BioSCape: Foliar Trait Maps derived from AVIRIS-NG imagery

Documentation Revision Date: 2026-11-04

Dataset Version: 1

Summary

This dataset provides maps of 20 foliar functional traits and 11 trait-relevant spectral indices derived from AVIRIS-NG imagery that was acquired during the Biodiversity Survey of the Cape (BioSCape) project in October-November 2023. Foliar traits were predicted with permutational partial least squares regression (PLSR) models trained with vegetation samples collected concurrently with image acquisition. The estimated foliar traits include concentrations of aluminum, boron, calcium, carbon, cellulose, copper, hemicellulose, iron, lignin, magnesium, manganese, nitrogen, phosphorus, potassium, sodium, sulfur, and zinc along with leaf mass per area and d13C and d15N isotope ratios. Each of the 20 trait data sets has associated pixelwise uncertainty estimations, expressed in units of the trait. The trait-relevant spectral indices are related to traits not sampled during image acquisition, such as pigment concentrations. The data are provided in cloud-optimized GeoTIFF format.

This dataset holds 196,884 cloud-optimized GeoTIFFs.

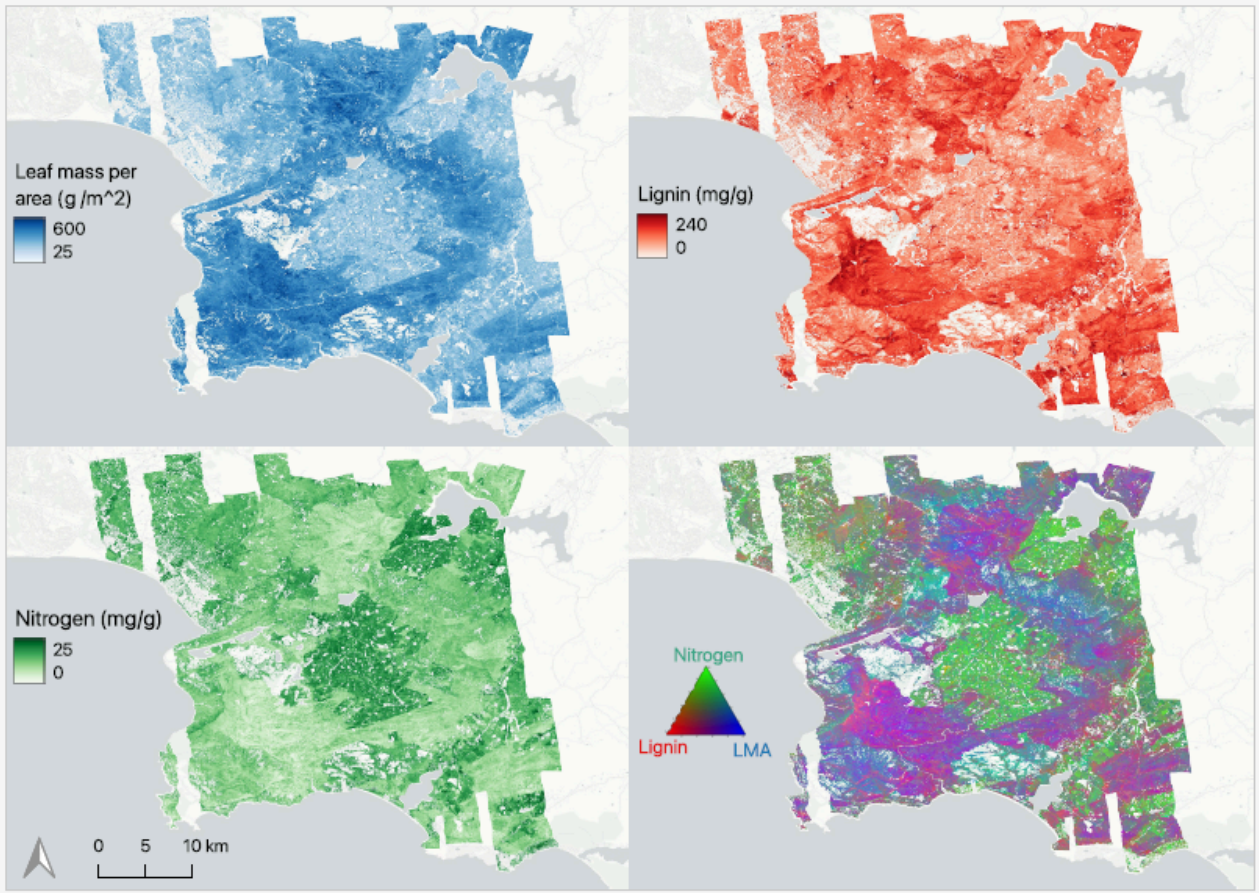


Figure 1. Example mosaicked trait maps over the Kogelberg region, one of epicenters of plant diversity and endemism in the Cape Floristic Region. The mosaics are created taking the lowest trait prediction uncertainty between overlapping tiles. Counter-clockwise from upper left are trait maps for lignin, leaf mass per area, nitrogen and finally a ternary map. The ternary map is created based on normalized values of the three traits relative to each other. Redder pixels represent areas where lignin is more prevalent relative to the other two traits, greener pixels are areas where nitrogen is more prevalent, and bluer pixels are areas where leaf mass per area is more prevalent.

Citation

Frye, H.A., D. Euston-Brown, K.R. Kovach, J.A. Slingsby, Z. Ye, and P.A. Townsend. 2026. BioSCape: Foliar Trait Maps derived from AVIRIS-NG imagery. ORNL DAAC, Oak Ridge, Tennessee, USA. https://doi.org/10.3334/ORNLDAAC/2521

Table of Contents

1. Dataset Overview
2. Data Characteristics
3. Application and Derivation
4. Quality Assessment
5. Data Acquisition, Materials, and Methods

6. References

1. Dataset Overview

This dataset provides maps of 20 foliar functional traits and 11 trait-relevant spectral indices derived from AVIRIS-NG imagery that was acquired during the Biodiversity Survey of the Cape (BioSCape) project in October-November 2023. Foliar traits were predicted with permutational partial least squares regression (PLSR) models trained with vegetation samples collected concurrently with image acquisition. The estimated foliar traits include concentrations of aluminum, boron, calcium, carbon, cellulose, copper, hemicellulose, iron, lignin, magnesium, manganese, nitrogen, phosphorus, potassium, sodium, sulfur, and zinc along with leaf mass per area and d13C and d15N isotope ratios. Each of the 20 trait data sets has associated pixelwise uncertainty estimations, expressed in units of the trait. The trait-relevant spectral indices are related to traits not sampled during image acquisition, such as pigment concentrations.

Project: Biodiversity Survey of the Cape (BioSCape)

The Biodiversity Survey of the Cape (BioSCape) is a research project that uses remote sensing and field data to understand the distribution, function, and importance of biodiversity on land and in the water around the Greater Cape Floristic Region (GCFR). Located in southwestern South Africa along the Atlantic coast, the GCFR is one of the world's 36 biodiversity hotspots. BioSCape is a collaboration between the US and South Africa. The team conducted an integrated field and airborne campaign in the hyperdiverse region in 2023. Researchers collected laser altimetry LiDAR data and UV/visible to short wavelength infrared (UVSWIR) and thermal imaging spectroscopy data over terrestrial and aquatic targets using four airborne instruments: AVIRIS-NG, PRISM, LVIS, and HyTES. This airborne dataset is unique in its size and scope and unprecedented in its instrument combination and level of detail. These airborne data are accompanied by a range of biodiversity-related field observations.

Related Dataset:

Kovach, K.R., Z. Ye, H. Frye, and P.A. Townsend. 2025. BioSCape: AVIRIS-NG L2B Enhanced Surface Reflectance. ORNL DAAC, Oak Ridge, Tennessee, USA. https://doi.org/10.3334/ORNLDAAC/2385

- This reflectance product is used as an input for PLSR models and index calculations.

Acknowledgments:

Funding for this project was provided by NASA's BioSCape project (grant 80NSSC22K0905). Additional funding was provided by the National Science Foundation's ASCEND Biology Integration Institute (BII) (grant DBI 2021898). This research was done using services provided by the OSG Consortium, which is supported by the National Science Foundation (awards 2030508, 2323298).

2. Data Characteristics

Spatial Coverage: Greater Cape Floristic Region (GCFR), South Africa

Spatial Resolution: nominal 5 m; varies from 1.9 to 6.3 m among flight scenes

Temporal Coverage: 2023-10-22 to 2023-11-26

Temporal Resolution: One time estimate

Site Boundaries: Latitude and longitude are given in decimal degrees.

Site	Westernmost Longitude	Easternmost Longitude	Northernmost Latitude	Southernmost Latitude
Greater Cape Floristic Region, South Africa	17.7891	26.2557	-31.2363	-34.9161

Data File Information

This dataset holds 196,884 cloud-optimized GeoTIFFs. The products include maps of 20 foliar functional traits predicted by permutational partial least squares regression (PLSR) models (n=145,840), 11 trait-relevant spectral indices (n=47,398), and model-training mask files (n=3646). They are organized by AVIRIS-NG flight scene (tile).

All GeoTIFFs use a projected UTM coordinate system based on WGS 84 datum; the UTM zones are 33S, 34S, and 35S (EPSG: 32733, 32734, 32735, respectively).

Functional trait maps from PLSR models

These PLSR files are named **<flight prefix>_<scene>*_L2A_OE_*<ver>*_RFL_ORT_geocorr_v1_plsr_*<trait>*.tif***, where

- <flight prefix> = flight line identifier, *angYYYYMMDDthhmmss*, encoding the date and time by year (YYYY), month (MM), day (DD), hour (hh), minute (mm), and second (ss) of the flight (e.g., ang20231022t092801).
- <scene> = scene ID along the flight line as a three digit number (e.g., "000", "001", "002", etc.)
- "*L2A_OE*" refers to the L2A reflectance product used for model input
- <ver> = unique eight character identifier of full heritage versioning.
- *"RFL_ORT_geocorr_v1*" indicates orthorectified surface reflectance with version 1 topographic, BRDF, and glint corrections applied.
- <trait> = the predicted foliar trait listed in Table 1.

These GeoTIFFs have two bands: Band 1 is the mean prediction across the 200 PLSR models. Band 2 is the standard deviation across those models, a measure of model uncertainty (not total prediction error).

For each PLSR file, an additional file is included that indicates whether the PLSR *predicted* value was inside (value 1) or outside (value 0) of the *observed* values for the community weighted mean trait values in the model training data (542 plots) (Table 1). Those files are named **<flight prefix>_<scene>_*L2A_OE*_<ver>_*RFL_ORT*_geocorr_v1_plsr_<trait>_range_mask.*tif***.

Example file names:
ang20231022t092801_000_L2A_OE_0b4f48b4_RFL_ORT_geocorr_v1_plsr_Aluminum_range_mask.tif
ang20231022t092801_000_L2A_OE_0b4f48b4_RFL_ORT_geocorr_v1_plsr_Aluminum.tif

Trait-relevant spectral indices

The trait-relevant spectral indices files provide 11 commonly derived vegetation indices not represented in the PLSR trait models: NDVI, EVI, NIRv, PRI, mARI, EWT, CIRE, CHL, CCI, CAR, and ANT (Table 2).
The files are named **<flight prefix>_<scene>_*L2A_OE*_<ver>_RFL_ORT_geocorr_v1_<index>.*tif***, where

- <index> = one of the indices listed Table 2.

Example file names:
ang20231022t092801_000_L2A_OE_0b4f48b4_RFL_ORT_geocorr_v1_ANT.tif
ang20231022t092801_000_L2A_OE_0b4f48b4_RFL_ORT_geocorr_v1_CAR.tif
ang20231022t092801_000_L2A_OE_0b4f48b4_RFL_ORT_geocorr_v1_CCI.tif

Model-training masks

The model-training mask applies the same vegetation and shadow filter used to select PLSR training pixels to every pixel in the scene. Pixels with NDVI ≥ 0.4 (computed from the AVIRIS-NG bands nearest 798 and 678 nm) and reflectance ≥ 0.1 at the band nearest 808 nm have a value of 1. All other pixels are 0, and pixels outside the image are 255 (nodata). The trait and index maps are not masked. Users should apply this mask, and each trait's range mask, before analysis. Across all scenes, about 34% of valid image pixels pass the mask, so sparsely vegetated and shadowed areas are mostly excluded.
These files are named **<flight prefix>_<scene>_*L2A_OE*_<ver>_RFL_ORT_geocorr_v1_masks.*tif***

Example file name: ang20231022t092801_000_L2A_OE_0b4f48b4_RFL_ORT_geocorr_v1_masks.tif

Table 1. Mapped foliar traits in the PLSR GeoTIFFs. Minimum and maximum observed are based on directly measured values of vegetation samples. These values are used to generate community-weighted means, and these ranges represent the most extreme individual plant measurements, not necessarily the range of values of community weighted means that the models were trained from. Macro- and micro-nutrients were derived from inductively coupled plasma optical emission spectroscopy, structural carbohydrates were derived from sequential digestion, and the isotopes and carbon values were derived from isotope ratio mass spectrometry.

Foliar Trait	Units	Minimum Observed Sample Value	Maximum Observed Sample Value
Aluminum	mg g^{-1}	0.0005*	1.3423
Boron	mg g^{-1}	0.0002	0.2013
Calcium	mg g^{-1}	0.10	77.40
Carbon	mg g^{-1}	320.40	688.20
Cellulose	mg g^{-1}	39.98	494.78
Copper	mg g^{-1}	0.0005*	0.0145
d13C (isotope ratio in per mille)	0.001 (‰)	-31.90	-13.19
d15N (isotope ratio in per mille)	0.001 (‰)	-9.89	13.51
Hemicellulose (plus bound proteins)	mg g^{-1}	22.36	498.23
Iron	mg g^{-1}	0.0005*	1.4283
lma (leaf mass per area)	g m^{-2}	37.63	3475.26**
lignin (Lignin and recalcitrants)	mg g^{-1}	8.21	356.52
Magnesium	mg g^{-1}	0.16	18.98
Manganese	mg g^{-1}	0.0005*	1.7545
Nitrogen	mg g^{-1}	2.00	46.50
Phosphorous	mg g^{-1}	0.10	4.37
Potassium	mg g^{-1}	1.00	49.30

Foliar Trait	Units	Minimum Observed Sample Value	Maximum Observed Sample Value
Sodium	mg g^{-1}	0.01	130.51
Sulfur	mg g^{-1}	0.31	10.79
Zinc	mg g^{-1}	0.0005*	0.2355

*Denotes a detection limit floor. **The maximum value for leaf mass per area is extremely high because culms from the Restionaceae family are included in this measure. These culms are the primary photosynthetic tissue of this prevalent vegetation component and are included in this measure.

Table 2. Foliar trait-relevant spectral indices and their derivation. $R_{<xxx>}$ denotes reflectance at wavelength of <*xxx*> nm.

Index (abbreviation)	Band math	Theoretical valid ranges	Reference
Anthocyanin index (ANT)	$(1 / R_{530–570} − 1 / R_{690–710}) × R_{760–800}$	0 to 50	Gitelson et al. (2006)
Carotenoid index (CAR)	$(1 / R_{510–520} − 1 / R_{690–710}) × R_{760–800}$	0 to 50	Gitelson et al. (2006)
Chlorophyll index (green) (CHL)	$(1 / R_{540–560} − 1 / R_{760–800}) × R_{760–800}$	0 to 50	Gitelson et al. (2006)
Chlorophyll Carotenoid Index (CCI)	$(R_{532} − R_{630}) / (R_{532} + R_{630})$	-1 to 1	Caplan and Huemmrich (2025), Gamon et al. (2016)
Chlorophyll Index Red Edge (CIRE)	$(R_{800} / R_{705}) − 1$	-1 to 1	Caplan & Huemmrich (2025); Gitelson and Solovchenko (2017)
Enhanced Vegetation Index (EVI)	$2.5 × ((R_{860} − R_{660}) / (R_{860} + 6 × R_{660} − 7.5 × R_{480} + 1))$	-1 to 1	Huete et al. (2002)
Equivalent Water Thickness (EWT)	$1 − (R_{1193} / R_{1126})$	-1 to 1	Schneider et al. (2017); Underwood (2003)
Modified Anthocyanin Reflectance Index (mARI)	$(1 / R_{550} − 1 / R_{705}) × R_{800}$	-1 to 1	Caplan and Huemmrich (2025); Gitelson et al. (2001)
Photochemical Reflectance Index (PRI)	$(R_{531} − R_{570}) / (R_{531} + R_{570})$	-1 to 1	Gamon et al. (1997)
Normalized difference vegetation index (NDVI)	$(R_{800} − R_{660}) / (R_{800} + R_{660})$	-1 to 1	Huete et al. (2002)
Near-infrared reflectance of *vegetation* (NIRv)	$(((R_{800} − R_{660}) / (R_{800} + R_{660})) − 0.08) × R_{800}$	-1 to 1	Badgley et al. (2017)

3. Application and Derivation

The Biodiversity Survey of the Cape (BioSCape) project acquired airborne image spectroscopy across the Cape Floristic Region using the AVIRIS-NG sensor (Cardoso et al., 2025). Concurrent to these flights, field teams inventoried vegetation and collected top of canopy samples to characterize the foliar traits of the dominant vegetation. These collections were used to train partial least squares regression models that were applied to the imagery to create trait maps. Together with the spectral indices, which stand in for traits that were not measured (such as pigments), these maps enable researchers to understand landscape-level distributions of foliar traits within a topographically complex biodiversity hotspot.

4. Quality Assessment

Partial least squares regression (PLSR) models were used to predict foliar traits. Model performance was assessed using root mean square error (RMSE), range normalized RMSE (NRMSE), and Nash-Sutcliffe Efficiency (NSE) (Table 3). RMSE is provided in units of the measurement and represents the approximate average error across the data. NRMSE is the RMSE divided by the range of observed values in the validation data. For example, an NRMSE of 0.067 means a typical error of about 6.7% of the trait's observed range. NSE is a R^2-like metric of model predictive skill that compares independent observations to model predictions. NSE ranges from negative infinity to 1, with NSE>0 interpreted as predictions that are better than the data mean trait value (i.e., acceptable, NSE=1 indicates perfect performance), while values <0 indicate poor model performance relative to the mean. All traits had mean NRMSE <0.2 and NSE >0, so every model predicted better than the trait mean. However, NSE values of 0.09 to 0.41 show that much of the plot-to-plot variation was not captured. Maps are best suited to landscape-scale patterns, not individual pixels. Both NRMSE and NSE should be considered when evaluating model reliability.

Table 3 reports performance values based on models trained using the infrared region (1000-2450 nm). The predictive performance of using the full spectrum (450-2450 nm; Figure 2) was also assessed. Given that performance was often similar between the two spectral ranges, the infrared region models were chosen for implementation because they were more parsimonious and avoided issues that may arise from correlation with pigments that

strongly affect the visible region. However, the full spectrum model was used for lignin, which showed modest improvement with a full spectrum model (likely leveraging the red edge).

Table 3. Summary of PLSR model performance with root mean squared error (RMSE) range normalized root mean squared error (NRMSE) and Nash-Sutcliffe Efficiency (NSE).

Foliar trait (unit)	RMSE	NRMSE	NSE
Sulfur (mg g^{-1})	0.437	0.067	0.241
Sodium (mg g^{-1})	6.869	0.078	0.205
Calcium (mg g^{-1})	4.226	0.081	0.26
Manganese (mg g^{-1})	0.047	0.093	0.096
d13C (‰)	1.638	0.093	0.086
Magnesium (mg g^{-1})	1.38	0.096	0.37
Zinc (mg g^{-1})	0.004	0.097	0.181
d15N (‰)	1.79	0.108	0.347
Carbon (mg g^{-1})	48.075	0.109	0.244
Nitrogen (mg g^{-1})	5.205	0.109	0.313
Iron (mg g^{-1})	0.035	0.115	0.09
Potassium (mg g^{-1})	2.405	0.118	0.413
Copper (mg g^{-1})	0.001	0.136	0.293
Boron (mg g^{-1})	0.008	0.137	0.249
Leaf mass per area (g m^{-2})	136.826	0.143	0.224
Aluminum (mg g^{-1})	0.029	0.154	0.133
Phosphorous (mg g^{-1})	0.331	0.156	0.399
Cellulose (mg g^{-1})	62.13	0.164	0.247
Hemicellulose (mg g^{-1})	70.598	0.176	0.169
Lignin(mg g^{-1})	53.961	0.192	0.182

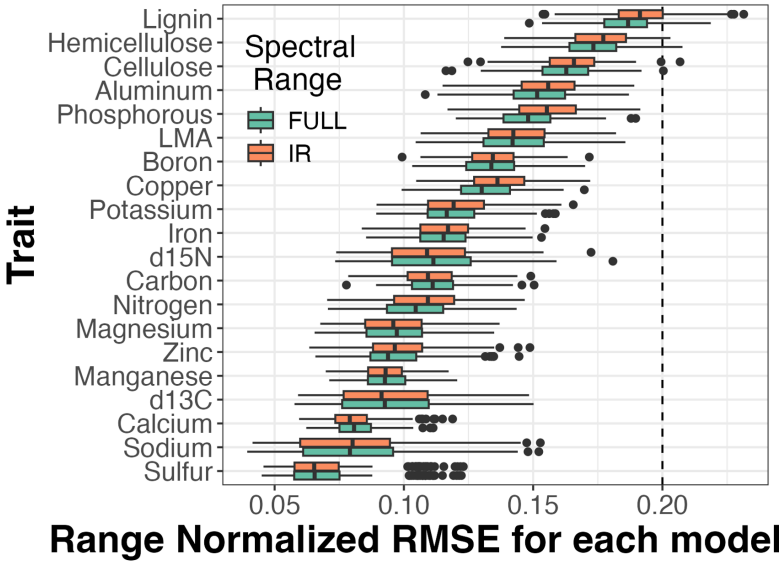


Figure 2. Comparison of model performance between full range (450–2450 nm) and infrared (1000–2450 nm) models across 200 model iterations.

5. Data Acquisition, Materials, and Methods

Foliar traits maps were created by applying partial least squares regression (PLSR) model coefficients to the AVIRIS-NG L2B enhanced reflectance product (Kovach et al., 2025). These PLSR models were trained using extracted AVIRIS-NG pixels and community-weighted mean traits derived from sample collections across 542 plots (Figure 3).

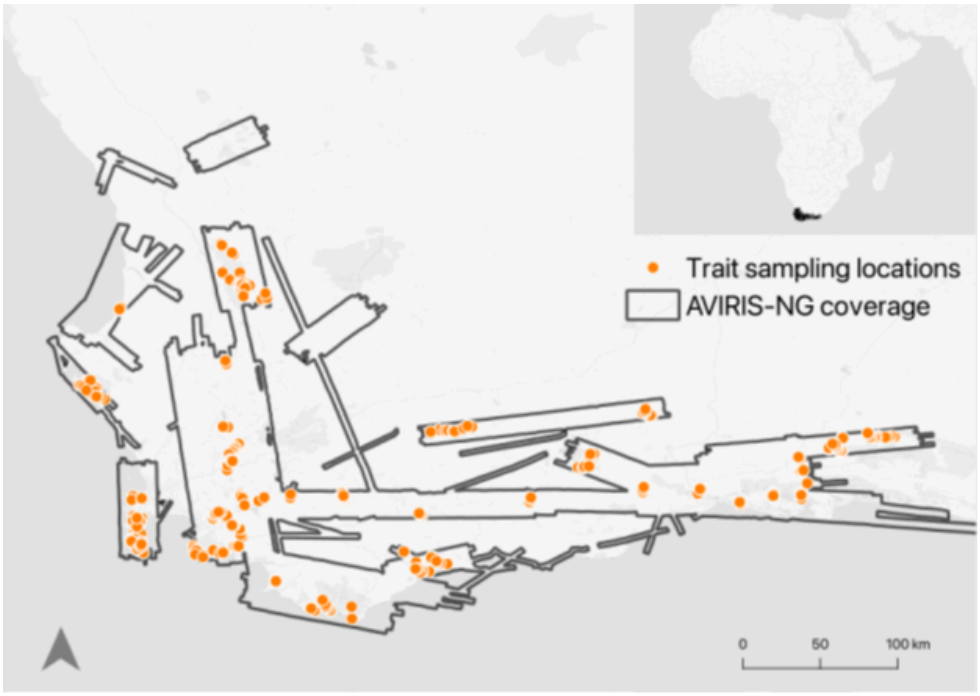


Figure 3. Map of trait sampling plots overlaid with a representation of the AVIRIS-NG coverage.

Plots consisted of either point locations marking the centers of 10-m circular plots or homogenous vegetation patches represented by hand delineated polygons. Pixels were extracted over point data using a 3 x 3 pixel window centered on the plot or all pixels whose centers fell within the polygon. Extracted pixels were first matched to community weighted mean traits based on image acquisition date. This resulted in a median 9-day difference between image acquisition date and trait collections, with a maximum of 27 days (Figure 4). In addition, 57 plots were collected the year following image acquisition to visit moist high elevation locations that were not logistically accessible during the main sampling efforts concurrent to image acquisitions in 2023. Rainfall totals were similar between years, and exploratory data analysis (not shown) did not reveal any trends that suggested that these plot values behaved differently than the rest of the data collected in 2023. Pixels were removed if NDVI (798, 678 nm) was below 0.4 or reflectance at 808 nm was below 0.1, to exclude sparse vegetation and shadow. The pixel with the highest NDVI values (i.e., greenest) was then retained for model development. Training data that included all pixels per plots was also tested and did not show strong differences between the best NDVI pixel selected training data.

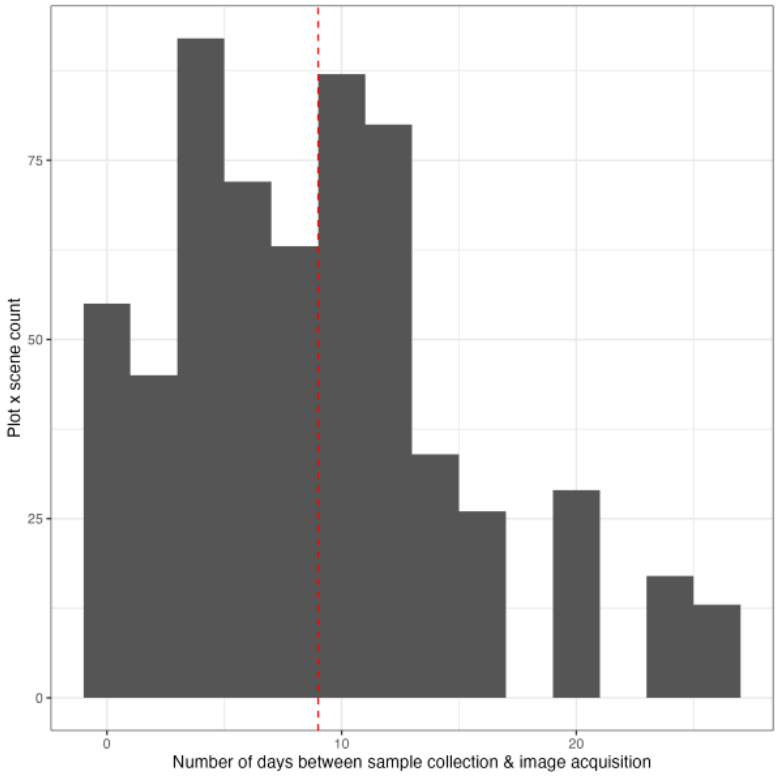


Figure 4. Distribution of the lowest temporal difference between field sample collection and image acquisition, withholding plots visited in 2024.

To train and validate PLSR models, an ensemble permutational approach was used. The data were first split to retain 85% of the data for training, withholding 15% for model validation. The initial 85% was further divided (85/15 split) to determine the optimal number of components used in the PLSR model. This inner splitting was repeated 30 times and the average optimal number of components across 30 splits was used to train a model. The model was validated using the initially withheld 15% of the data. This process was repeated 200 times, yielding 200 models per trait.Other model settings were: no y transformation, replication within samples allowed, and the best on average number of components selected from the inner model loop. To create foliar trait maps, these 200 models are applied to each image tile (scene); the mean prediction was stored as band 1 and the standard deviation of model predictions stored as band 2 in the tile GeoTIFF.

Trait-relevant spectral indices were created using the band math in Table 2. The closest AVIRIS-NG band center was selected based on the input band math, with a tolerance up to 10 nm. In the case of the Gitelson et al. (2006) indices, a broad band reflectance across a range of wavelengths was used in the index calculation.

Future versions of this dataset will include additional traits, notably nonstructural carbohydrates (e.g., sugars, starches, phenolics). Additional field data sets derived from fresh (green) leaf spectroscopy and chlorophyll meter measurements may be integrated into the index data. Finally, future versions will also include estimations of transformed versions of the data such as specific leaf area (SLA), the inverse of LMA, and stoichiometric ratios such as C:N and N:P. These metrics can be estimated from the data provided here but may differ from modeled estimates due to nonlinear relationships with spectra.

6. References

Badgley, G., C.B. Field, and J.A. Berry. 2017. Canopy near-infrared reflectance and terrestrial photosynthesis. Science Advances 3:e1602244. https://doi.org/10.1126/sciadv.1602244

Caplan, S., and K.F. Huemmrich. 2025. Unveiling PACE OCI's hyperspectral terrestrial data products. Remote Sensing Letters 16:422–433. https://doi.org/10.1080/2150704X.2025.2470905

Cardoso, A.W., E.L. Hestir, J.A. Slingsby, C.J. Forbes, G.R. Moncrieff, W. Turner, A.L. Skowno, J. Nesslage, P.G. Brodrick, K.D. Gaddis, and A.M. Wilson. 2025. The biodiversity survey of the Cape (BioSCape), integrating remote sensing with biodiversity science. npj Biodiversity 4:2. https://doi.org/10.1038/s44185-024-00071-5

Gamon, J.A., K.F. Huemmrich, C.Y. S. Wong, I. Ensminger, S. Garrity, D.Y. Hollinger, A. Noormets, and J. Peñuelas. 2016. A remotely sensed pigment index reveals photosynthetic phenology in evergreen conifers. Proceedings of the National Academy of Sciences 113:13087-13092. https://doi.org/10.1073/pnas.1606162113

Gamon, J.A., L. Serrano, and J.S. Surfus. 1997. The photochemical reflectance index: an optical indicator of photosynthetic radiation use efficiency across species, functional types, and nutrient levels. Oecologia 112:492-501. https://doi.org/10.1007/s004420050337

Gitelson, A.A., G.P. Keydan, and M.N. Merzlyak. 2006. Three-band model for noninvasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves. Geophysical Research Letters 33:L11402. https://doi.org/10.1029/2006GL026457

Gitelson, A.A., M.N. Merzlyak, and O.B. Chivkunova. 2001. Optical properties and nondestructive estimation of anthocyanin content in plant leaves. Photochemistry and Photobiology 74:38-45. https://doi.org/10.1562/0031-8655(2001)0740038OPANEO2.0.CO2

Gitelson, A., and A. Solovchenko. 2017. Generic Algorithms for Estimating Foliar Pigment Content. Geophysical Research Letters 44:9293-9298. https://doi.org/10.1002/2017GL074799

Huete, A., K. Didan, T. Miura, E.P. Rodriguez, X. Gao, and L.G. Ferreira. 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sensing of Environment 83:195–213. https://doi.org/10.1016/S0034-4257(02)00096-2

Kovach, K.R., Z. Ye, H. Frye, and P.A. Townsend. 2025. BioSCape: AVIRIS-NG L2B Enhanced Surface Reflectance. ORNL Distributed Active Archive Center. https://doi.org/10.3334/ORNLDAAC/2385

Schneider, F.D., F. Morsdorf, B. Schmid, O.L. Petchey, A. Hueni, D.S. Schimel, and M.E. Schaepman. 2017. Mapping functional diversity from remotely sensed morphological and physiological forest traits. Nature Communications 8:1441. https://doi.org/10.1038/s41467-017-01530-3

Underwood, E., S. Ustin, and D. DiPietro. 2003. Mapping nonnative plants using hyperspectral imagery. Remote Sensing of Environment 86:150-161. https://doi.org/10.1016/S0034-4257(03)00096-8