coverage and open access. Glob Change Biol 26(1):119–188

Kothari S, Schweiger AK (2022) Plant spectra as integrative measures of plant phenotypes. J Ecol 110:1365–2745

Kothari S, Beauchamp-Rioux R, Blanchard F, Crofts AL, Girard A, Guilbeault-Mayers X, Hacker PW, Pardo J, Schweiger AK, Demers-Thibeault S, Bruneau A, Coops NC, Kalacska M, Vellend M, Laliberté E (2023) Predicting leaf traits across functional groups using reflectance spectroscopy. New Phytol 238(2):549–566

Latimer AM, Silander JA, Cowling RM (2005) Neutral ecological theory reveals isolation and rapid speciation in a biodiversity hot spot. Science 309(5741):1722–1725

Laughlin DC, Lusk CH, Bellingham PJ, Burslem DFRP, Simpson AH, Kramer-Walter KR (2017) Intraspecific trait variation can weaken interspecific trait correlations when assessing the whole-plant economic spectrum. Ecol Evol 7(21):8936–8949

Manning J (2013) The Extra Cape flora. Number 2 in plants of the Greater Cape floristic region/John Manning and Peter Goldblatt. SANBI, Pretoria

Manning J, Goldblatt P (eds) (2012) Plants of the greater cape floristic region. Number 29 in Strelitzia. SANBI, Biodiversity for Life, Pretoria

Meggill B, Enquist B, Weiher E, Westoby M (2006) Rebuilding community ecology from functional traits. Trends Ecol Evol 21(4):178–185

Meireles JE, Cavender-Bares J, Townsend PA, Ustin S, Gamon JA, Schweiger AK, Schaeppman ME, Asner GP, Martin RE, Singh A, Schrodt F, Chlus A, O’Meara BC (2020) Leaf reflectance spectra capture the evolutionary history of seed plants. New Phytol 228(2):485–493

Mitchell N, Moore TE, Mollmann HK, Carlson JE, Mocko K, Martinez-Caberra H, Adams C, Silander DA, Jones CS, Schlichting CD, Holsinger KE (2015) Functional traits in parallel evolutionary radiations and trait-environment associations in the cape floristic region of South Africa. Am Nat 185(4):525–537

Mitchell N, Lewis PO, Lemmon EM, Lemmon AR, Holsinger KE (2017) Anchored phylogenomics improves the resolution of evolutionary relationships in the rapid radiation of Protea L. Am J Bot 104(1):102–115

Mucina L, Rutherford MC (eds) (2006) The vegetation of South Africa, Lesotho and Swaziland. Number 19 in Strelitzia. South African National Biodiversity Institute, Pretoria. OCLC: ocn137259974

Myers N, Mittermeier RA, Mittermeier CG, Da Fonseca GA, Kent J (2000) Biodiversity hotspots for conservation priorities. Nature 403(6772):853–858

Orcival JM, Joffre R, Rambal S (1999) Exploring the relationships between reflectance and anatomical and biochemical properties in Quercus ilex leaves. New Phytol 143(2):351–364
Pau S, Nippert JB, Slapikas R, Griffith D, Bachle S, Helliker BR, O’Connor RC, Riley WJ, Still CJ, Zaricor M (2022) Poor relationships between NEON Airborne Observation Platform data and field-based vegetation traits at a mesic grassland. Ecology 103(2):e03590

Peñuelas J, Filella I, Biel C, Serrano L, Savé R (1993) The reflectance at the 950–970 nm region as an indicator of plant water status. Int J Remote Sens 14(10):1887–1905

Penuelas J, Pinol J, Ogaya R, Filella I (1997) Estimation of plant water concentration by the reflectance Water Index WI (R900/R970). Int J Remote Sens 18(13):2869–2875

Pérez-Harguindeguy N, Díaz S, Garnier E, Lavorel S, Poore J, Jaureguiberry P, Basset-Harte MS, Cornelissen JHC (2016) Corrigendum to: New handbook for standardised measurement of plant functional traits worldwide. Aust J Bot 64(8):715–716

Pirie MD, Oliver EGH, Mugrabi de Kuppler A, Gehrke B, Le Maître NC, Kandziora M, Bellstedt DU (2016) The biodiversity hotspot as evolutionary hot-bed: spectacular radiation of Erica in the Cape Floristic Region. BMC Evol Biol 16(1):190

Poorter H, Niinemets U, Poorter L, Wright JJ, Villar R (2009) Causes and consequences of variation in leaf mass per area (LMA): a meta-analysis. New Phytol 182(3):565–588

Pu R, Ge S, Kelly NM, Gong P (2003) Spectral absorption features as indicators of water status in coast live oak (Quercus agrifolia) leaves. Int J Remote Sens 24(9):1799–1810

Ramsay J (2005) Functional data analysis. Encyclopedia of statistics in behavioral science

Ramsay JO, Silverman BW (2007) Applied functional data analysis: methods and case studies. Springer, Berlin

Reich PB, Walters MB, Ellsworth DS (1997) From tropics to tundra: global convergence in plant functioning. Proc Natl Acad Sci 94(25):13730–13734

Reich PB, Ellsworth DS, Walters MB, Vose JM, Gresham C, Violin JC, Bowman WD (1999) Generality of leaf trait relationships: a test across six biomes. Ecology 80(6):1955–1969

Ripullone F, Rivelli AR, Baraldi R, Guarini R, Guerrieri R, Magnani F, Peñuelas J, Raddi S, Borghetti M (2011) Effectiveness of the photochemical reflectance index to track photosynthetic activity over a range of forest tree species and plant water statuses. Funct Plant Biol 38(3):177

Rodríguez-Pérez JR, Riaño D, Carlisle E, Ustin S, Smart DR (2007) Evaluation of hyperspectral reflectance indexes to detect grapevine water status in vineyards. Am J Enol Viticult 58(3):302–317

Schleuning M, Neuschulz EL, Allbrecht J, Bender IM, Bowler DE, Dehling DM, Fritz SA, Hof C, Mueller T, Nowak L, Sorensen MC, Böhning-Gaese K, Kissling WD (2020) Trait-based assessments of climate-change impacts on interacting species. Trends Ecol Evol 35(4):319–328

Schliep EM, Hoeting JA (2013) Multilevel latent gaussian process model for mixed discrete and continuous multivariate response data. J Agric Biol Environ Stat 18(4):492–513

Schliep EM, Gelfand AE, Mitchell RM, Aiello-Lammens ME, Silander JA Jr (2018) Assessing the joint behaviour of species traits as filtered by environment. Methods Ecol Evol 9(3):716–727

Schulze RE (1997) South African atlas of agrohydrology and climatology: contribution towards a final report to the water research commission on project 492. Technical Report TT82-96, Water Resource Commission, Pretoria, South Africa

Schweiger AK, Cavender-Bares J, Townsend PA, Hobbie SE, Madritch MD, Wang R, Tilman D, Gamon JA (2018) Plant spectral diversity integrates functional and phylogenetic components of biodiversity and predicts ecosystem function. Nat Ecol Evol 2(6):976–982

Seelig H, Hoehn A, Stodieck LS, Klaus DM, Adams WW III, Emery WJ (2008) The assessment of leaf water content using leaf reflectance ratios in the visible, near-, and short-wave-infrared. Int J Remote Sens 29(13):3701–3713

Serbin SP, Wu J, Ely KS, Kruger EL, Townsend PA, Meng R, Wolfe BT, Chlus A, Wang Z, Rogers A (2019) From the Arctic to the tropics: multibiome prediction of leaf mass per area using leaf reflectance. New Phytol 224(4):1557–1568

Shiklomanov AN, Dietze MC, Viskari T, Townsend PA, Serbin SP (2016) Quantifying the influences of spectral resolution on uncertainty in leaf trait estimates through a Bayesian approach to RTM inversion. Remote Sens Environ 183:226–238
Singh A, Serbin SP, McNeil BE, Kingston CC, Townsend PA (2015) Imaging spectroscopy algorithms for mapping canopy foliar chemical and morphological traits and their uncertainties. Ecol Appl 25(8):2180–2197

Thenot F, Méthly M, Winkel T (2002) The photochemical reflectance index (PRI) as a water-stress index. Int J Remote Sens 23(23):5135–5139

Verboom GA, Archibald JK, Bakker FT, Bellstedt DU, Conrad F, Dreyer LL, Forest F, Galley C, Goldblatt P, Henning JF, Mummenhoff K, Linder HP, Muasya AM, Oberlander KC, Savolainen V, Snijman DA, Niet Tvd, Nowell TL (2009) Origin and diversification of the Greater Cape flora: ancient species repository, hot-bed of recent radiation, or both? Mol Phylogenet Evol 51(1):44–53

Wang Z, Chlus A, Geygan R, Ye Z, Zheng T, Singh A, Couture JJ, Cavender-Bares J, Kruger EL, Townsend PA (2020) Foliar functional traits from imaging spectroscopy across biomes in eastern North America. New Phytol 228(2):494–511

Wang Z, Townsend PA, Kruger EL (2022) Leaf spectroscopy reveals divergent inter- and intra-species foliar trait covariation and trait-environment relationships across NEON domains. New Phytol 235(3):923–938

White PA, Keeler DG, Rupper S (2021) Hierarchical integrated spatial process modeling of monotone west antarctic snow density curves. Ann Appl Stat 15(2):556–571

White PA, Frye H, Christensen MF, Gelfand AE, Silander JA Jr (2022) Spatial functional data modeling of plant reflectances. Ann Appl Stat 16:1919–1936

White PA, Keeler DG, Sheanshang D, Rupper S (2022) Improving piecewise linear snow density models through hierarchical spatial and orthogonal functional smoothing. Environmetrics 33(5):e2726

Wright JP, Sutton-Grier A (2012) Does the leaf economic spectrum hold within local species pools across varying environmental conditions? Funct Ecol 26(6):1390–1398

Wright JJ, Reich PB, Westoby M, Ackerly DD, Baruch Z, Bongers F, Cavender-Bares J, Chapin T, Cornelissen JH, Diemer M et al (2004) The worldwide leaf economics spectrum. Nature 428(6985):821–827

Wüest RO, Litsios G, Forest F, Lexer C, Linder HP, Salamin N, Zimmermann NE, Pearman PB (2016) Resprouter fraction in Cape Restionaceae assemblages varies with climate and soil type. Funct Ecol 30(9):1583–1592

Yang X, Tang J, Mustard JF, Wu J, Zhao K, Serbin S, Lee J-E (2016) Seasonal variability of multiple leaf traits captured by leaf spectroscopy at two temperate deciduous forests. Remote Sens Environ 179:1–12

Zhang H (2004) Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. J Am Stat Assoc 99(465):250–261

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

**Authors and Affiliations**

Philip A. White¹² · Michael F. Christensen³ · Henry Frye⁴ · Alan E. Gelfand³ · John A. Silander Jr.⁴

✉ Philip A. White
pwhite@stat.byu.edu

Michael F. Christensen
michael.f.christensen@duke.edu

Henry Frye
henry.frye@uconn.edu

Alan E. Gelfand
alan@duke.edu
