Imaging Spectroscopy Derived Maps of 17 Plant Functional Traits for the CHEESEHEAD Study Area

Ting Zheng*, Eliceo Ruiz, Brendan Heberlein, Zhiwei Ye, Philip A. Townsend

Ting Zheng: tzheng39@wisc.edu, 0000-0003-4728-6627, Department of Forest and Wildlife Ecology, University of Wisconsin-Madison. 1630 Linden Drive, Madison, WI 53706.

Eliceo Ruiz: eruiz@mail.ubc.ca, Faculty of Forestry & Environmental Stewardship, the University of British

Brendan Heberlein: bheberlein@wisc.edu, Department of Forest and Wildlife Ecology, University of Wisconsin-Madison.

Zhiwei Ye: ye6@wisc.edu, 0000-0001-5673-7034, Department of Forest and Wildlife Ecology, University of Wisconsin-Madison.

Philip A. Townsend: ptownsend@wisc.edu, 0000-0001-7003-8774, Department of Forest and Wildlife Ecology, University of Wisconsin-Madison.

*: lead and corresponding author.

Award ID: NSF Award #1822420: **Chequamegon Heterogeneous Ecosystem Energy-balance Study Enabled by a High-density Extensive Array of Detectors (CHEESEHEAD)**

1 Data Set Description

This dataset provides maps of 17 plant functional traits (leaf mass per area (LMA), non-structural carbohydrates (NSC), and foliar concentration of boron, carbon, cellulose, fiber, iron, lignin, magnesium, nitrogen, phenolics, phosphorus, potassium, starch, sugars, sulfur, and zinc) across a 10km × 10km forested region in Northern Wisconsin, USA. The maps were derived from HySpex imagery collected during the Chequamegon Heterogeneous Ecosystem Energy-balance Study Enabled by a High-density Extensive Array of Detectors 2019 project (CHEESEHEAD19, Butterworth et al., 2020) between June and August 2019 (https://doi.org/10.26023/NMQG-NGWE-3400). In specific, we mapped traits for the following dates: June 29th, July 11th and 13th, August 4th and 6th, and August 30th. Among these dates, acquisitions for June 29th and August 30th covered the entire study region, while both July and early August flights took two days to survey the region due to cloudy weather.

In total, the dataset contains 3332 trait files, consisting of 1666 ENVI binary files and 1666 corresponding header files.

We also included the outline of each flightline as KML files to help user identify which flightlines cover their region of interest without downloading the whole trait dataset.

Table 1. Flight acquisition dates and image spatial resolutions

Flight date	Number of flightlines	Raw resolution		Map resolution (m)
		VNIR(m)	SWIR(m)	
June 29, 2019	15	1.5	3.21	1
July 11, 2019	10	0.92	2.03	1
July 13, 2019	15	0.91	1.99	1
August 04, 2019	18	0.92	2.03	1
August 06, 2019	15	0.71	1.56	1
August 30, 2019	25	0.91	1.99	1

Data version: v1-06/12/2026

Data Status: Final

Time period and Data Frequency: see Table 1

Physical location: latitude: 45.9° N – 46° N; longitude: 90.21° W – 90.34° W

2 Instrument Description

This dataset is derived from HySpex imagery collected during the CHEESEHEAD campaign, for details on the HySpex sensor, please refer to the readme of the HySpex Reflectance Dataset: https://doi.org/10.26023/NMQG-NGWE-3400

3 Data Collection and Processing:

3.1 HySpex

Altitude between 1000m – 2000m, yielding a raw resolution between 0.7m – 1.5 m for the VNIR camera and 1.56m- 3.21m resolution for the SWIR camera. The raw data were orthorectified and atmospherically corrected to surface reflectance using an in-house processing pipeline (HyPro, https://github.com/EnSpec/hypro). All the reflectance data were processed to 1m spatial resolution for consistency. Reflectance from the SWIR camera were resampled using nearest neighbor to match the higher spatial resolution from the VNIR camera. The reflectance

data were further processed to remove the bidirectional reflectance distribution function (BRDF) effects using FlexBRDF (Queally et al., 2024).

3.2 Field data collection

Foliar samples were collected within two weeks of the image acquisition. We sampled in total 122 plots covering 83 broadleaf plots with 13 species, 35 conifer plots with 8 species, and 4 grass plots. The broadleaf and conifer plots were selected in two ways: 1. A single tree with a bigger than 2m crown radius or 2. A patch of mono-culture trees. For grass plots, we sampled from homogeneous grass patches. We sampled top of canopy foliar for the tree plots using a combination of pole pruner and customized air guns (Notch APTA: https://www.treestuff.com/apta-air-powered-tree-access/). For evergreen conifers, current year needles and old needles were sampled separately, and the portion of each age class on the top of canopy branch were estimated in the field. For grass plots, we clipped the material from 3 50cm*50cm quadrat. We recorded the location of each plot using a high precision GPS unit (GPS; Trimble Geo 7X; Trimble Inc., Sunnyvale, CA, USA).

3.3 Sample processing and plot trait estimation

Fresh foliar samples were put into a Ziploc bag with a piece of wet paper towel to maintain a moist condition. The Ziploc bags were stored in an ice cooler and transferred to the lab within 6 hours. Inside the lab, each foliar sample was divided into two parts, the first one is used to estimate leaf mass per area. It usually contains about 5-10 leaves for broadleaf and 20-50 needles for conifers. The area of these samples was estimated based on scans from a flat-bed scanner and the dry mass was obtained after 48-hour drying session in an oven set to 70˚C. The second part usually contains about 16g of fresh material, they were ground to 20mesh using a willey mill after 48-hour drying session in an oven set to 70˚C. The ground samples were packed into 3D printed trays and measured using an ASD Fieldspec 3 spectrometer ASD spectrometer for reflectance (ASD Inc., Boulder, CO, USA). We took six reflectance measurements per sample. The reflectance measurements were then jump-corrected using SpecDAL (https://github.com/EnSpec/SpecDAL) and averaged to yield one spectrum per sample.

We predicted 16 foliar function traits by applying existing leaf-level spectroscopic models (Wang et al., 2020; Zheng et al., 2024) to the processed reflectance spectra. For broadleaf and grass samples, the predicted traits are taken as the plot traits, while for evergreen conifers, the plot traits are the weighted average of traits from each age class.

3.4 Training data compilation

We overlay the plot locations on top of the HySpex images acquired most closely to the plot sampling date and manually delineate the boundaries of each plot. We then extracted spectra for all pixels within the plot boundary. We averaged the spectra after removing outliers following Zheng et al. (2024). The plot trait data and the plot spectra data were then joined to form one training dataset.

3.5 Model training:

We employed an ensemble approach for Partial Least Squared Regression (PLSR) model training. We first identified outliers in the dataset as the 2% of samples with the highest absolute residuals from 100 cross-validation permutations. After removing outliers, the dataset was split into 20% testing and 80% training subsets. The training subset was randomly divided into a calibration set (75%) and a validation set (25%). PLSR models were fitted on the calibration set

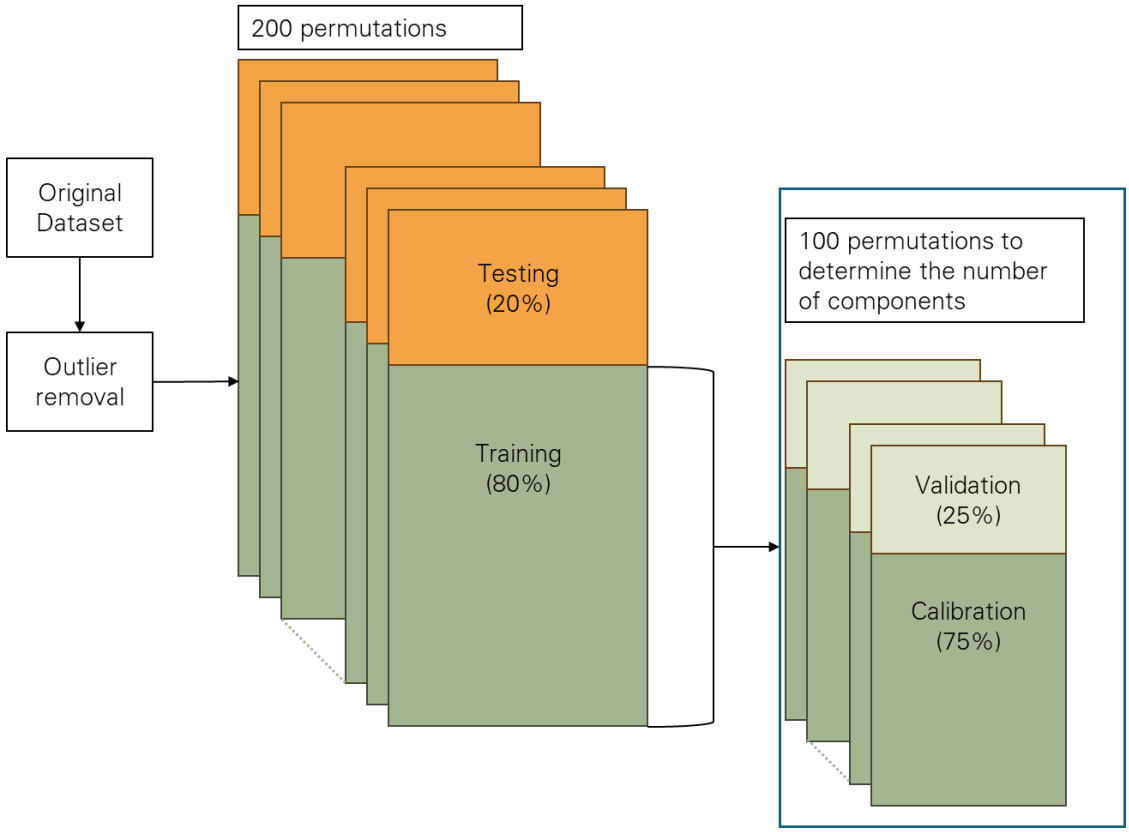

Fig. 1. The Partial Least Squared Regression (PLSR) model training workflow

and evaluated on the validation set. For each training subset, 100 permutations were performed, and the prediction residual error sum of squares (PRESS; Chen et al., 2004) was used to determine the optimal number of orthogonal spectral vectors. A final PLSR model was then fitted to the outlier-free training subset using the selected number of components and evaluated on the testing subset. This procedure was repeated across 200 different 20/80 splits, resulting in an ensemble of 200 PLSR models (Fig. 1).

3.6 Model performance

Table 2. PLSR model performance for different foliar functional traits. Metrics are reported for both training and testing subsets

trait	unit	Training				Testing			
		R^2	*RMSE*	*NRMSE*	*bias*	R^2	*RMSE*	*NRMSE*	*bias*
Boron	mg g^{-1}	0.47 ±0.02	0.01 ±0.00	16.18	0.00	0.35 ±0.14	0.01 ±0.00	19.59	0.00
Carbon	mg g^{-1}	0.61 ±0.07	12.31±1.12	12.17	0.00	0.41 ±0.17	14.65±1.63	17.59	0.15
Cellulose	mg g^{-1}	0.69 ±0.05	28.98±2.59	12.38	0.00	0.56 ±0.12	33.25±3.97	16.79	0.23
Fiber	mg g^{-1}	0.65 ±0.03	62.99±2.41	15.72	0.00	0.55 ±0.10	70.24±6.39	18.81	0.59
Iron	mg g^{-1}	0.53 ±0.05	0.01 ±0.00	13.73	0.00	0.38 ±0.15	0.01 ±0.00	18.00	0.00
Lignin	mg g^{-1}	0.64 ±0.05	46.85±2.85	15.33	0.00	0.49 ±0.13	54.09±5.33	19.25	1.46
LMA	g m^{-2}	0.84 ±0.03	33.92±3.17	9.39	0.00	0.70 ±0.10	43.73±5.51	14.19	0.49
Magnesium	mg g^{-1}	0.53 ±0.02	0.44 ±0.01	16.96	0.00	0.44 ±0.11	0.47 ±0.04	19.84	0.01
Nitrogen	mg g^{-1}	0.65 ±0.04	3.81 ±0.20	14.34	0.00	0.50 ±0.12	4.44 ±0.56	19.20	0.11
NSC	mg g^{-1}	0.56 ±0.07	33.53±2.37	11.91	0.00	0.37 ±0.16	39.56±3.98	17.98	0.90
Phenolics	mg g^{-1}	0.63 ±0.07	25.29±2.24	12.16	0.00	0.45 ±0.14	29.91±3.01	18.11	0.03
Phosphorus	mg g^{-1}	0.50 ±0.05	0.40 ±0.02	15.93	0.00	0.32 ±0.14	0.45 ±0.05	20.55	0.00
Potassium	mg g^{-1}	0.49 ±0.05	3.14 ±0.12	13.34	0.00	0.36 ±0.19	3.44 ±0.37	18.57	0.02
Starch	mg g^{-1}	0.59 ±0.06	8.20 ±0.61	11.48	0.00	0.32 ±0.20	10.21±1.47	17.74	0.43
Sugar	mg g^{-1}	0.59 ±0.06	33.27±2.47	12.26	0.00	0.39 ±0.13	39.72±4.01	18.48	0.08
Sulfur	mg g^{-1}	0.49 ±0.04	0.29 ±0.01	14.72	0.00	0.36 ±0.11	0.32 ±0.04	19.36	0.00
Zinc	mg g^{-1}	0.64 ±0.05	0.01 ±0.00	13.58	0.00	0.44 ±0.18	0.01 ±0.00	19.16	0.00

4 Data Format

Format: ENVI image format, Float32, 32-bit single-precision (one binary file and one ascii header file)

ENVI format interleave type: BSQ (Band Sequential)

Missing data value: -9999

File naming conventions:

CHEESEHEAD_YYYYMMDD_LL_Refl_deshift_T

CHEESEHEAD_YYYYMMDD_LL_Refl_deshift_T.hdr

YYYYMMDD: airborne imagery acquisition date (year-month-day)

LL: airborne imagery flightline order (two digits)

T: name of the trait, see column ‘trait’ in Table 2.

Information of each band is listed in Table 3.

Table 3 Trait raster band description

Band	Band Name	Description	Data Type	Data Range	Unit
1	T_mean: T is the trait name	Model predicted ensemble average trait value	Float32	No explicit range on valid pixels -9999: nodata	See 'unit' column in Table 2
2	T_std: T is the trait name	Model predicted trait standard deviation from 200 permutations	Float32	(0,+∞): valid pixel -9999: nodata	See 'unit' column in Table 2
3	range_mask	trait range mask. The proper range for a certain trait is calculated based on the range from field observations: range_min = field_min - 0.25 * field_range range_max = field_max + 0.25 * field_range For traits that shouldn't have negative values, the range_min will be the maximum between 0 and calculated range_min	Float32	1: trait mean in the range 0: trait mean outside of the range -9999: nodata	NA
4	ndi	NDVI (Red: 660nm, NIR: 850nm) range mask (0.05 - 1) on pixels	Float32	1: NDVI in the range (0.05,1) 0: other valid pixels -9999: nodata	NA
5	neon_edge	30 pixel buffer from the edge of valid image	Float32	0: pixels within 30-pixel buffer of the valid image edge 1:other valid pixels that are at least 30 pixels away from the edge of the image -9999: nodata	NA

5 Data Remarks

N/A

6 References

Butterworth, B.J., Desai, A.R., Metzger, S., Townsend, P.A., Schwartz, M.D., Petty, G.W., Mauder, M., Vogelmann, H., Andresen, C.G., Augustine, T.J., Bertram, T.H., Brown, W.O.J., Buban, M., Cleary, P., Durden, D.J., Florian, C.R., Guzman, E.R., Iglinski, T.J.,

Kruger, E.L., Lantz, K., Lee, T.R., Meyers, T.P., Mineau, J.K., Olson, E.R., Oncley, S.P., Paleri, S., Pertzborn, R.A., Pettersen, C., Plummer, D.M., Riihimaki, L., Sedlar, J., Smith, E.N., Speidel, J., Stoy, P.C., Sühring, M., Thom, J.E., Turner, D.D., Vermeuel, M.P., Wagner, T.J., Wang, Z., Wanner, L., White, L.D., Wilczak, J.M., Wright, D.B., Zheng, T., 2020. Connecting Land-Atmosphere Interactions to Surface Heterogeneity in CHEESEHEAD 2019. https://doi.org/10.1002/essoar.10503532.1

Chen, S., Hong, X., Harris, C.J., Sharkey, P.M., 2004. Sparse Modeling Using Orthogonal Forward Regression With PRESS Statistic and Regularization. IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics) 34, 898–911. https://doi.org/10.1109/TSMCB.2003.817107

Wang, Z., Chlus, A., Geygan, R., Ye, Z., Zheng, T., Singh, A., Couture, J.J., Cavender-Bares, J., Kruger, E.L., Townsend, P.A., 2020. Foliar functional traits from imaging spectroscopy across biomes in eastern North America. New Phytologist 228, 494–511. https://doi.org/10.1111/nph.16711

Zheng, T., Ye, Z., Singh, A., Desai, A.R., Krishnayya, N.S.R., Dave, M.G., Townsend, P.A., 2024. Variability in Forest Plant Traits Along the Western Ghats of India and Their Environmental Drivers at Different Resolutions. Journal of Geophysical Research: Biogeosciences 129. https://doi.org/10.1029/2023JG007753

7 Appendix

Keyword: Hyperspectral Imagery, Vegetation, Plant Functional Traits, Temperate Forest Mix