AVIRIS-NG and field data to explore functional traits

Phil Townsend

Imaging Spectroscopy for Plant Traits

Phil Townsend

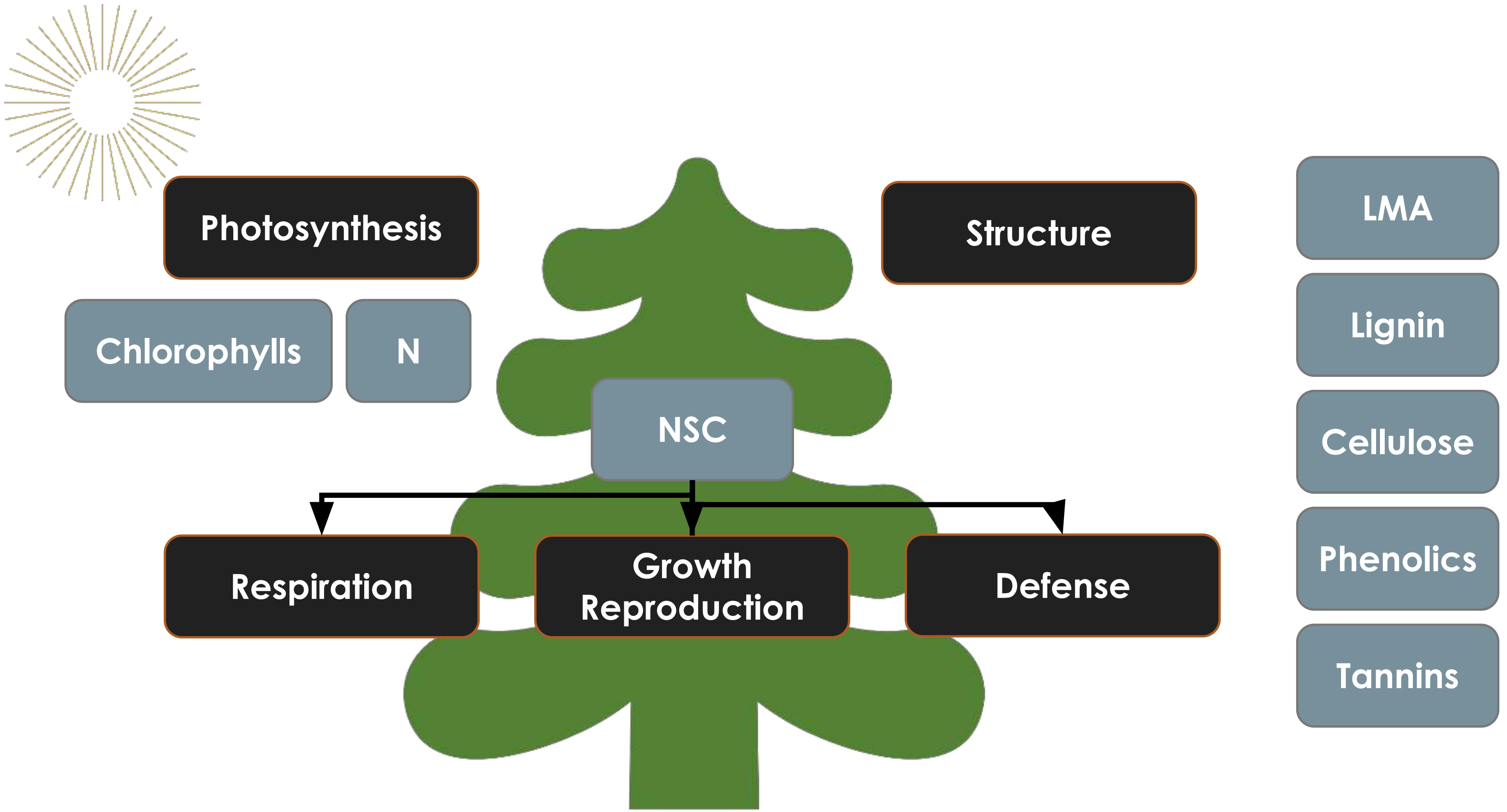
Acknowledgments: Adam Chlus, Phuong Dao, Henry Frye, Kyle Kovach, Shawn Serbin, Aditya Singh, Zhihui Wang, Sarah Wegmueller, Ting Zheng

Includes both:

AVIRIS-NG and field data to explore functional traits

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

BIOSCAPE


Photosynthesis
Chlorophylls
N
Structure
LMA
Lignin
Cellulose
Phenolics
Tannins
NSC
Respiration
Growth
Reproduction
Defense

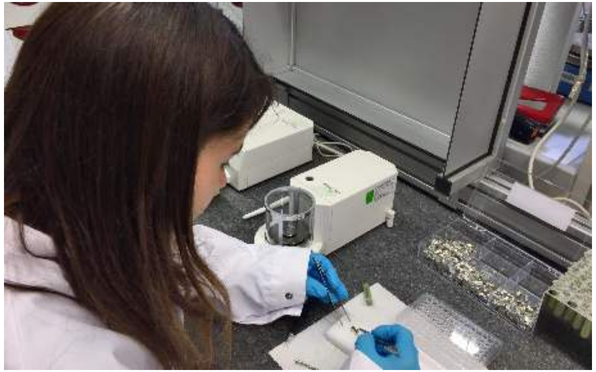
Sample Collection

- Sample sunlit foliage (top of canopy)

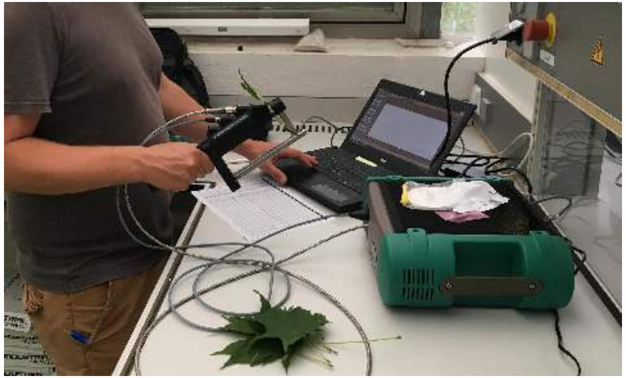
Sample Collection

Field assays

Lab assays



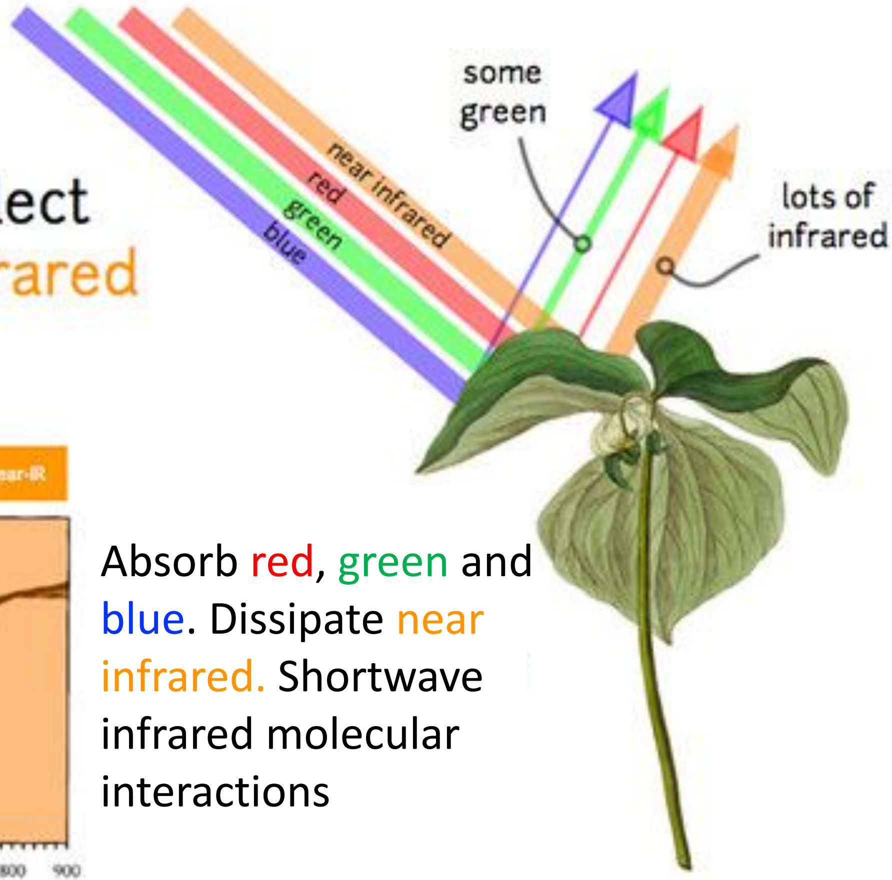
Fresh spectra

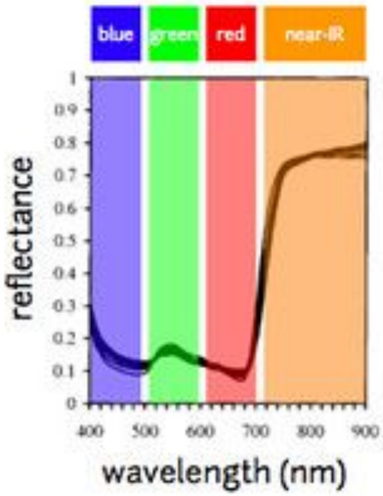


Dry spectra


HAND PACKED

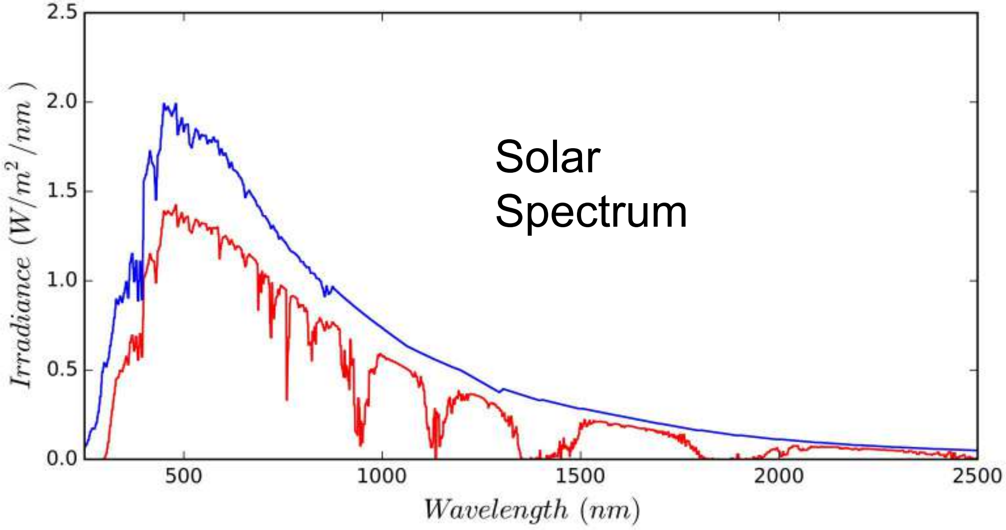
Why do plants reflect lots of infrared light?



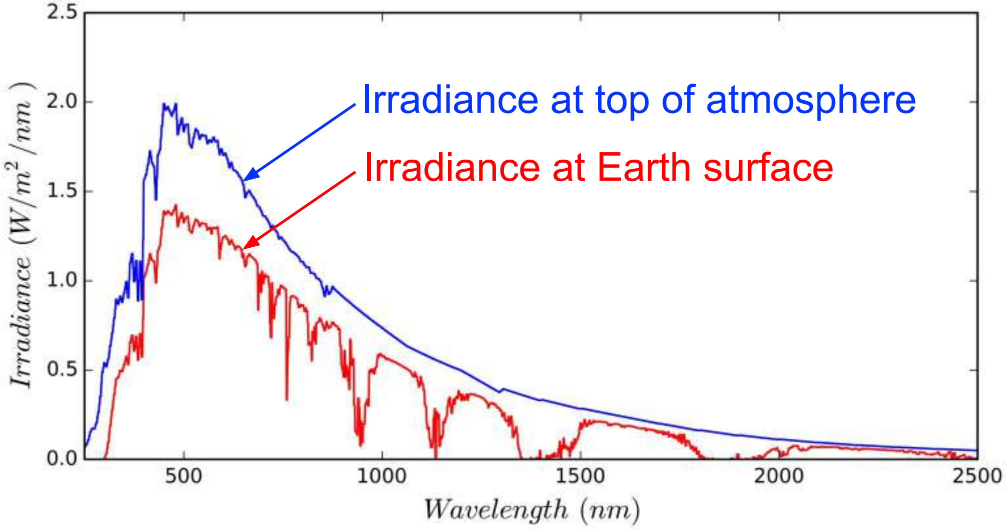


Absorb red, green and blue. Dissipate near infrared. Shortwave infrared molecular interactions

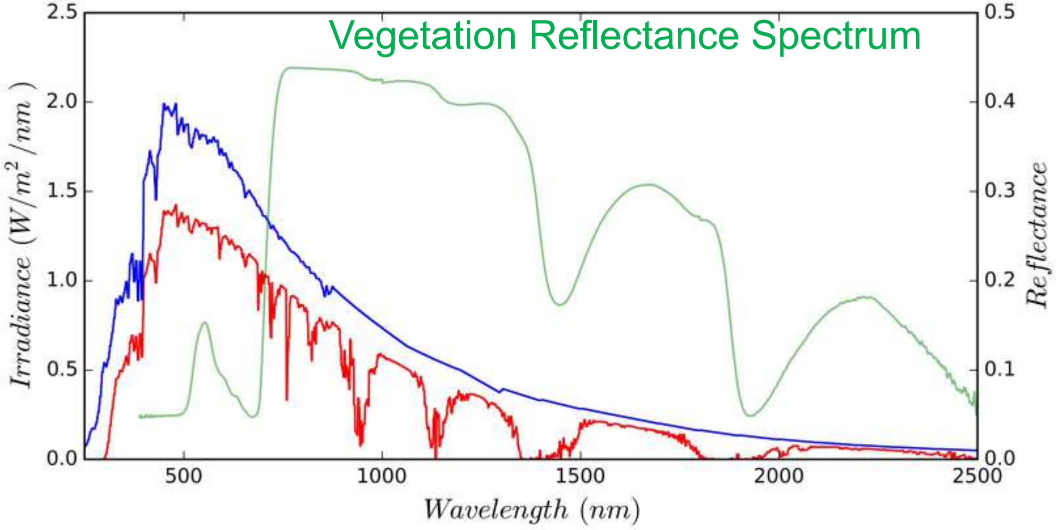
Photosynthesis and Radiation



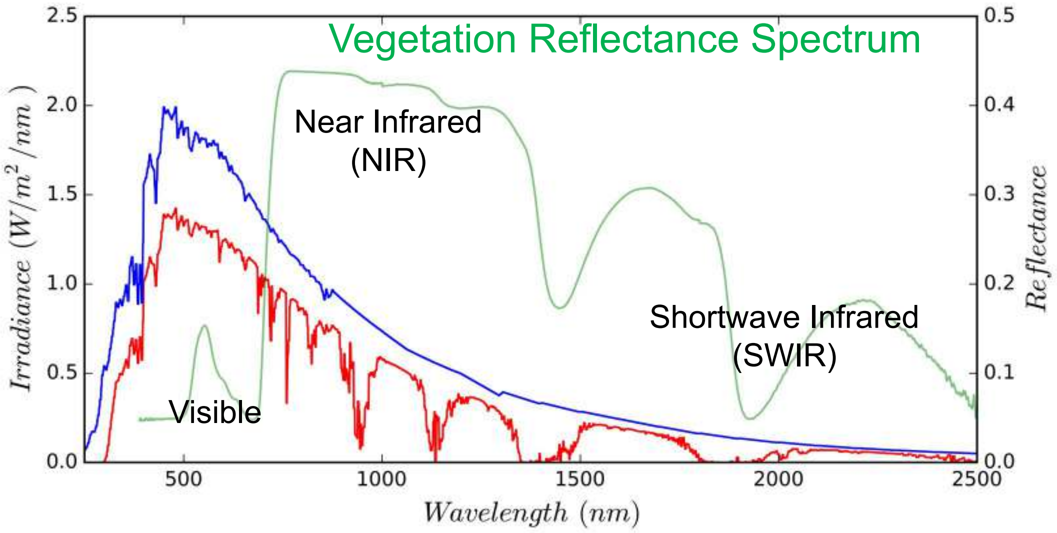
Photosynthesis and Radiation



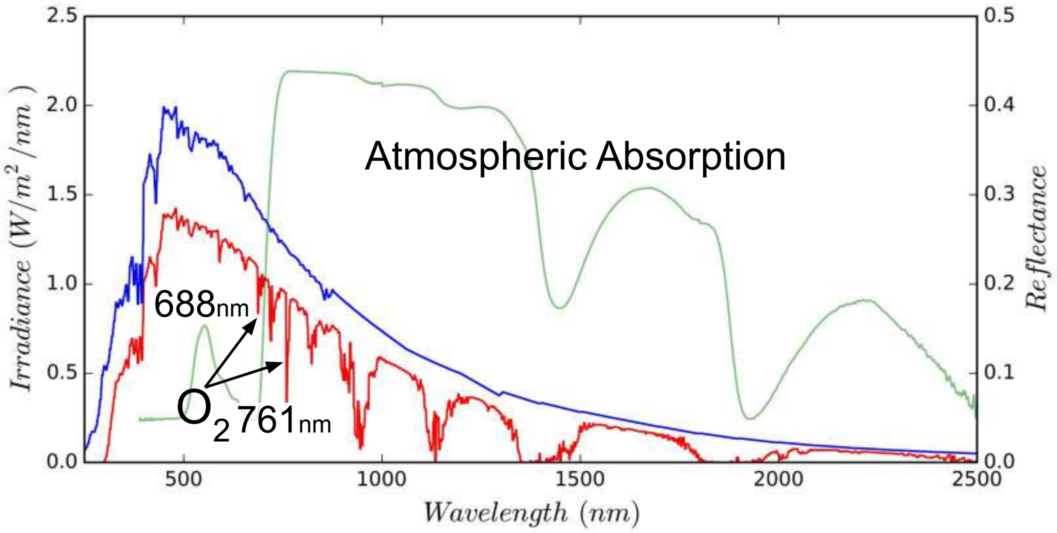
Photosynthesis and Radiation



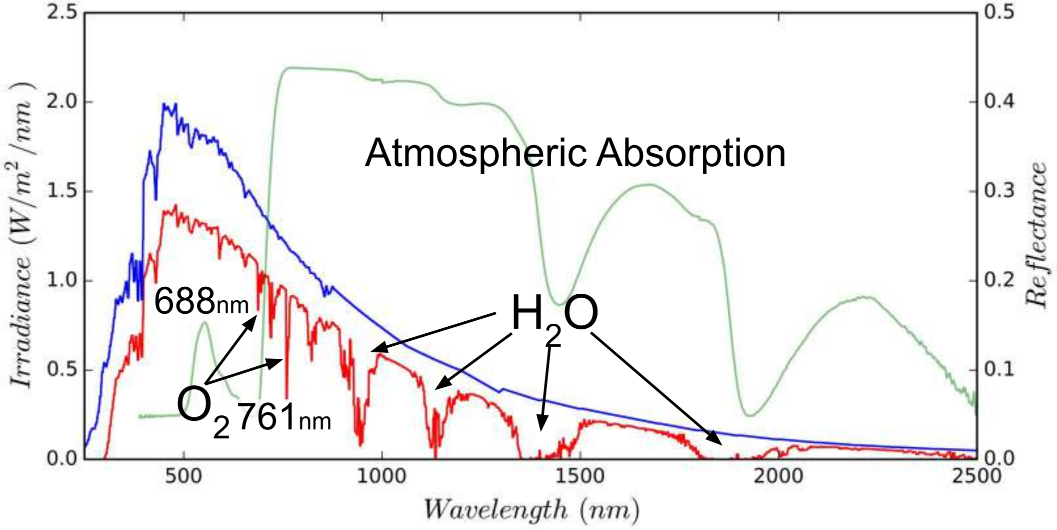
Photosynthesis and Radiation



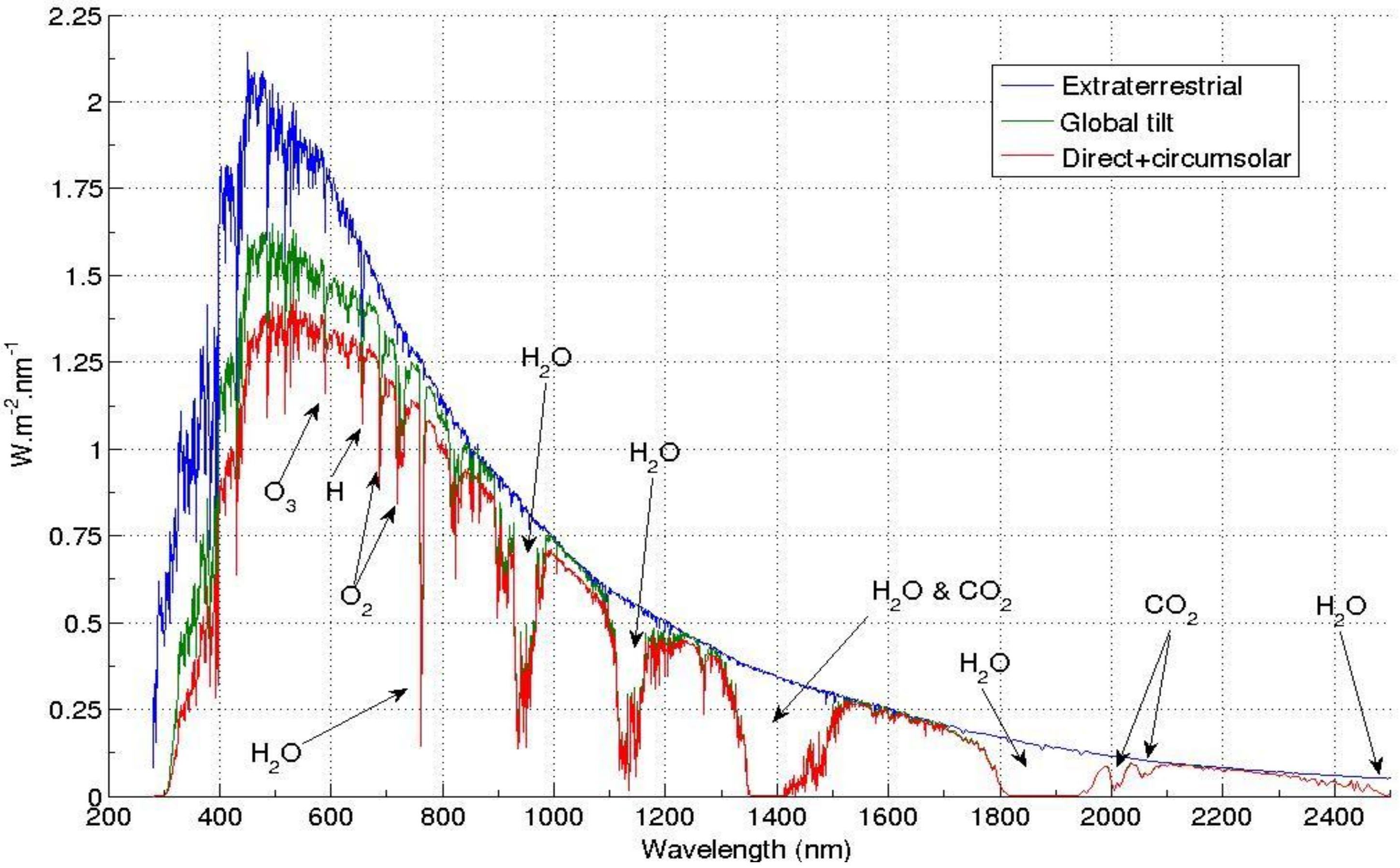
Photosynthesis and Radiation



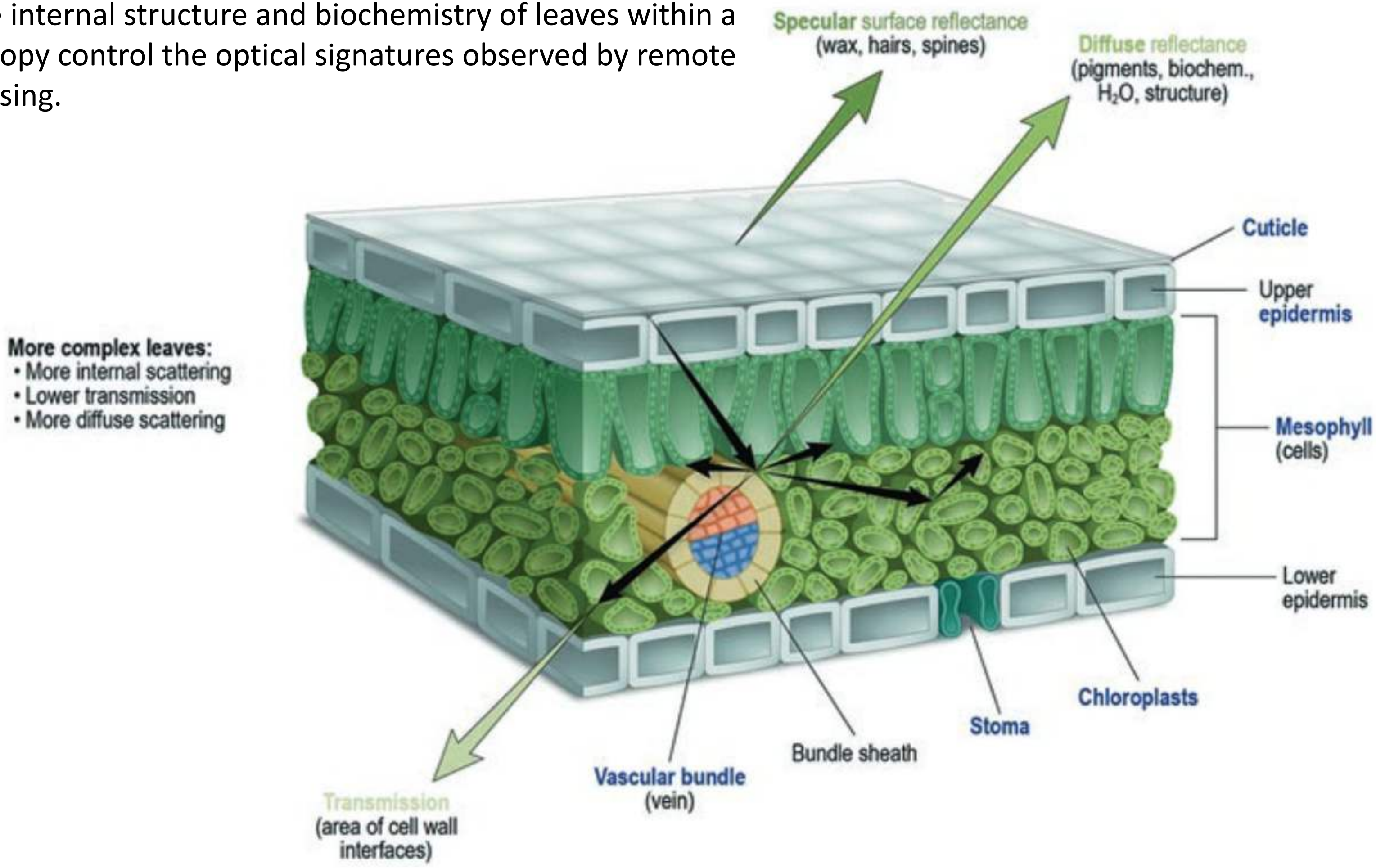
Photosynthesis and Radiation

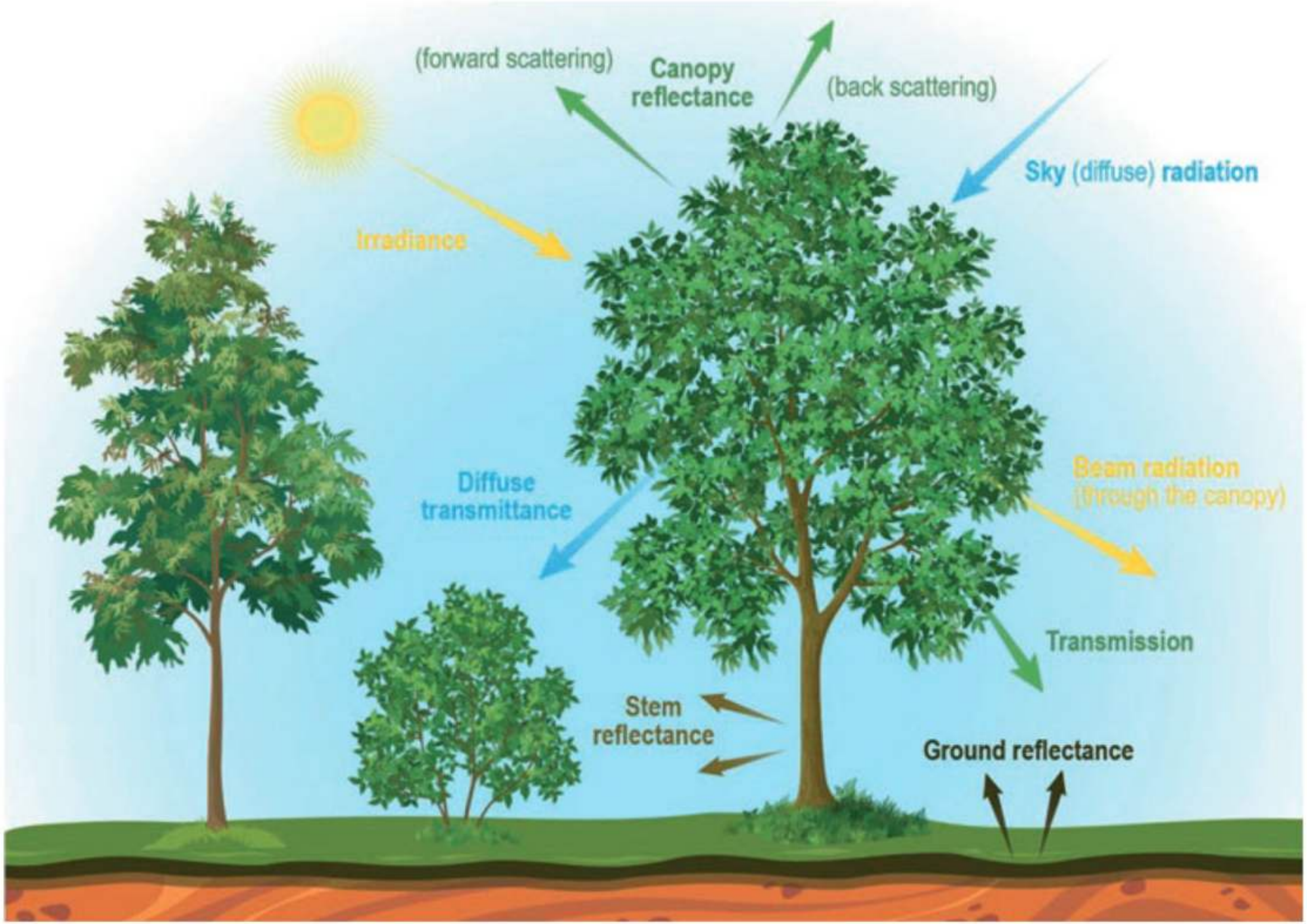


The sun is our light source…..



The internal structure and biochemistry of leaves within a canopy control the optical signatures observed by remote sensing.




(forward scattering)
Canopy reflectance
(back scattering)
Sky (diffuse) radiation
Irradiance
Diffuse transmittance
Beam radiation (through the canopy)
Transmission
Stem reflectance
Ground reflectance

Plant Traits from Imagery – An Analogy

4-band MULTISPECTRAL
Red, Green, Blue, and Near Infrared (NIR)

Most airborne imagery

8-12 band MULTISPECTRAL
Red, Green, Blue, NIR
+
Short wave infrared (SWIR), Red Edge, others

Most satellites

Hundreds of bands: HYPERSPECTRAL
Able to distinguish many more colors and tree traits

AVIRIS-NG, EMIT

Reflectance Spectroscopy

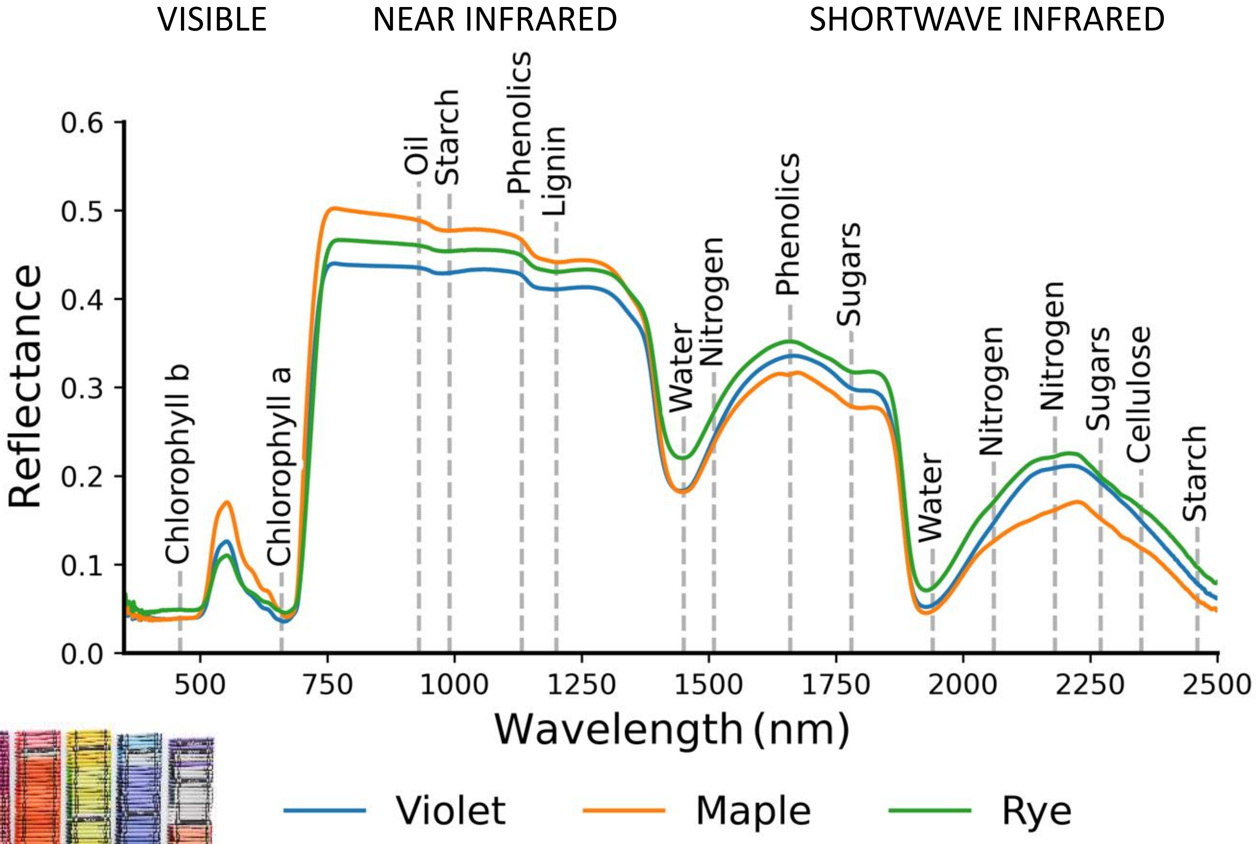


Figure: Adam Chlus

Reflectance Spectroscopy

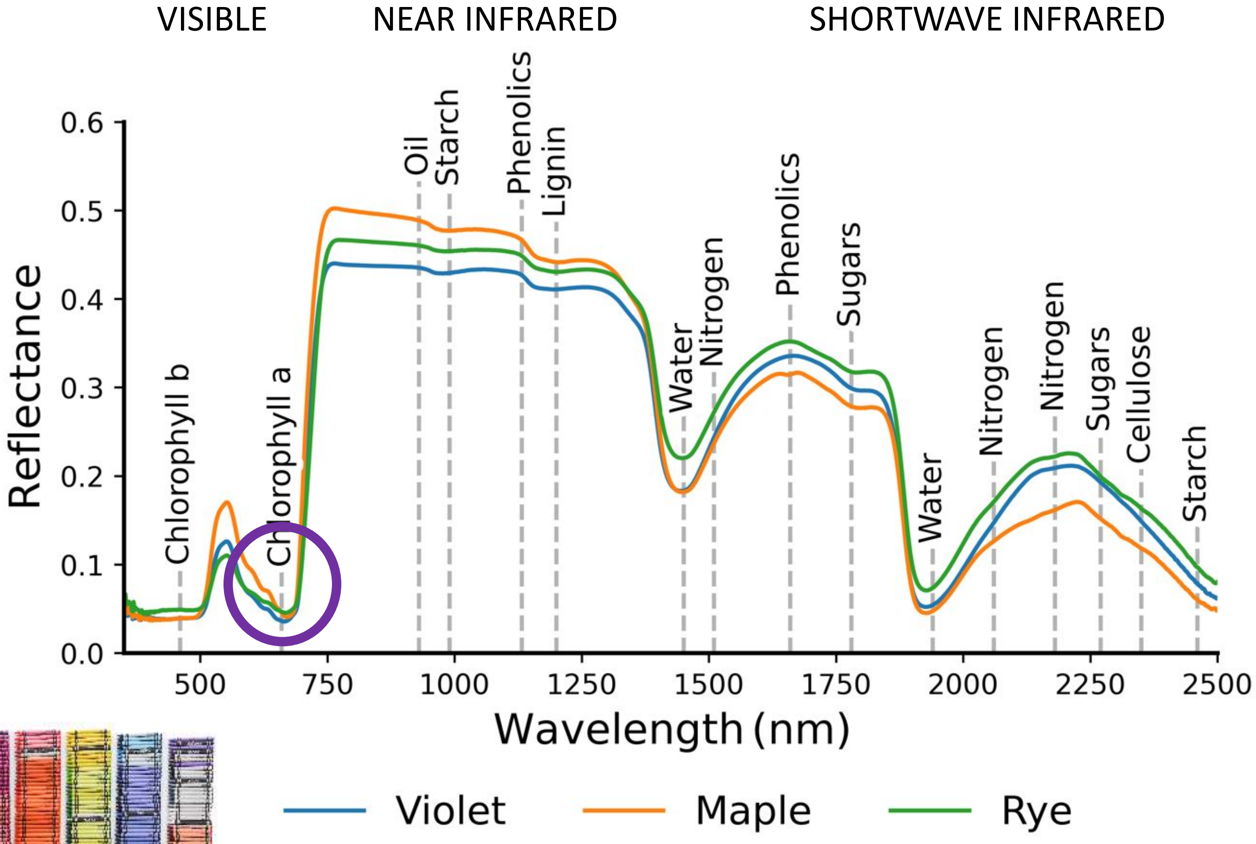


Figure: Adam Chlus

Reflectance Spectroscopy

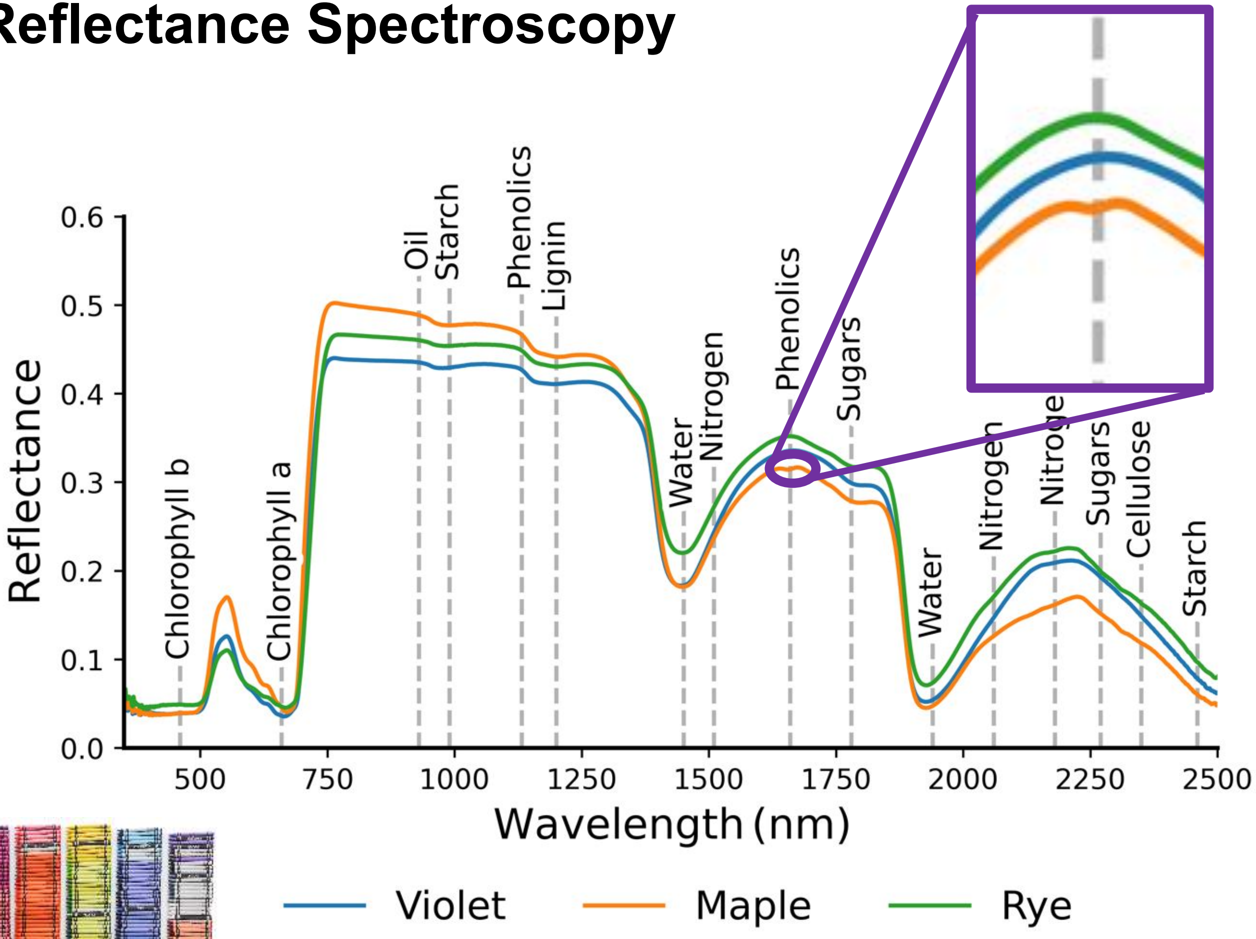
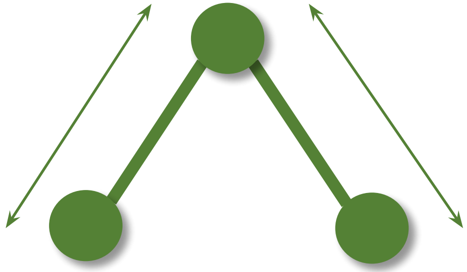


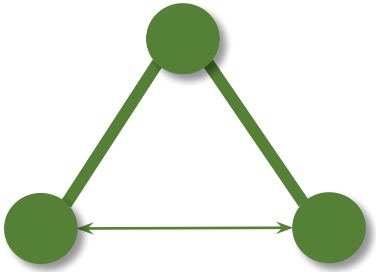
Figure: Adam Chlus

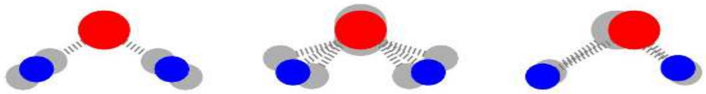
Vegetation spectroscopy: Why does this work?

Stretching



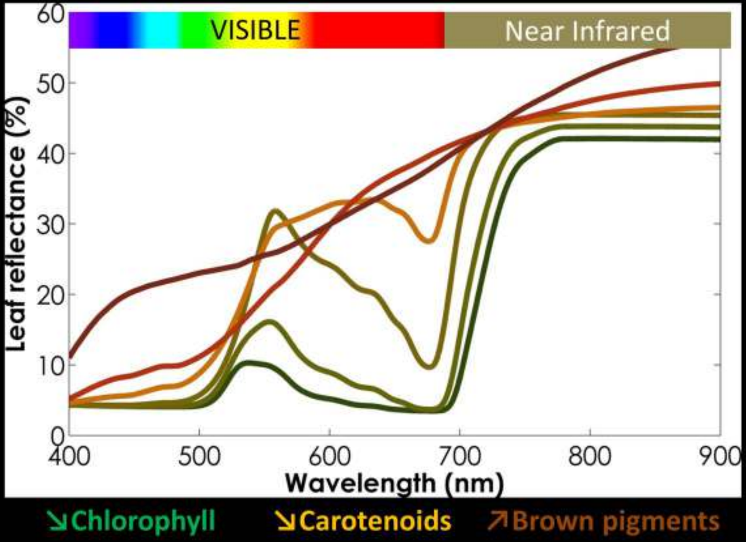
Bending & Twisting

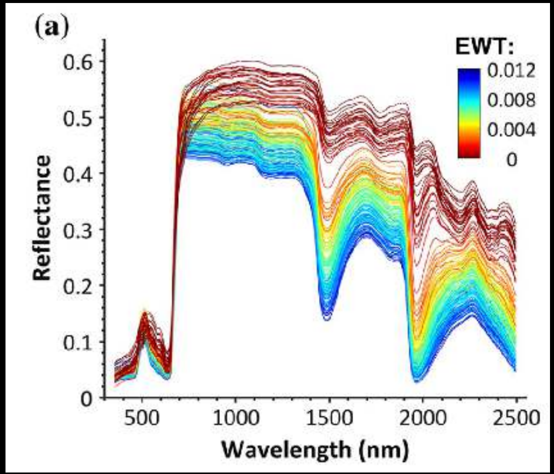




Harmonics of O-H, C-H and N-H stretches

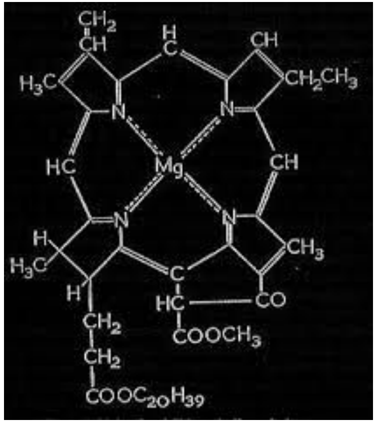
What can leaf/canopy optics

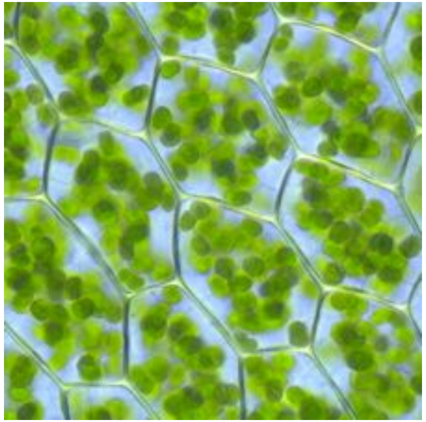


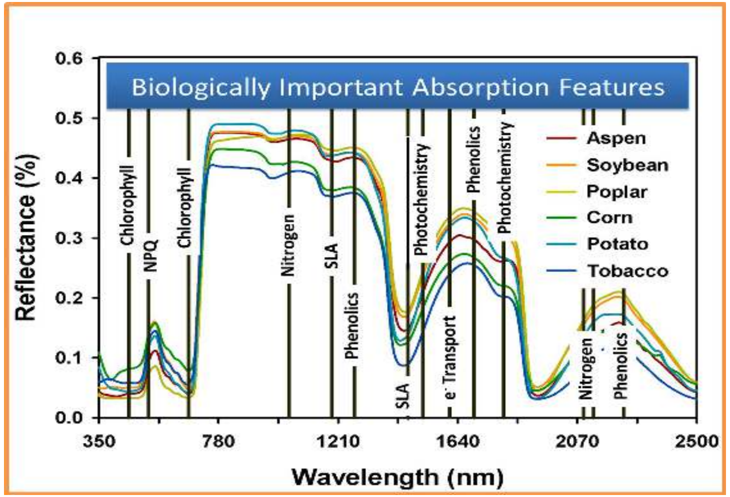


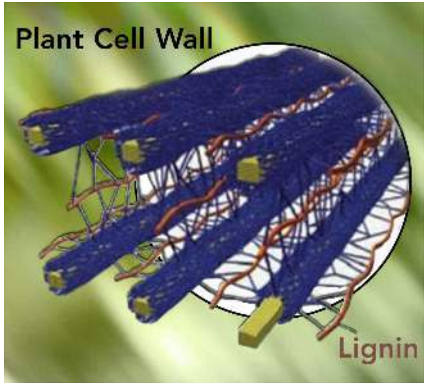
Drying experiment in oak leaves, Hill et al. 2019

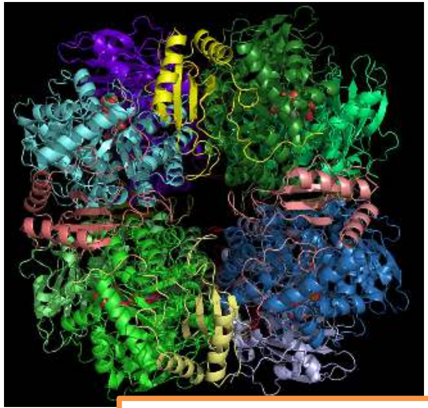
Leaf structure includes leaf constituents

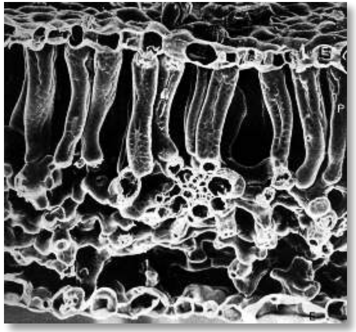


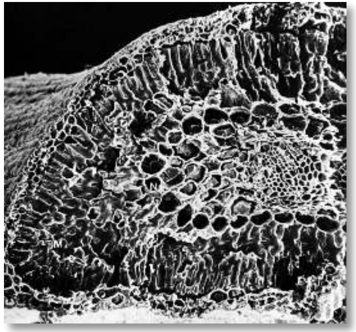


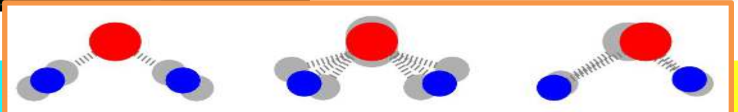




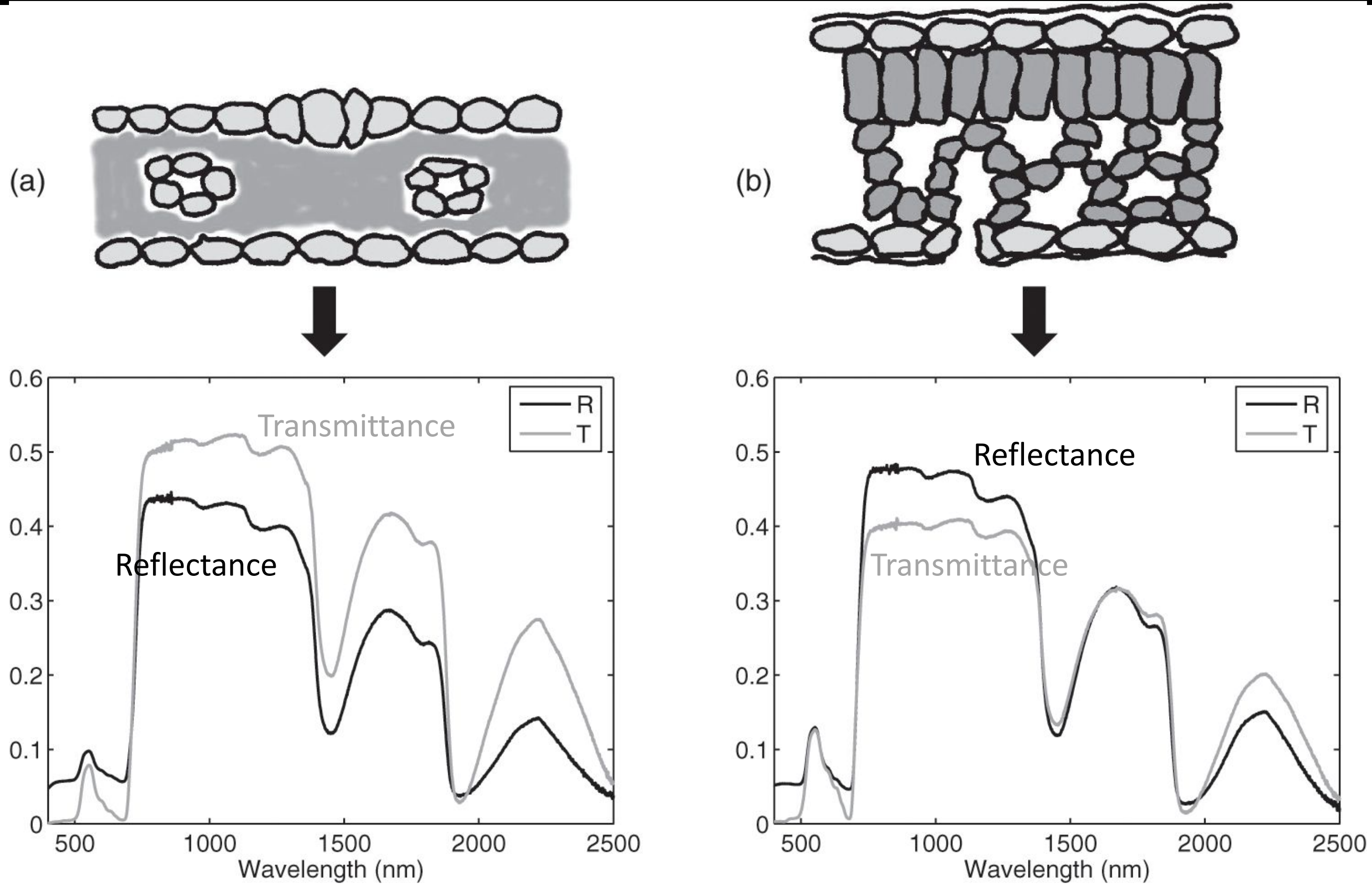




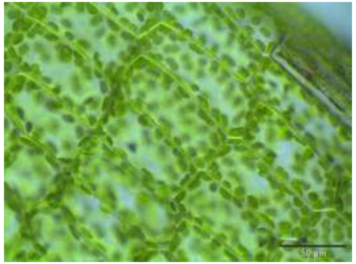


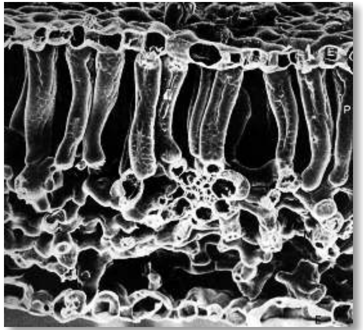


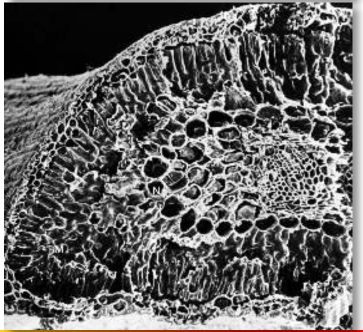
Leaf structure



Jacquemoud and Ustin 2019

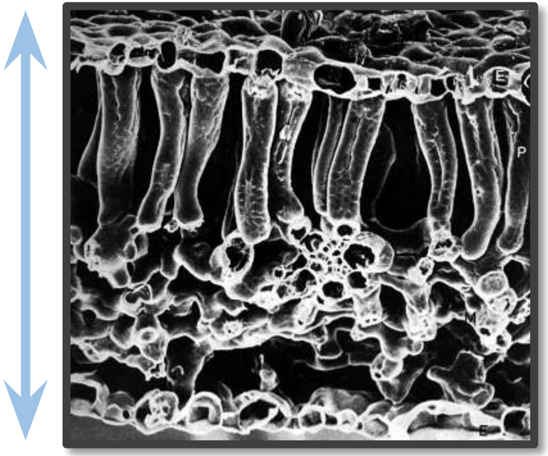




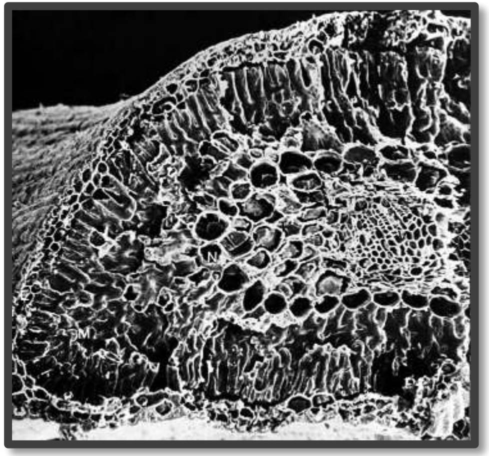


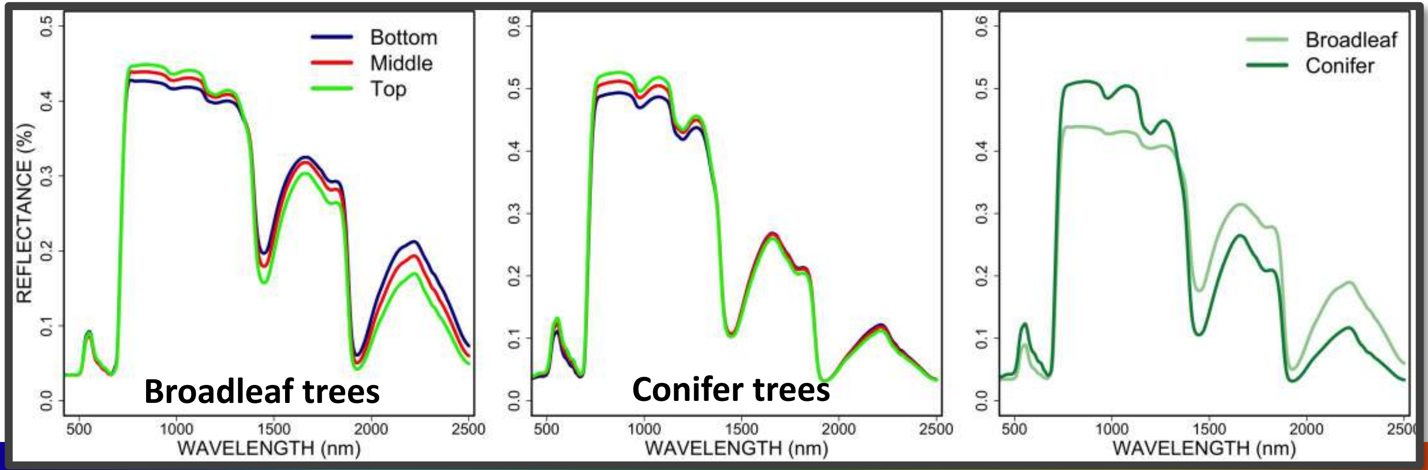
Leaf structure and spectra (leaf level)

Maple

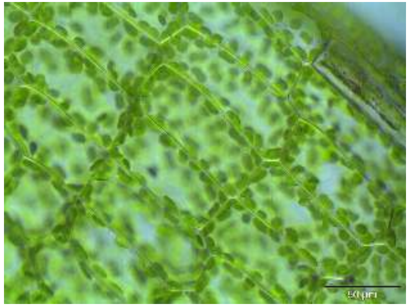


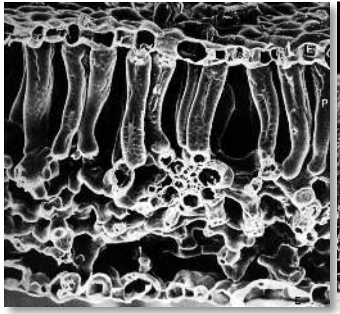
Pine

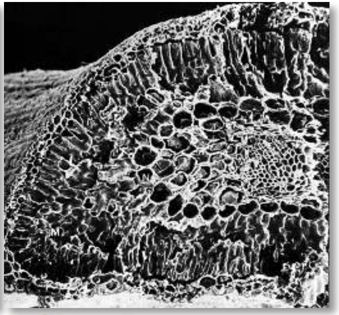




What are plants doing?
What’s different among plants?







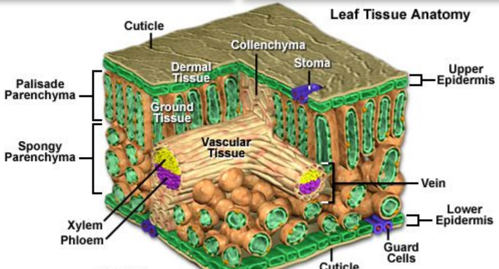
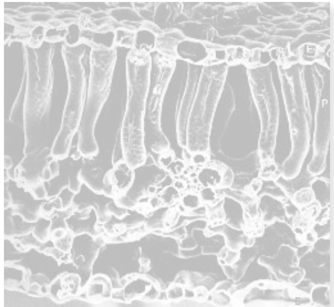
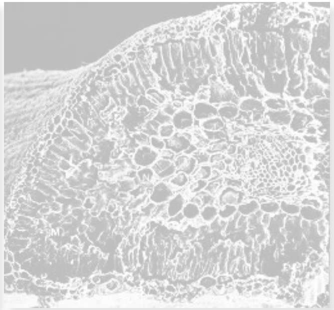


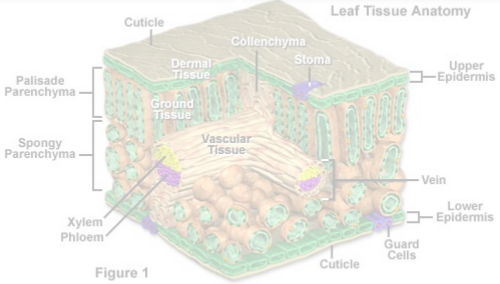
Figure 1

What are plants doing?
What's different among plants?

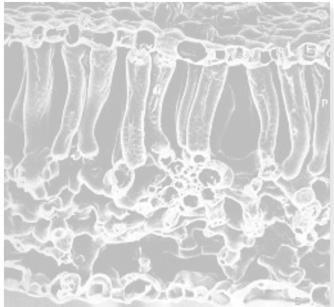
What are foliar functional traits and why do we care?

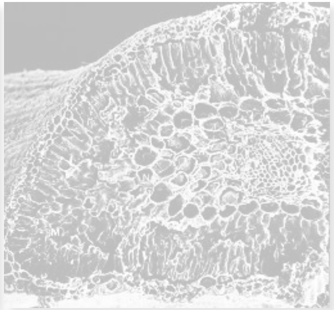


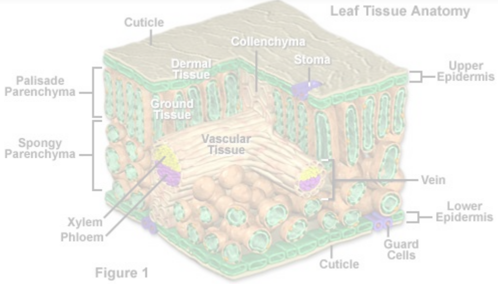


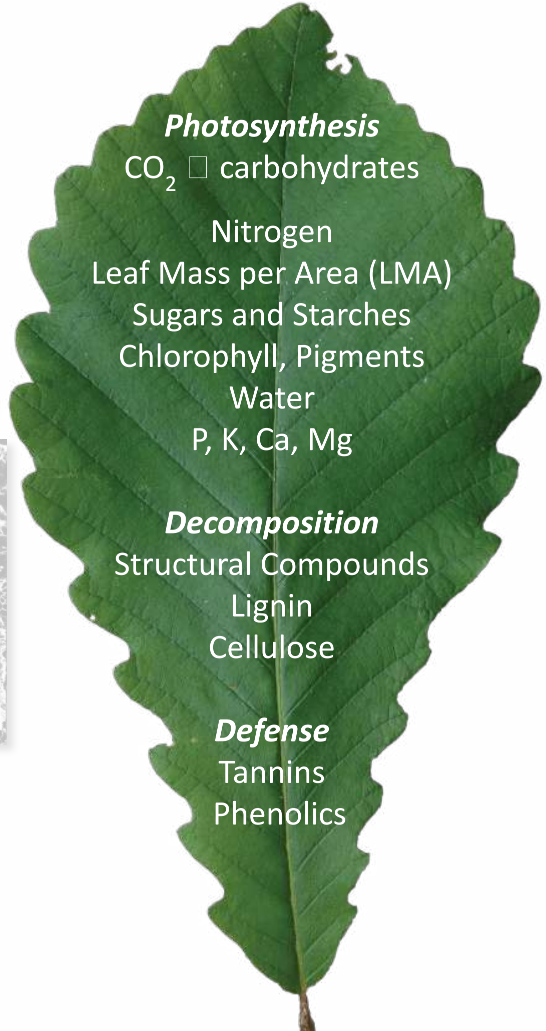


What are plants doing?
What's different among plants?

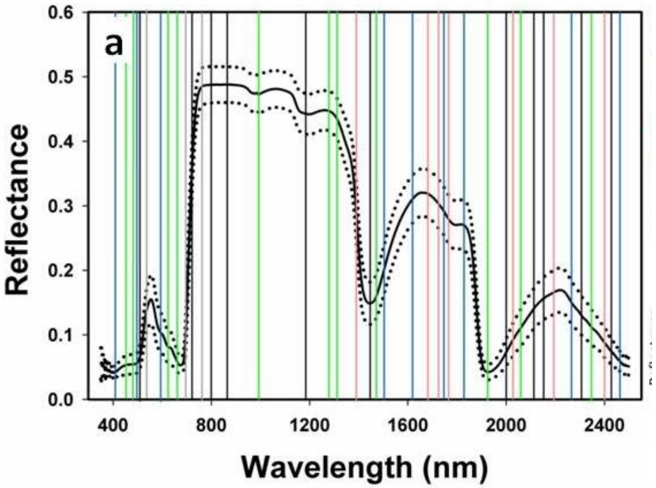




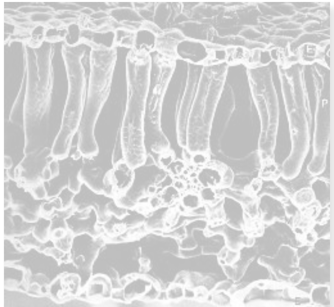


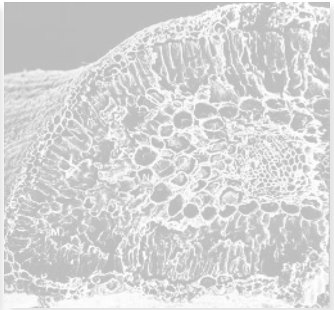


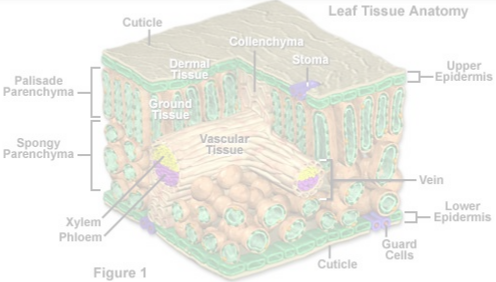
What are foliar functional traits and why do we care?

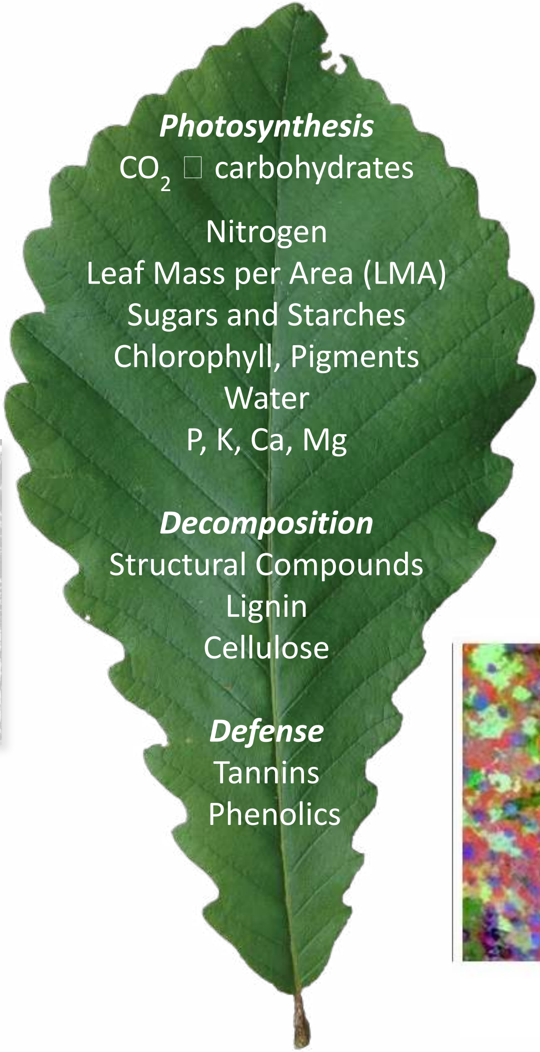


What are plants doing?
What's different among plants?









Photosynthesis

$CO_2 \rightarrow$ carbohydrates

Nitrogen

Leaf Mass per Area (LMA)

Sugars and Starches

Chlorophyll, Pigments

Water

P, K, Ca, Mg

Decomposition

Structural Compounds

Lignin

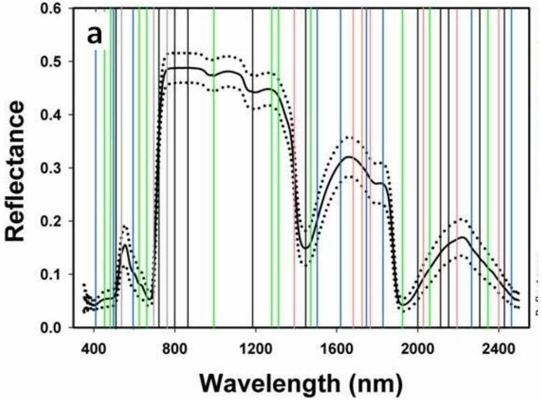
Cellulose

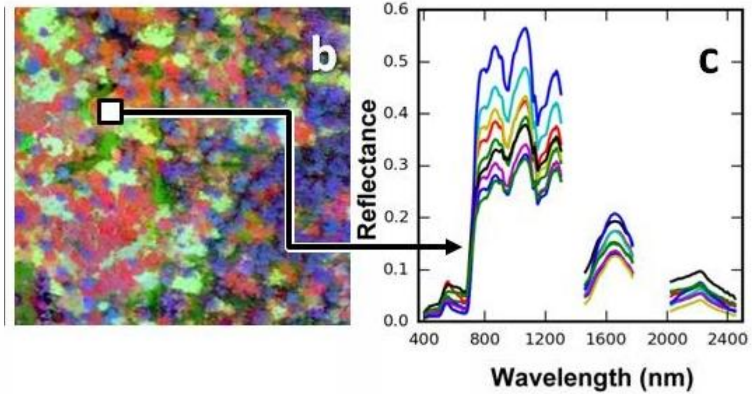
Defense

Tannins

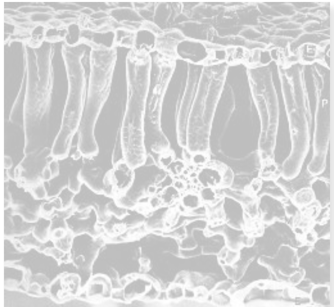
Phenolics

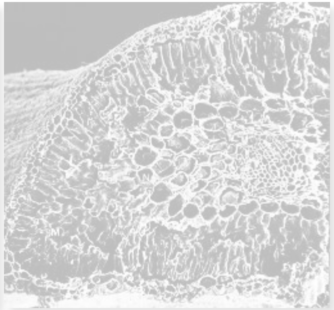
What are foliar functional traits and why do we care?

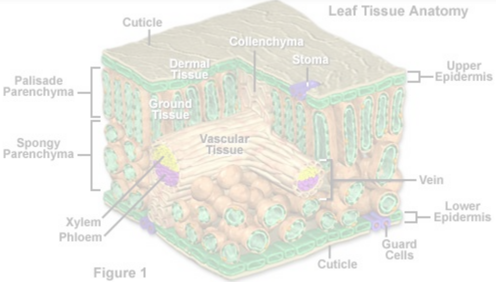




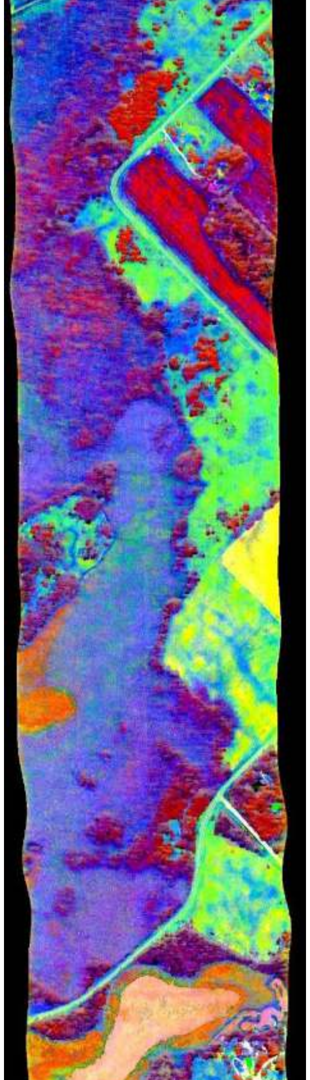
What are plants doing?
What's different among plants?







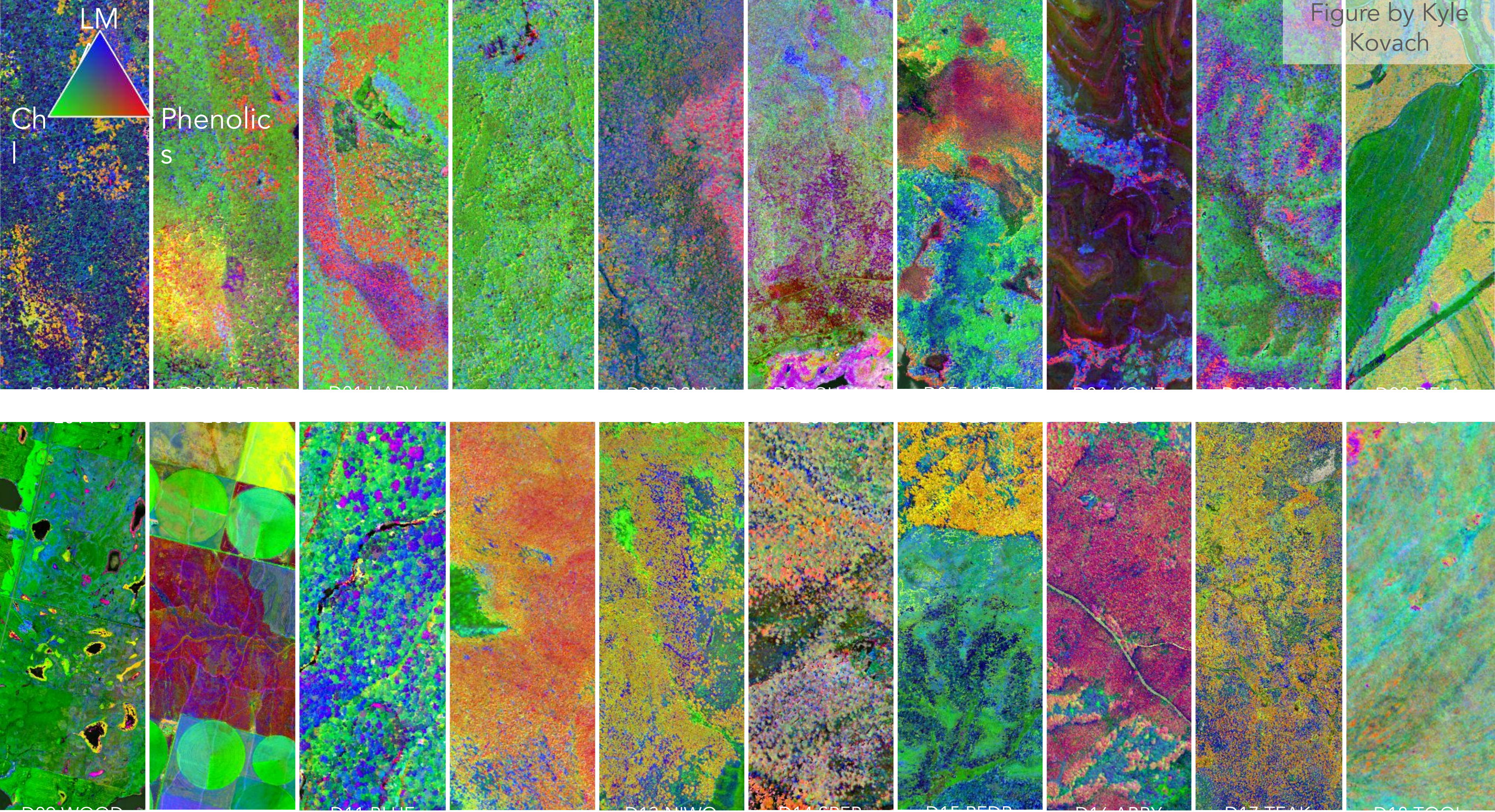
What are foliar functional traits
and why do we care?



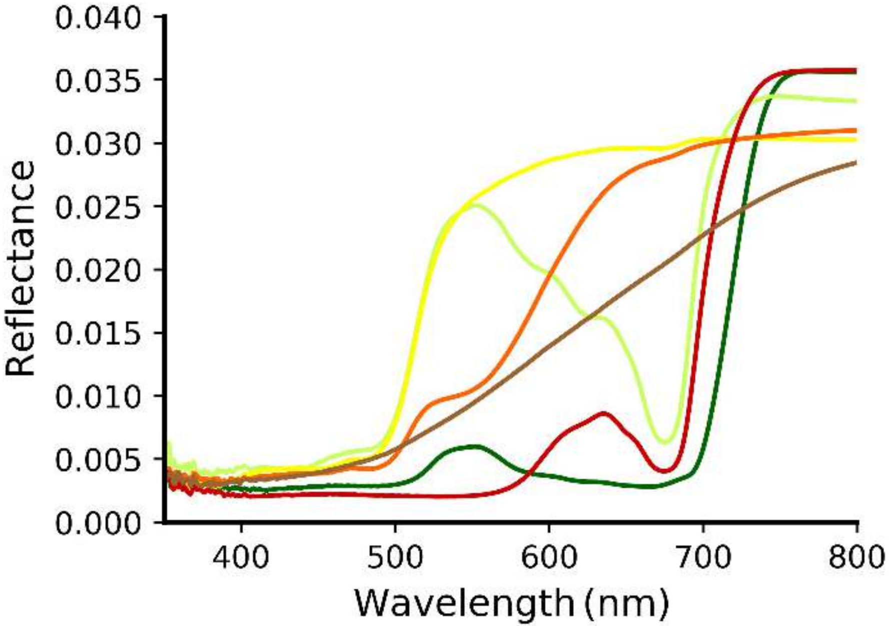
Single/mixed species
Homo/heterogeneous
Open/closed canopy

Wide climate range

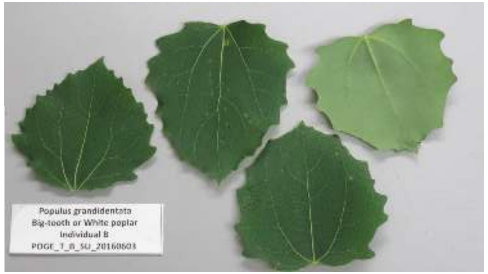
- Broadleaf tree
- Conifer tree
- Shrubland
- Grassland
- Forb


LM
Ch
l
Phenolic
s
Figure by Kyle
Kovach

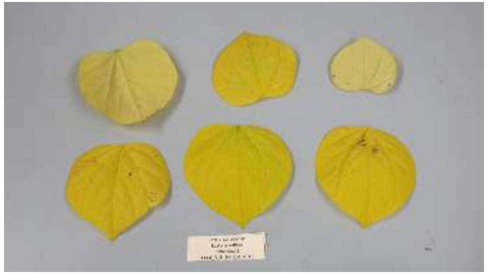
Colors

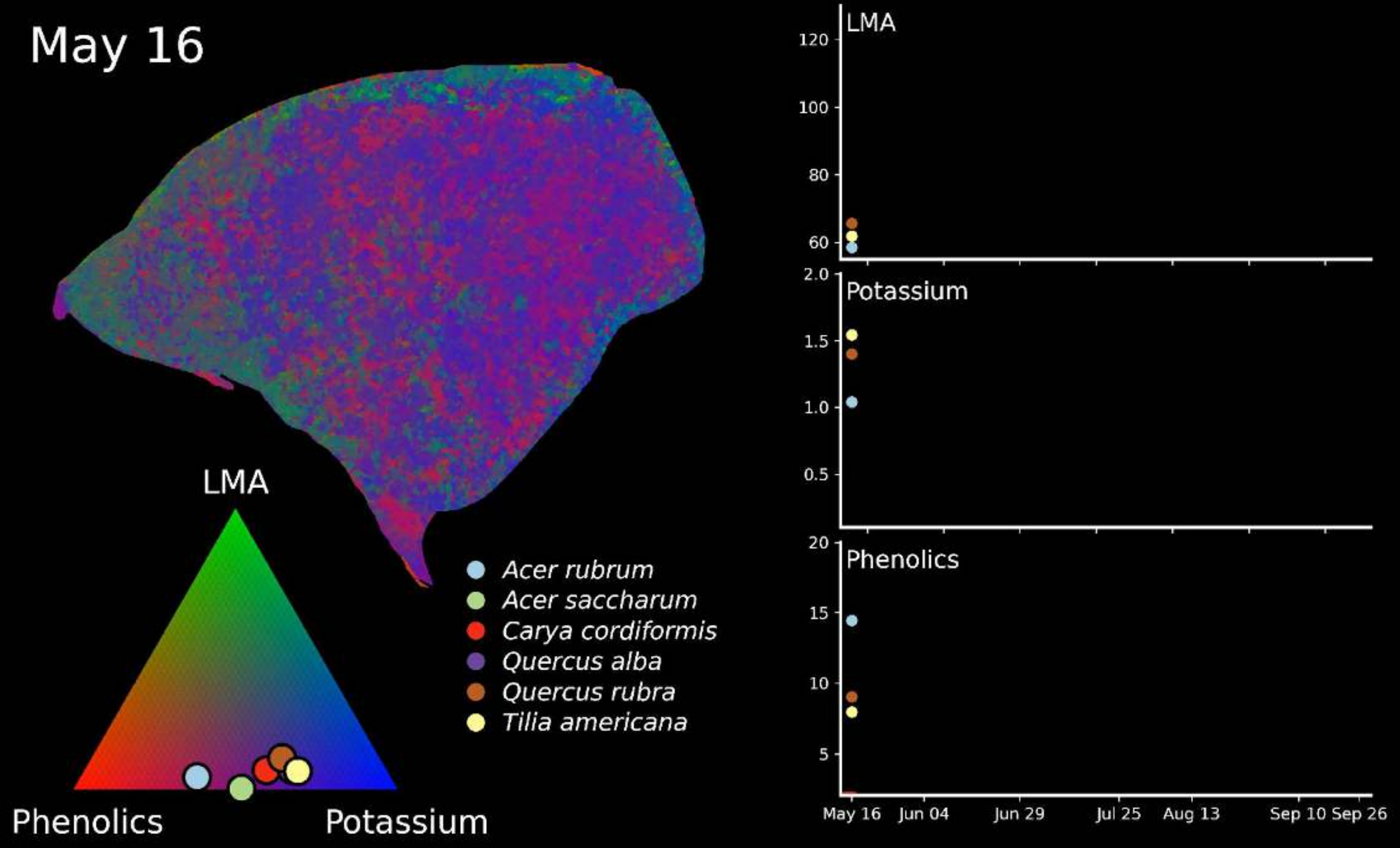


0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
Reflectance
400
500
600
700
800
Wavelength (nm)



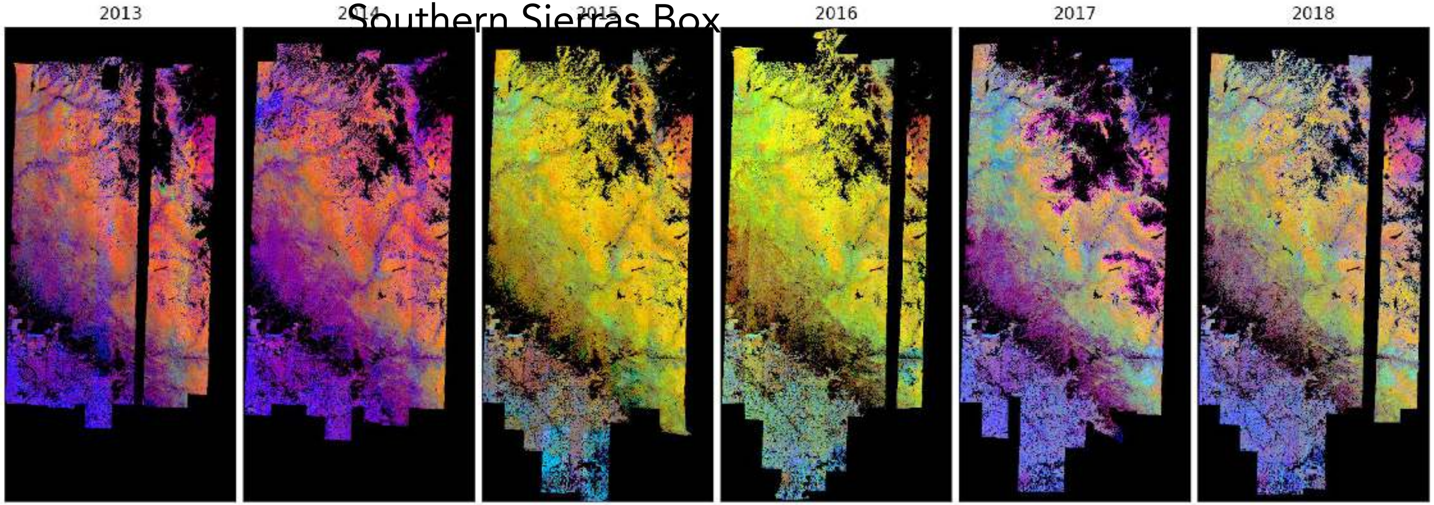
Populus grandidentata
Big-tooth or White poplar
Individual B
POGE_T_B_SU_20160603

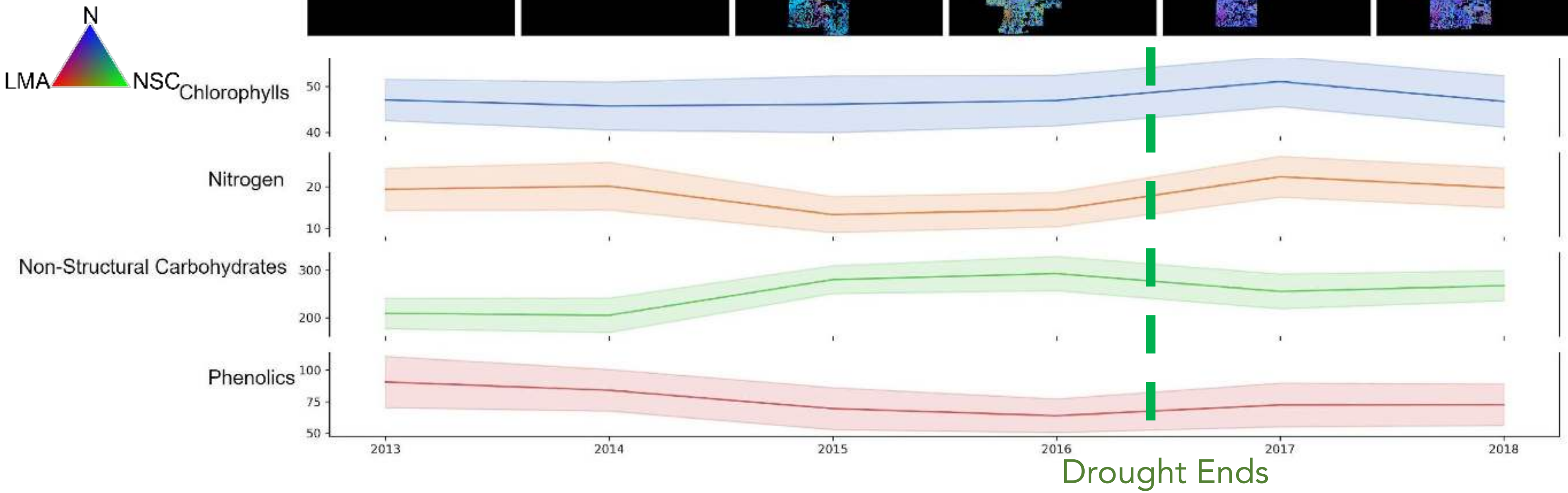



May 16
LMA
Phenolics
Potassium
Acer rubrum
Acer saccharum
Carya cordiformis
Quercus alba
Quercus rubra
Tilia americana
LMA
120
100
80
60
Potassium
2.0
1.5
1.0
0.5
Phenolics
20
15
10
5
May 16
Jun 04
Jun 29
Jul 25
Aug 13
Sep 10
Sep 26

Adam Chlus

Changes in Foliar Functional Traits from AVIRIS-Classic during the California Drought – Yosemite / Southern Sierras Box



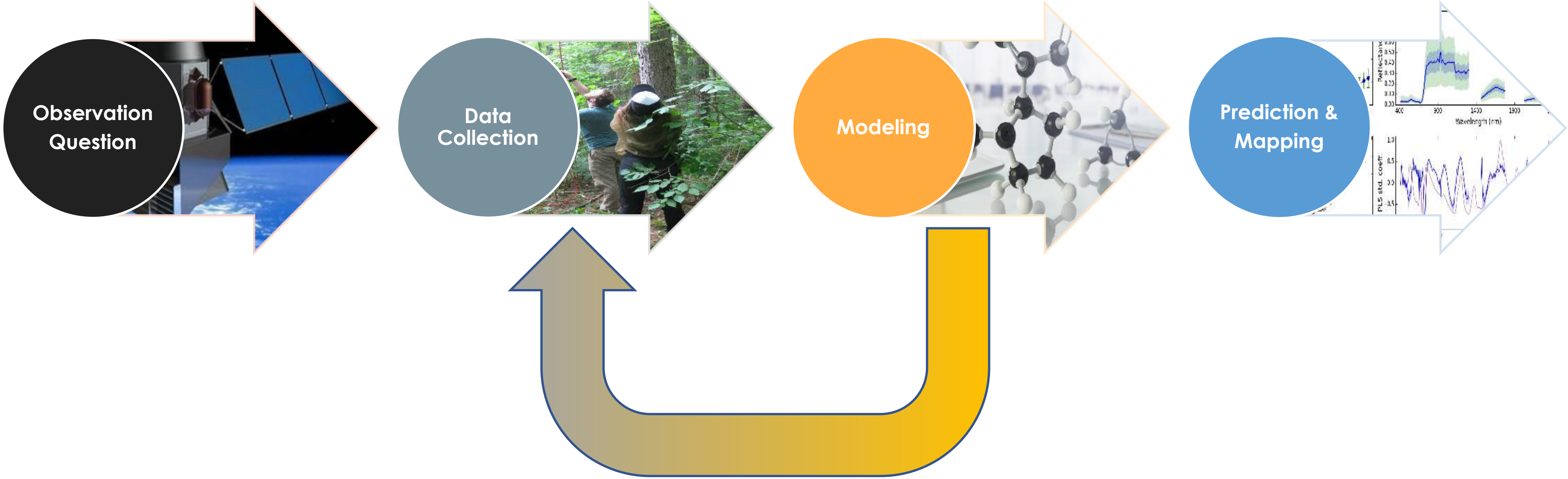


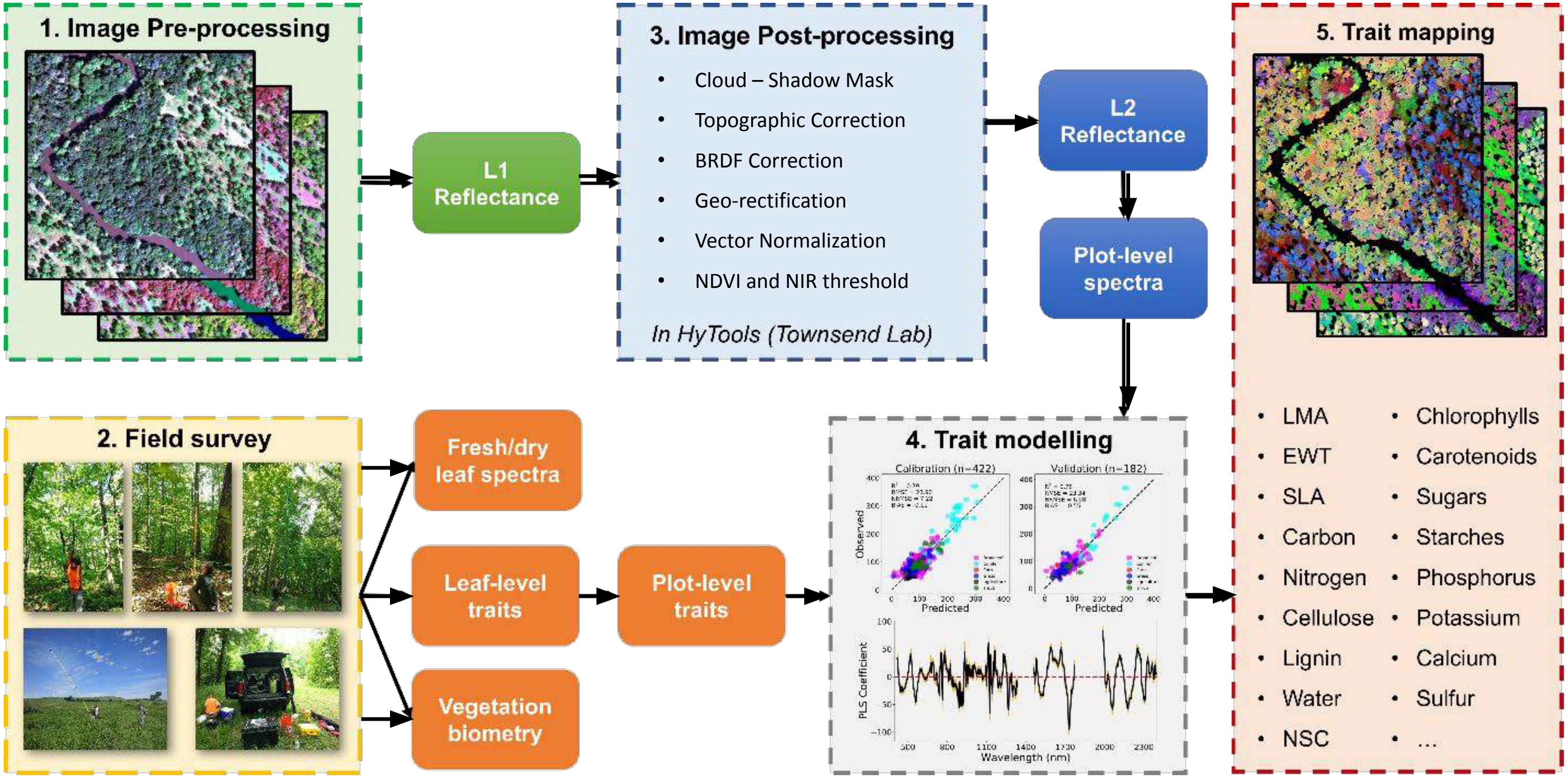

I HAVE QUESTIONS
LOTS OF QUESTIONS

How do we map traits from hyperspectral imagery?

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

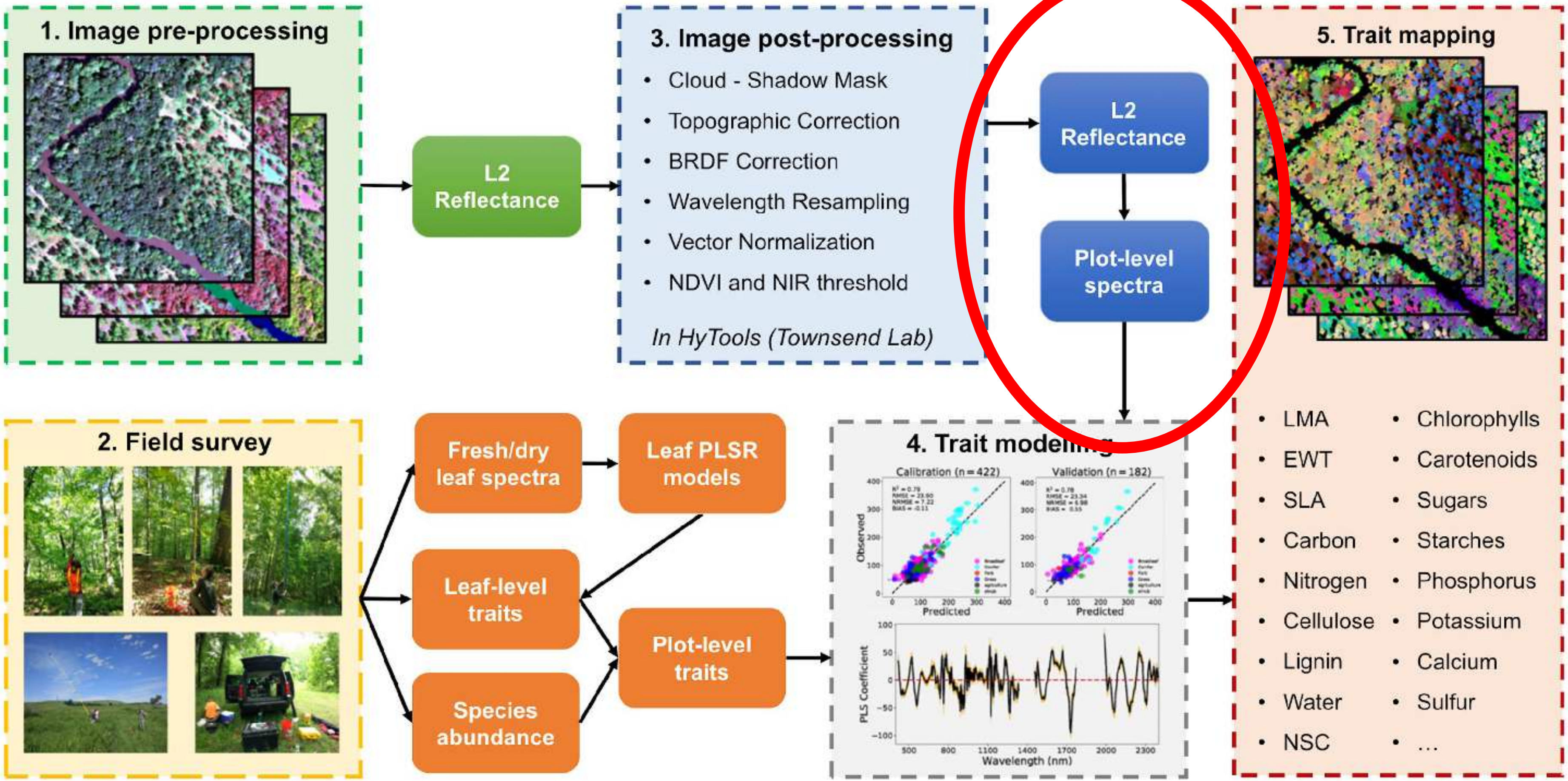
Designing Your Study Using Imaging Spectroscopy



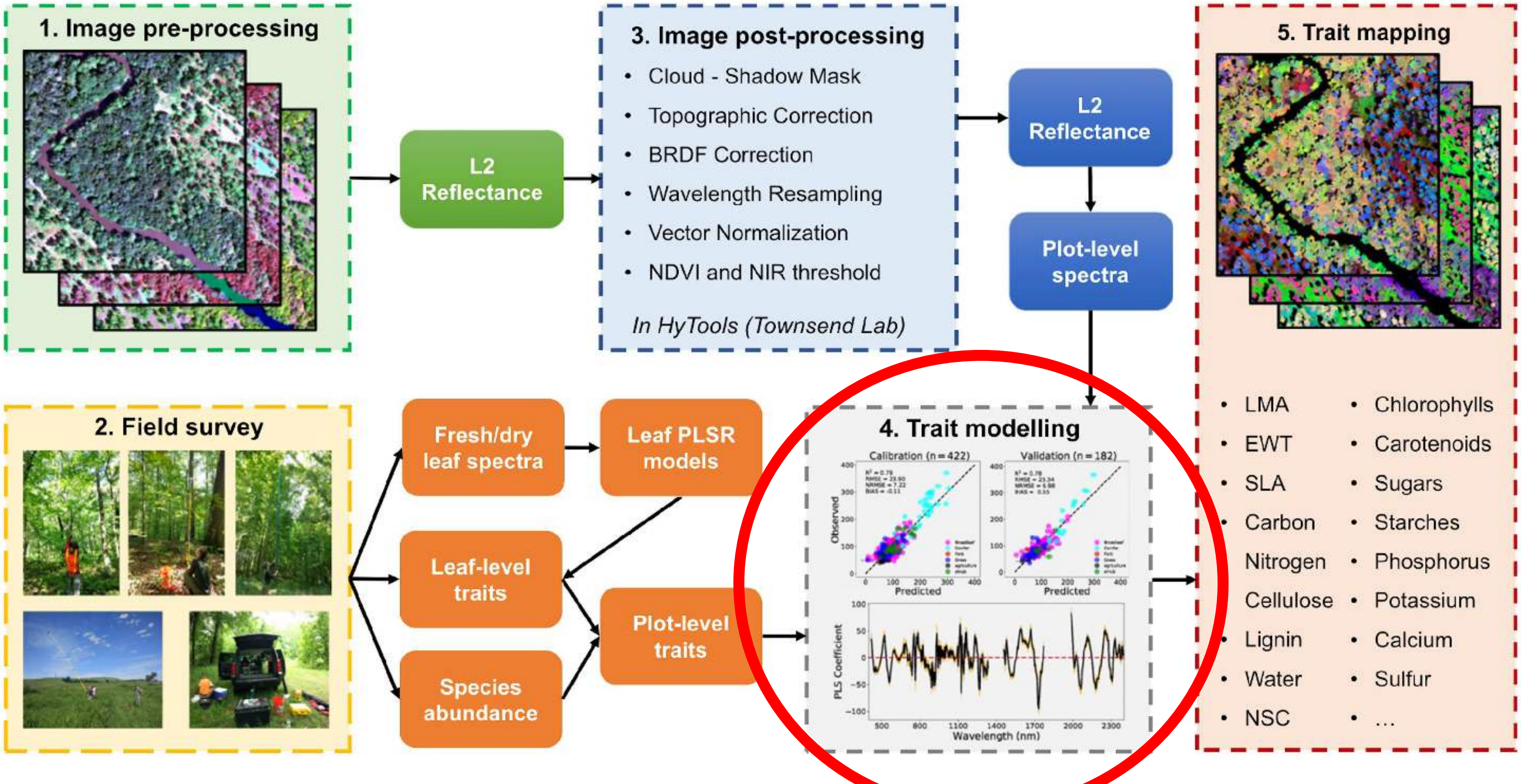


1. Image Pre-processing
L1 Reflectance
3. Image Post-processing
Cloud – Shadow Mask
Topographic Correction
BRDF Correction
Geo-rectification
Vector Normalization
NDVI and NIR threshold
In HyTools (Townsend Lab)
L2 Reflectance
Plot-level spectra
5. Trait mapping
2. Field survey
Fresh/dry leaf spectra
Leaf-level traits
Vegetation biometry
Plot-level traits
4. Trait modelling
Calibration (n=422)
Validation (n=182)
Observed
Predicted
PLS Coefficient
Wavelength (nm)
LMA
EWT
SLA
Carbon
Nitrogen
Cellulose
Lignin
Water
NSC
Chlorophylls
Carotenoids
Sugars
Starches
Phosphorus
Potassium
Calcium
Sulfur
…

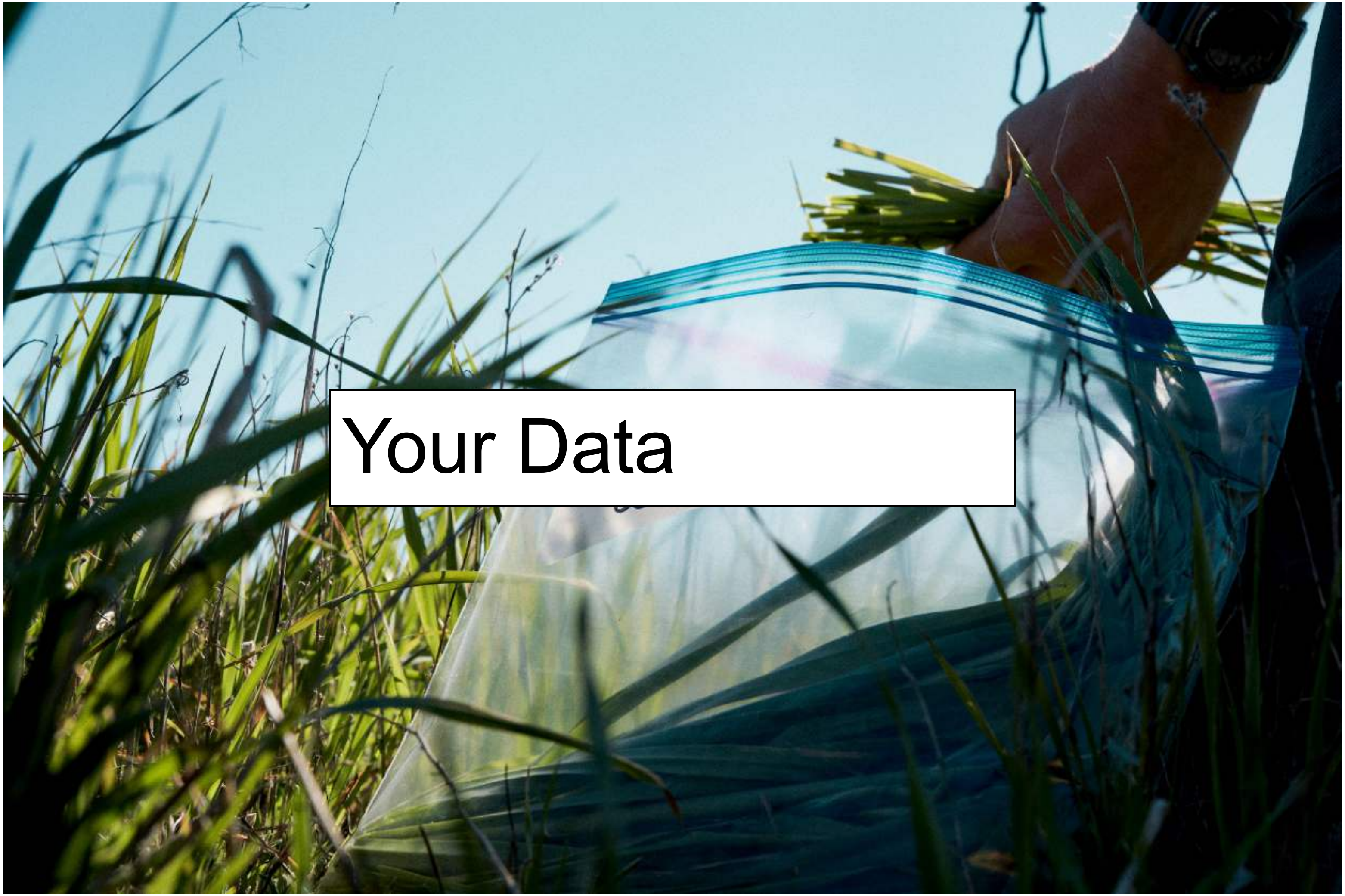
Modeling Overview



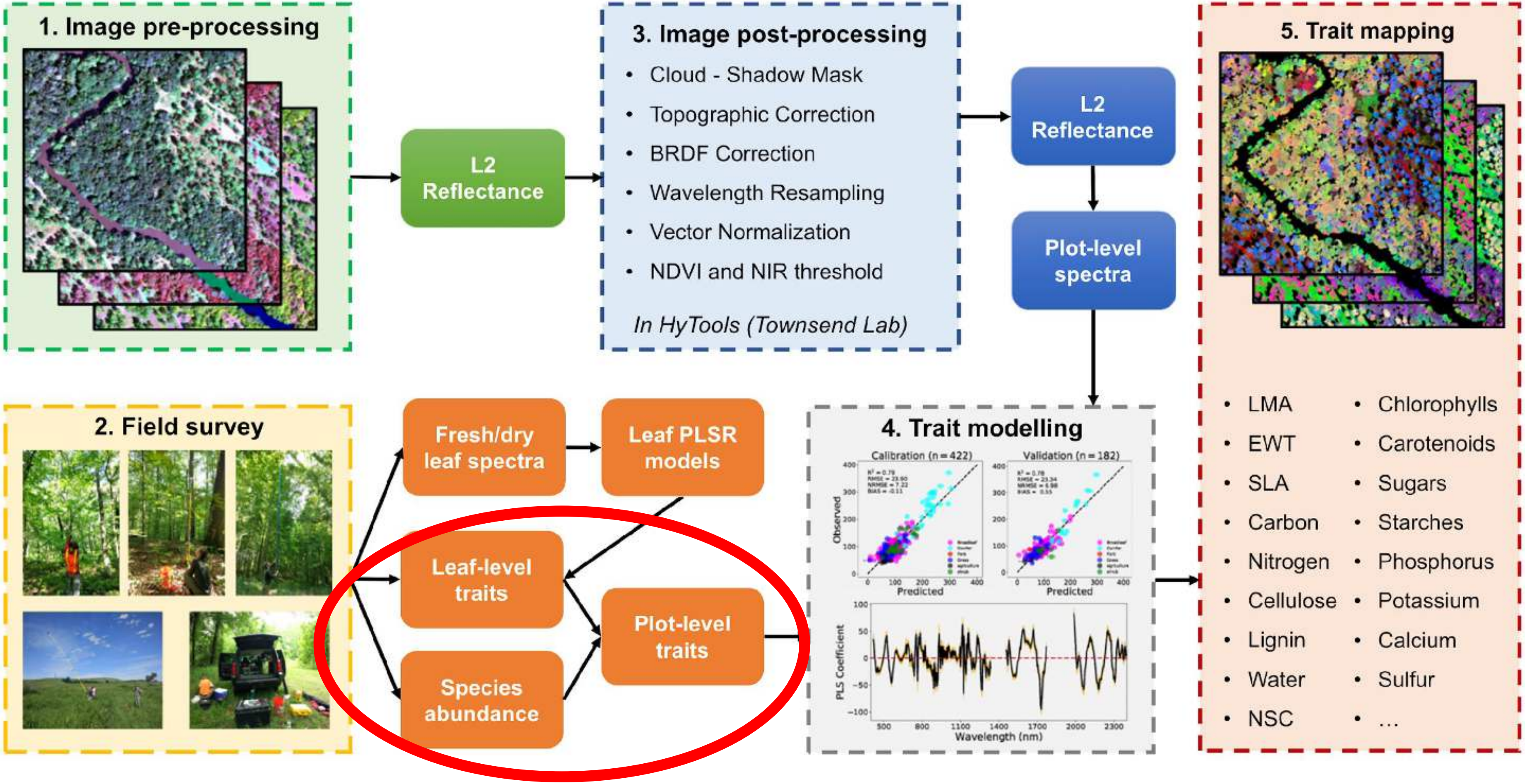
Modeling Overview



Your Data

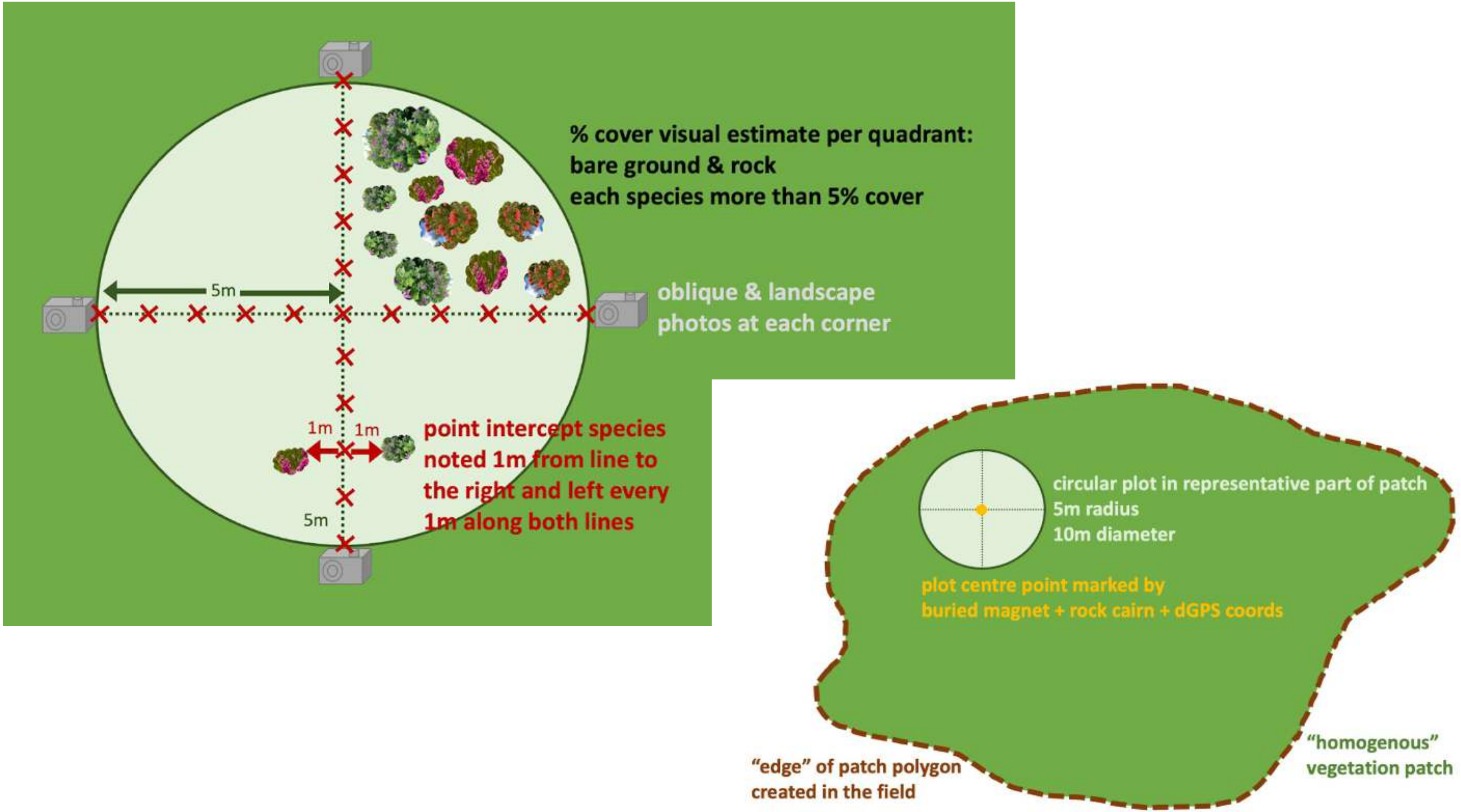


Community-Weighted Scaling



Plot Traits for Remote Sensing

- Community-weighted mean upscaling
  - Many possible definitions
  - Consider what the air/spaceborne sensor "sees"



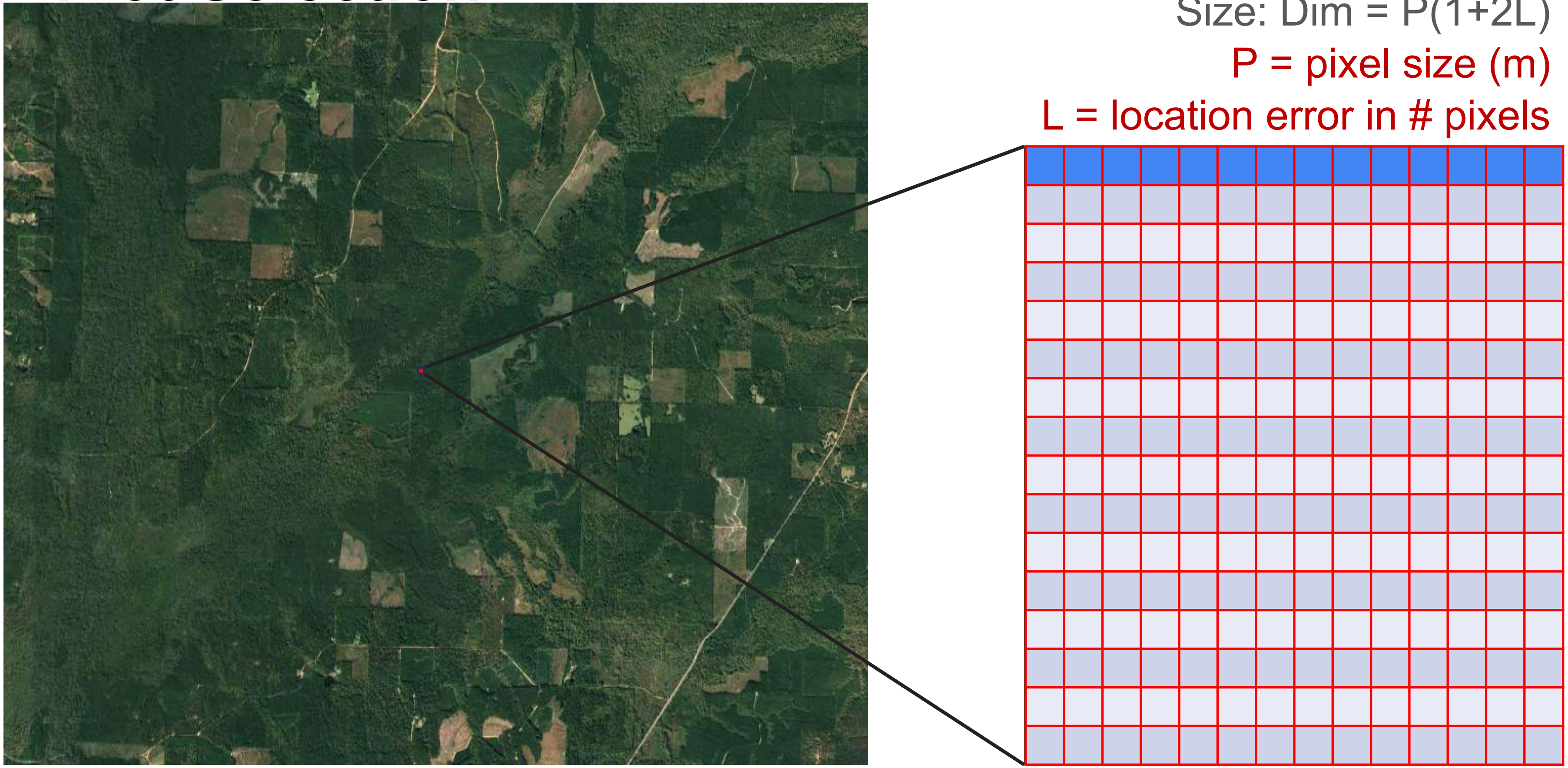
% cover visual estimate per quadrant:
bare ground & rock
each species more than 5% cover
5m
oblique & landscape
photos at each corner
1m
1m
point intercept species
noted 1m from line to
the right and left every
1m along both lines
5m
circular plot in representative part of patch
5m radius
10m diameter
plot centre point marked by
buried magnet + rock cairn + dGPS coords
"edge" of patch polygon
created in the field
"homogenous"
vegetation patch

Plot Selection

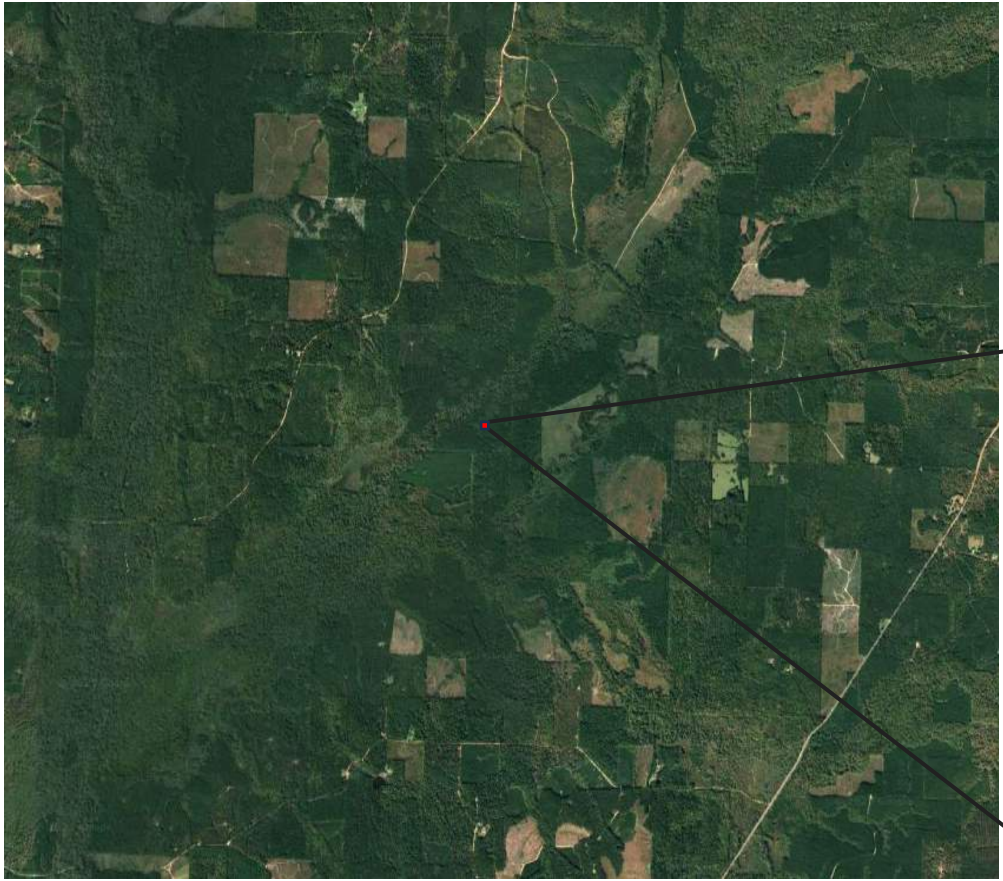
Size: Dim = P(1+2L)

P = pixel size (m)

L = location error in # pixels

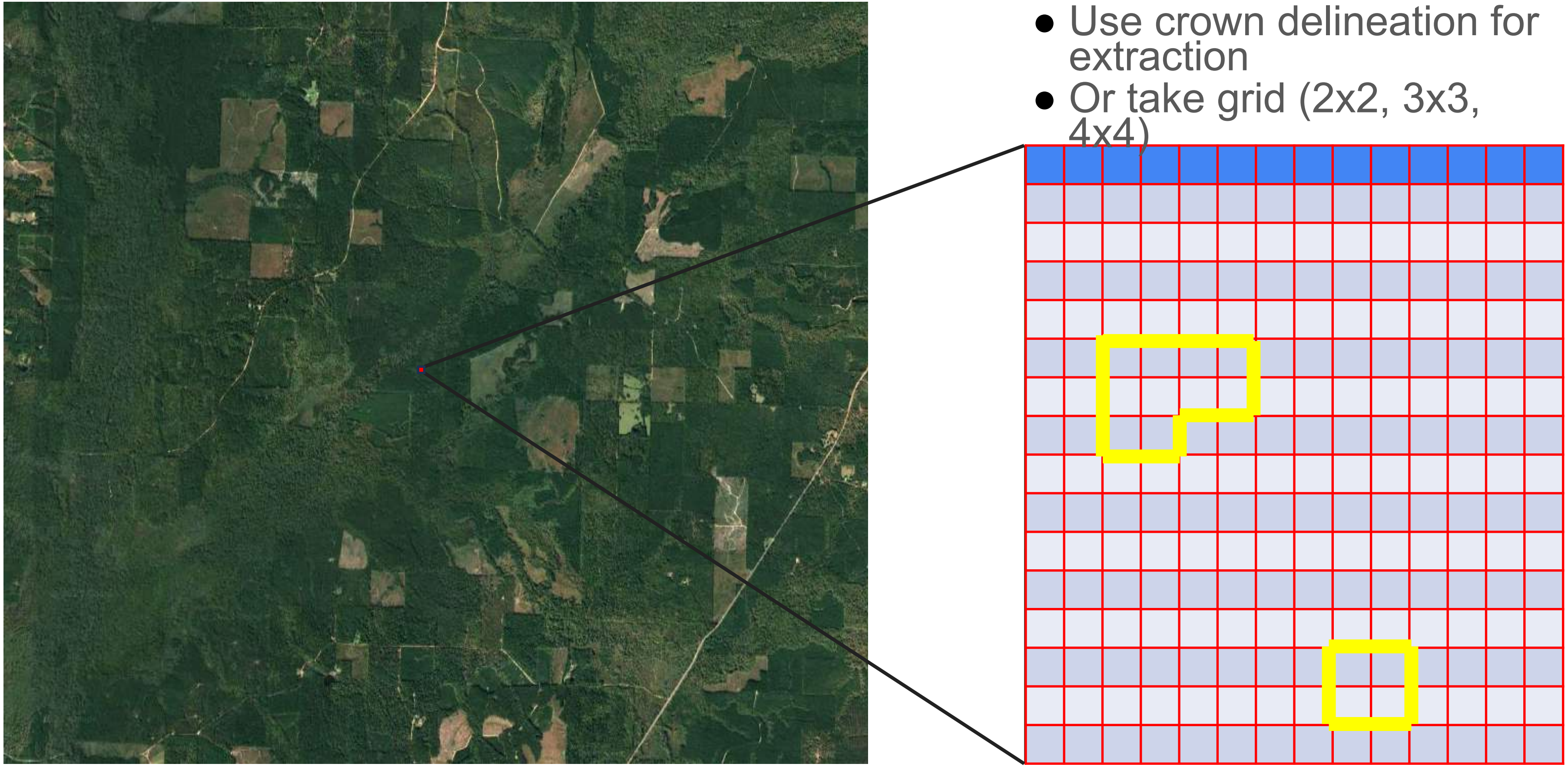


Plot Selection



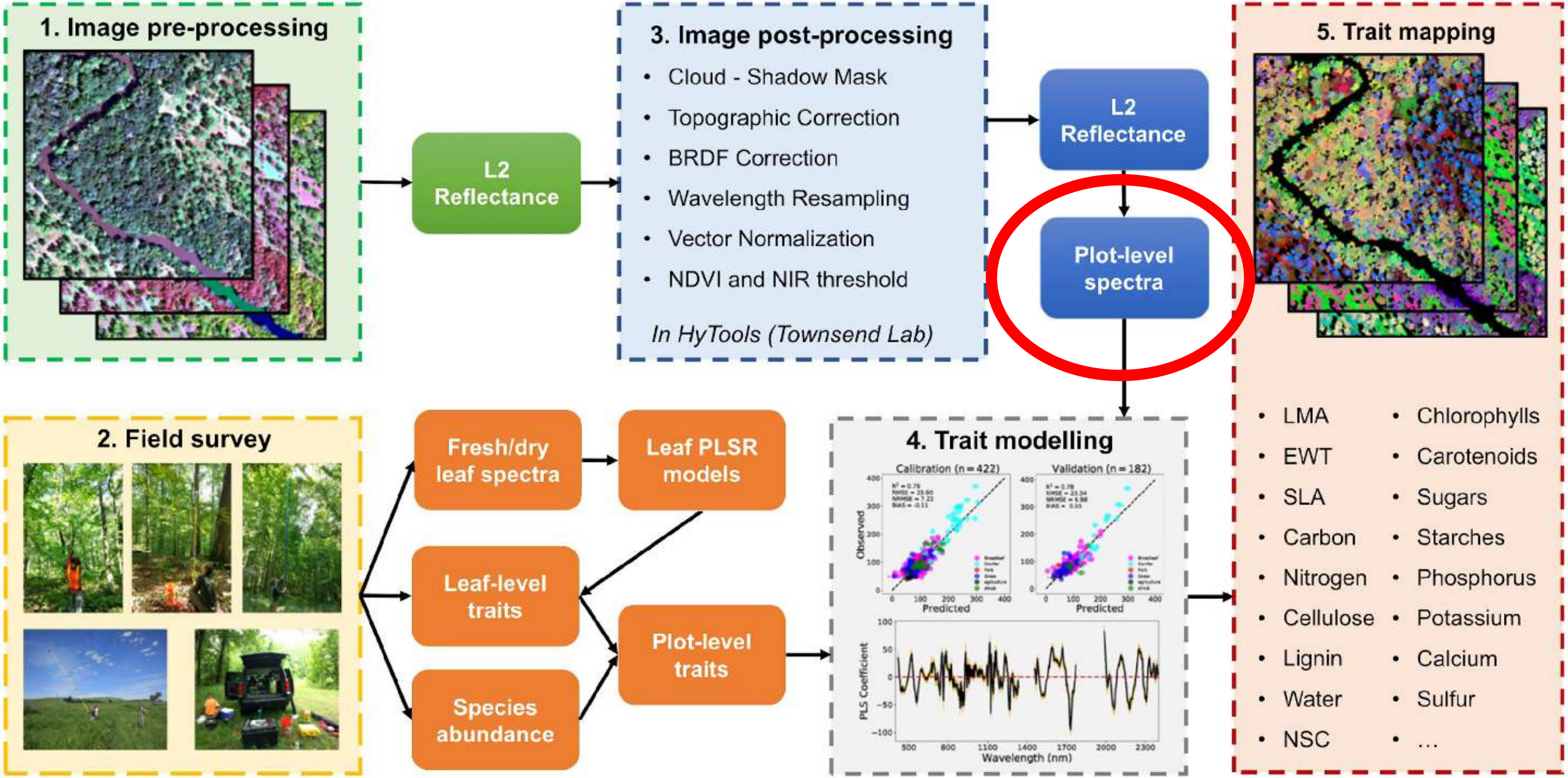
- Plot delineation
- GPS point for center of plot
- Single crown = one trait value
- Multiple species = community weighted mean
- All species with cover greater than 5% adding to at least 80% of total cover

Extract Pixels



- Use crown delineation for extraction
- Or take grid (2x2, 3x3, 4x4)

Modeling Overview



Creating Plot Spectra

- Create plot spectra using one of two methodologies

1. Average pixels (to create non-existent pixel)
2. Sample from pixels within plot

DON'T PSUEDOREPLICATE!

2340.417	2345.426	2350.436	2355.446	2360.456	2365.465	2370.475	2375.485	2380.495	2385.504	2390.514	2395.524	plotname	site	lon	lat
308	294	312	301	303	270	290	279	276	258	271	260	CHEQ01_PIBA	CHEQ	-90.0727	45.82754
131	121	138	131	134	129	155	127	130	130	134	128	CHEQ01_PIBA	CHEQ	-90.0727	45.82753
399	348	352	372	364	329	346	346	332	317	360	345	CHEQ01_PIBA	CHEQ	-90.0727	45.82753
515	543	492	516	521	511	489	467	444	441	451	461	CHEQ01_PIBA	CHEQ	-90.0727	45.82753
445	418	447	432	465	437	456	438	398	396	399	398	CHEQ01_PIBA	CHEQ	-90.0727	45.82753
136	126	140	144	158	134	177	157	127	137	126	129	CHEQ01_PIBA	CHEQ	-90.0727	45.82752
339	347	347	356	349	312	340	317	310	326	304	277	CHEQ01_PIBA	CHEQ	-90.0727	45.82752
345	354	319	338	332	326	310	296	314	298	280	311	CHEQ01_PIBA	CHEQ	-90.0727	45.82752
423	419	400	390	407	388	406	376	352	332	340	384	CHEQ01_PIBA	CHEQ	-90.0727	45.82752
110	127	105	89	89	110	117	119	122	77	116	105	CHEQ01_PIBA	CHEQ	-90.0727	45.82751
182	199	176	190	172	158	185	155	173	172	146	160	CHEQ01_PIBA	CHEQ	-90.0727	45.82751
108	106	116	113	95	102	92	136	68	78	95	122	CHEQ01_PIRE	CHEQ	-90.0767	45.82908
84	81	81	97	70	85	87	79	76	89	76	59	CHEQ01_PIRE	CHEQ	-90.0767	45.82908
153	148	155	152	146	110	122	140	99	115	110	150	CHEQ01_PIRE	CHEQ	-90.0767	45.82907
97	106	104	94	97	91	130	105	94	91	68	90	CHEQ01_PIRE	CHEQ	-90.0767	45.82907
82	83	105	85	103	86	86	104	86	82	97	99	CHEQ01_PIRE	CHEQ	-90.0767	45.82907
271	270	257	250	245	242	279	202	224	222	221	202	CHEQ01_PIRE	CHEQ	-90.0767	45.82906
135	132	133	145	115	115	164	123	108	160	122	118	CHEQ01_PIRE	CHEQ	-90.0767	45.82906
140	134	94	117	110	108	140	120	94	81	101	103	CHEQ01_PIRE	CHEQ	-90.0767	45.82906
307	263	264	261	257	255	261	239	233	226	225	240	CHEQ02_POTR	CHEQ	-90.0729	45.8285
226	203	181	198	213	199	203	182	161	155	140	161	CHEQ02_POTR	CHEQ	-90.0729	45.8285
258	238	217	231	213	219	186	229	194	204	215	199	CHEQ02_POTR	CHEQ	-90.0729	45.82849
216	231	183	196	204	224	190	226	187	170	175	179	CHEQ02_POTR	CHEQ	-90.0729	45.82849
399	385	376	376	364	343	355	331	318	297	288	297	CHEQ02_POTR	CHEQ	-90.0729	45.82849
255	236	232	230	224	180	209	201	190	187	206	189	CHEQ02_POTR	CHEQ	-90.0729	45.82849
253	245	231	247	220	240	255	218	210	180	204	194	CHEQ02_POTR	CHEQ	-90.0729	45.82849
409	393	381	390	366	340	337	344	332	334	347	309	CHEQ02_POTR	CHEQ	-90.0729	45.82849
385	383	383	396	381	354	363	349	310	305	323	316	CHEQ02_POTR	CHEQ	-90.0729	45.82848
379	362	368	372	349	351	351	337	320	298	313	287	CHEQ02_POTR	CHEQ	-90.0729	45.82848
425	397	382	382	378	363	387	341	358	360	354	327	CHEQ02_POTR	CHEQ	-90.0729	45.82848
376	345	364	382	357	336	369	340	339	342	308	291	CHEQ02_POTR	CHEQ	-90.0729	45.82848
372	369	382	333	333	340	348	321	287	292	299	288	CHEQ02_POTR_low	CHEQ	-90.0903	45.82448

Avg Plot CHEQ01_PIBA

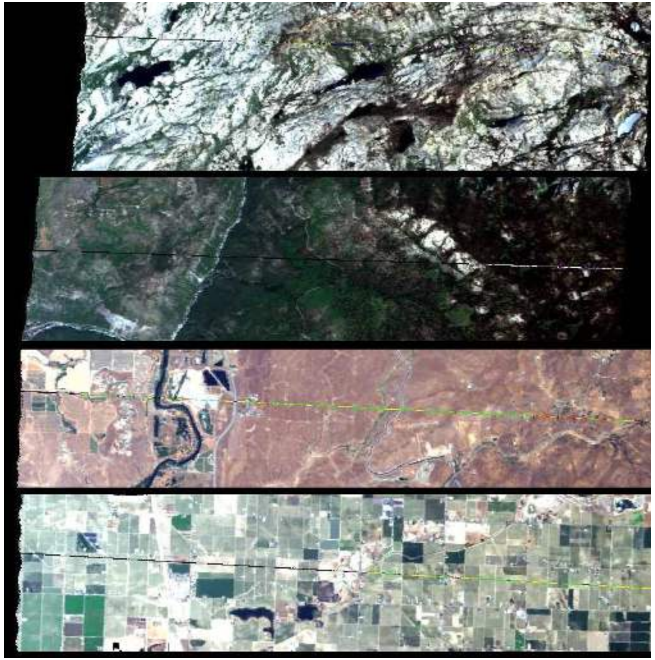
Avg Plot CHEQ01_PIRE

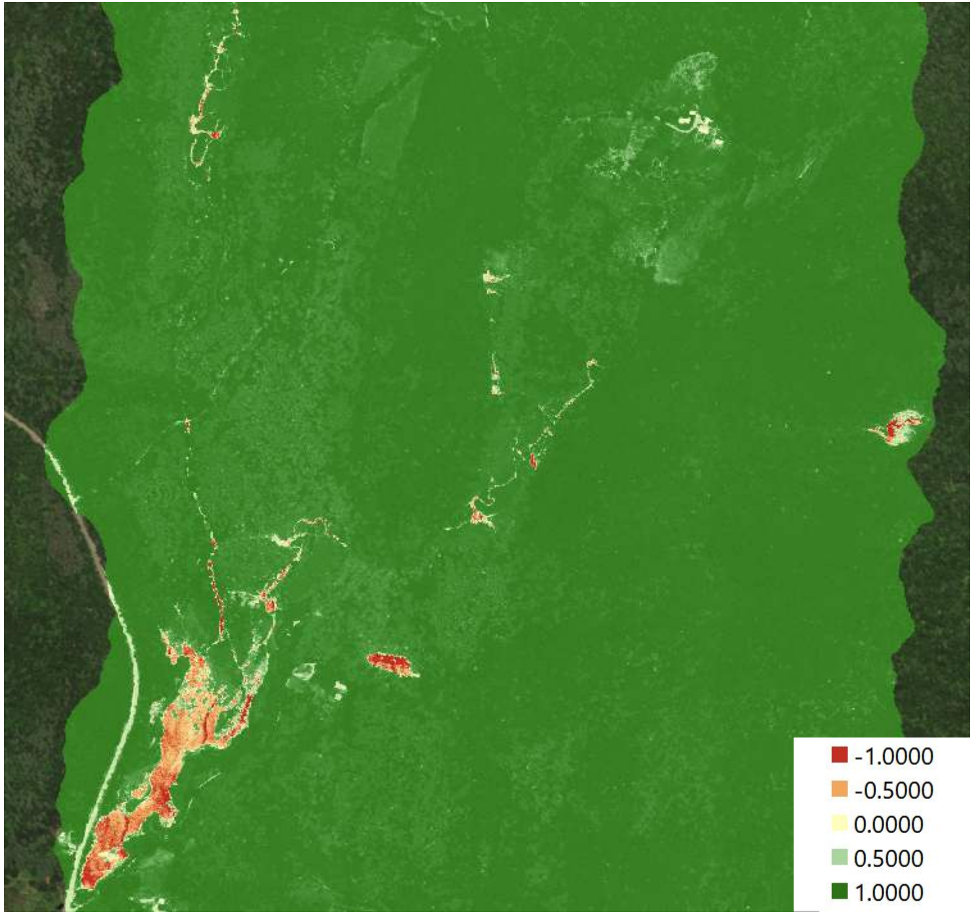
Avg Plot CHEQ02_POTR

Data Quality

- Refining data to build models

1. Assess outliers
2. Remove non-veg pixels with NDVI threshold
3. Remove potential shadow pixels with NIR threshold





Modeling

```python
def build_model(trait_name, n_outers, n_inners):
    init_transforms = [htt.LabelBasedSelector(column_name='pix_pos', column_labels='Center'), ## need missing_ok arg
                       htt.SubSampler(),
                       htt.WavelengthSelector(wanted_ranges=[(419.9, 1315.1), (1464.9, 1765.1), (1989.9, 2395.1)]),
                       htt.UnitMagnitudeNormalizer()]
    init_transform = htt.SpectrumDataTransformSequence(init_transforms)
    dataset = htt.SpectrumFrameDataset(data_csv=clean_dir/f'{trait_name}.csv',
                                       transform=init_transform)

    splitter = htt.SpectrumFrameDatasetSplitter(outer_params=htt.SPLIT_PARAMS_SHUFFLE(n_splits=n_outers, percent=80),
                                                calib_inner_params=htt.SPLIT_PARAMS_SHUFFLE(n_splits=n_inners, percent=80),
                                                deploy_inner_params=htt.SPLIT_PARAMS_ALLINONE(percent=100))
    split_data = splitter(sample_ids=dataset.sample_data()[0],
                          sample_labels=dataset.sample_data()[1])

    start_time = time.perf_counter()
    oi_n_comps, oi_rmses, oi_r2s, oi_biases = [], [], [], []
    oi_mnrmses, oi_qnrmses, oi_rnr

    for oi in range(n_outers):
        calib_train_transform = ht
        calib_train_dataloader = h
                                                    split_data=split_data[oi].all_calib_trains(),
                                                    transform=calib_train_transform)

        calib_valid_transform = htt.SpectrumDataTransformSequence([])
        calib_valid_dataloader = htt.SpectrumFrameDatasetDataLoader(dataset=dataset,
                                                                    split_data=split_data[oi].all_calib_valids(),
                                                                    transform=calib_valid_transform)

        deploy_train_transform = htt.SpectrumDataTransformSequence([])
        deploy_train_dataloader = htt.SpectrumFrameDatasetDataLoader(dataset=dataset,
                                                                     split_data=split_data[oi].all_deploy_trains(),
                                                                     transform=deploy_train_transform)

        test_transform = htt.SpectrumDataTransformSequence([])
        test_dataloader = htt.SpectrumFrameDatasetDataLoader(dataset=dataset,
                                                             split_data=split_data[oi].all_tests(),
                                                             transform=test_transform)
```

Methods

Surv Geophys (2019) 40:589–629
https://doi.org/10.1007/s10712-018-9478-y

CrossMark

Quantifying Vegetation Biophysical Variables from Imaging Spectroscopy Data: A Review on Retrieval Methods

Jochem Verrelst[1] · Zbyněk Malenovský[2,3,4] · Christiaan Van der Tol[5] · Gustau Camps-Valls[1] · Jean-Philippe Gastellu-Etchegorry[6] · Philip Lewis[7,8] · Peter North[9] · Jose Moreno[1]

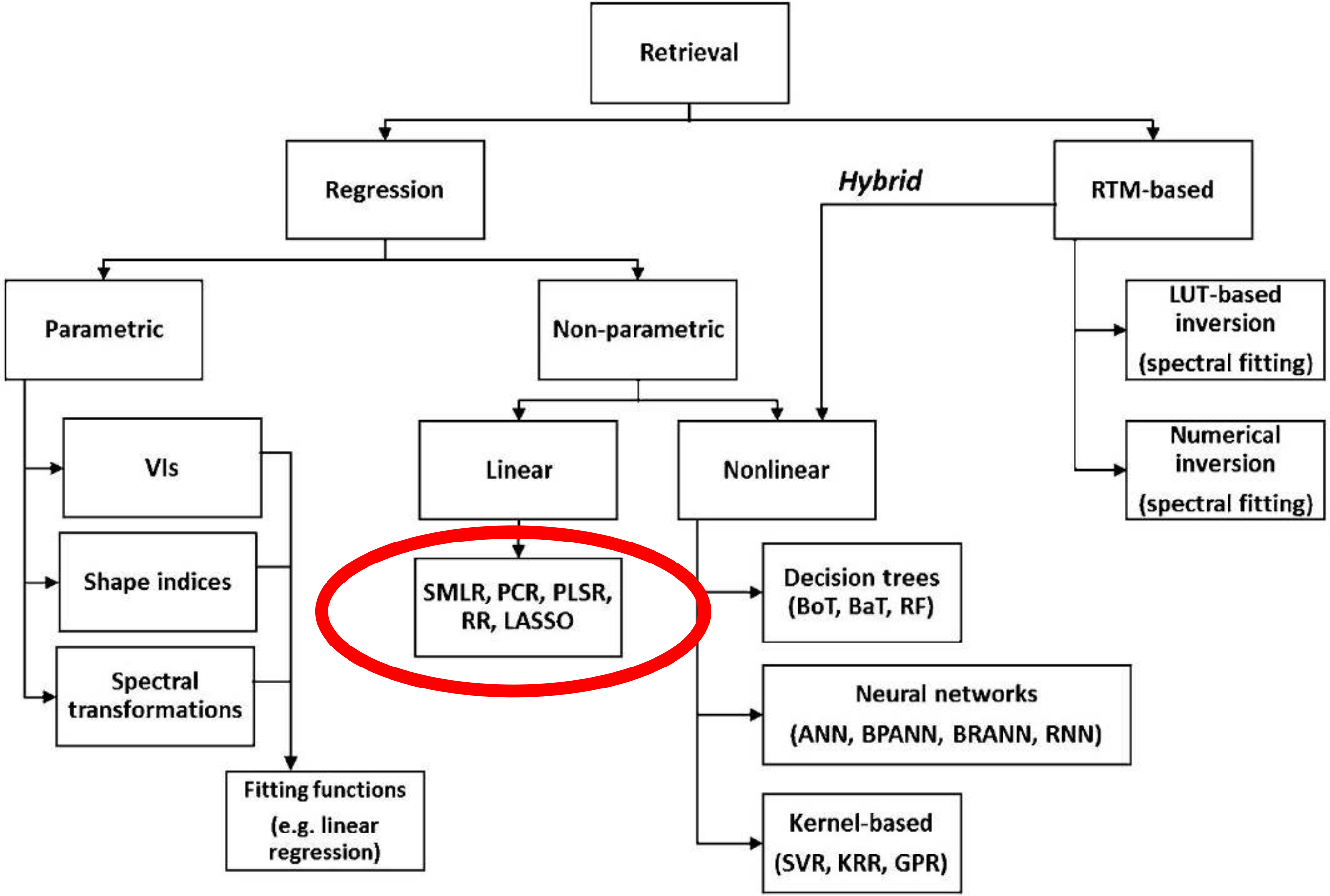
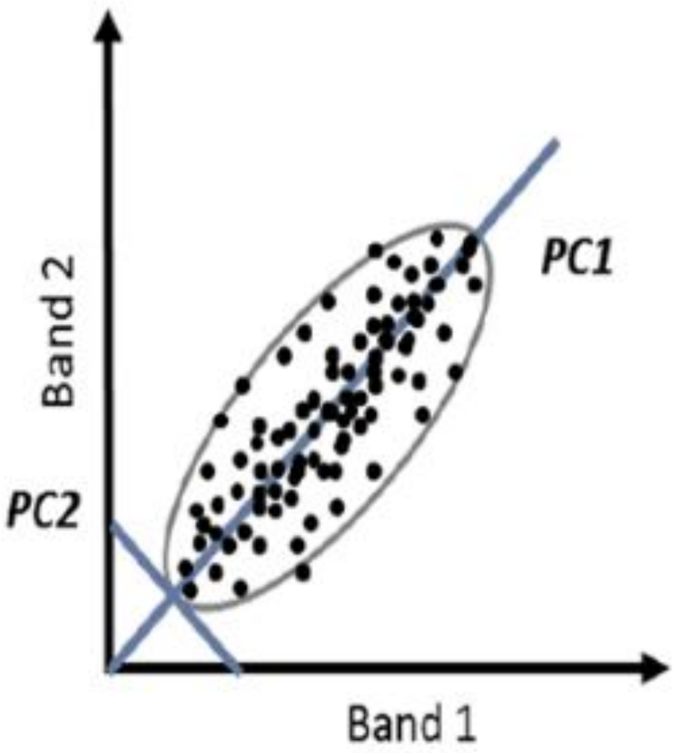


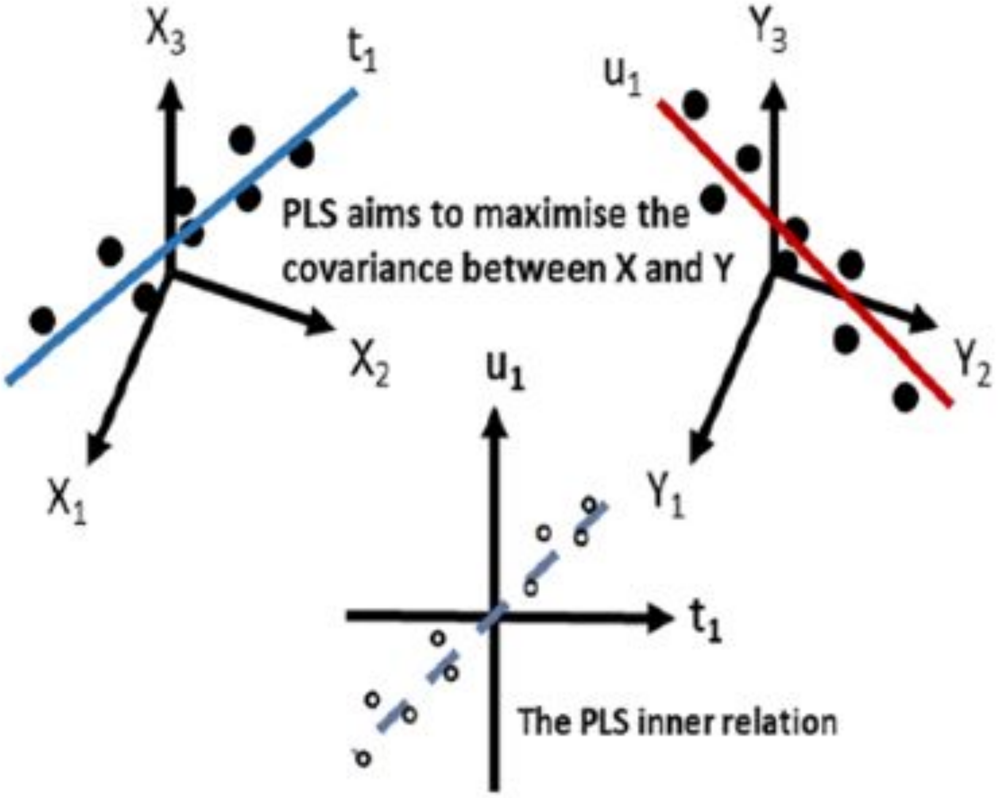
Fig. 9 Schematic overview of the main retrieval methods

Verrelst et al. 2019

PCA (a)


PC1
PC2
Band 2
Band 1

PLS (b)


X_3
t_1
X_2
X_1
PLS aims to maximise the covariance between X and Y
u_1
Y_3
Y_2
Y_1
u_1
t_1
The PLS inner relation

PLS: Projection to Latent Structures, a.k.a.
PLSR: Partial-Least-Squares Regression

Similar to PCA, but…

- maximizes covariance, instead of minimizing correlation
- incorporates the response variable, not just the predictors

Unlike OLS regression, does not assume predictors are error-free
Similar to Multiple Linear Regression, but handles predictor collinearity
□ able to handle many predictor variables with few response variables

$$\begin{bmatrix} W_{1,1} & \cdots & W_{m,1} \\ \vdots & \ddots & \vdots \\ W_{1,n} & \cdots & W_{m,n} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_m \end{bmatrix} = \begin{bmatrix} V_1 \\ \vdots \\ V_n \end{bmatrix}$$

Reflectance Spectroscopy

VISIBLE NEAR INFRARED SHORTWAVE INFRARED

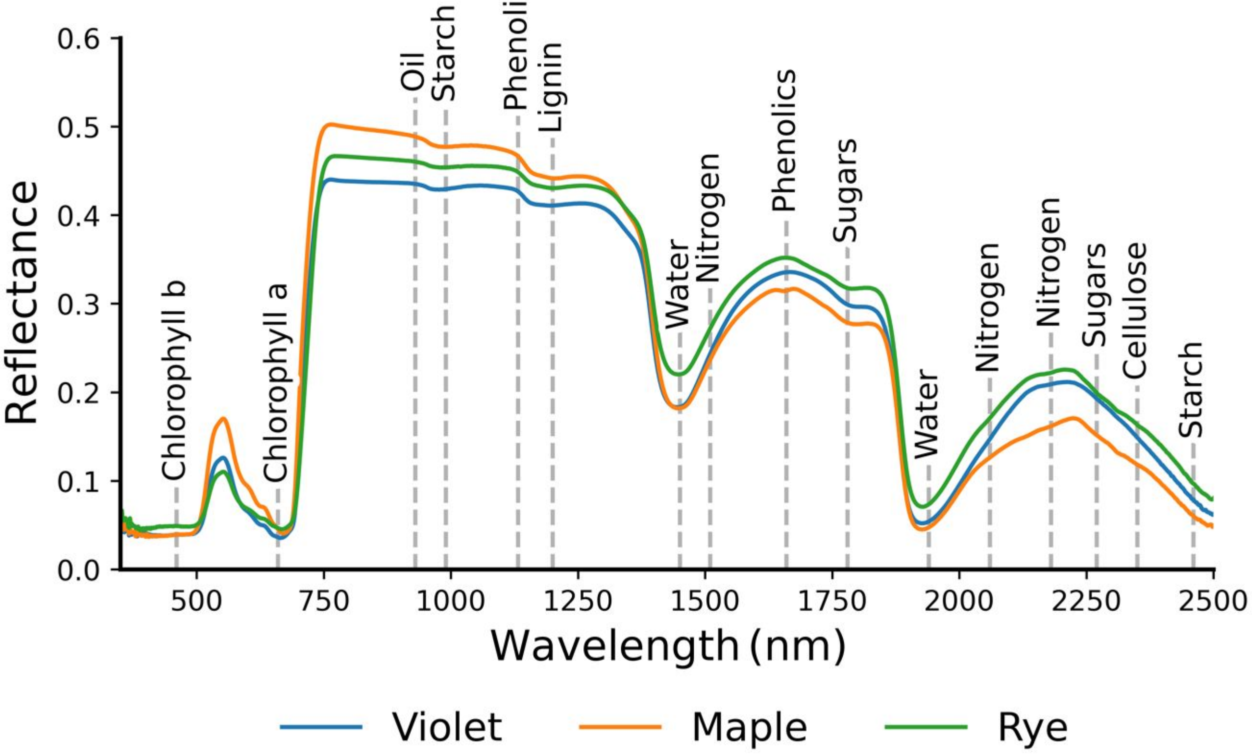
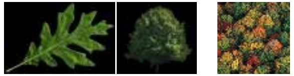
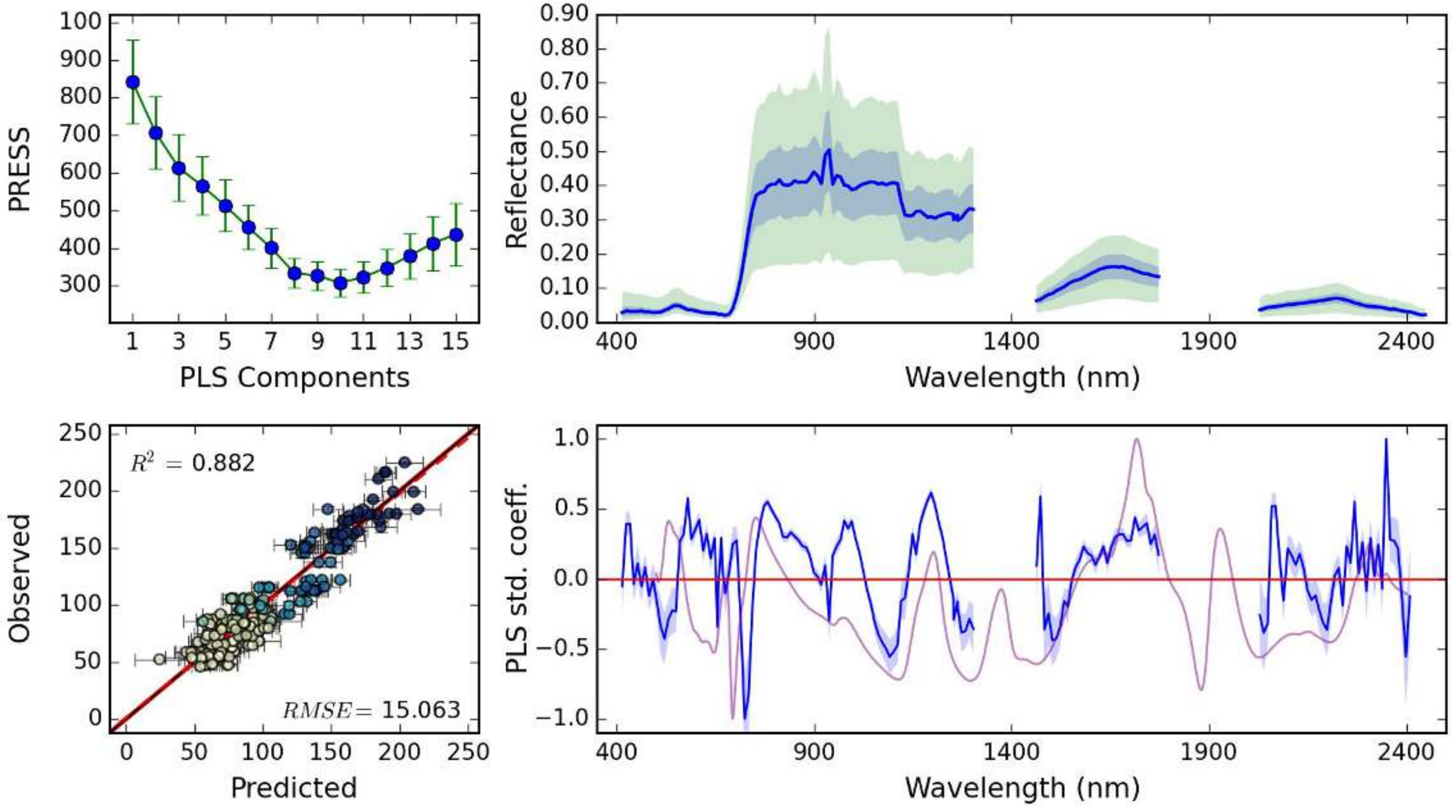


Figure: Adam Chlus



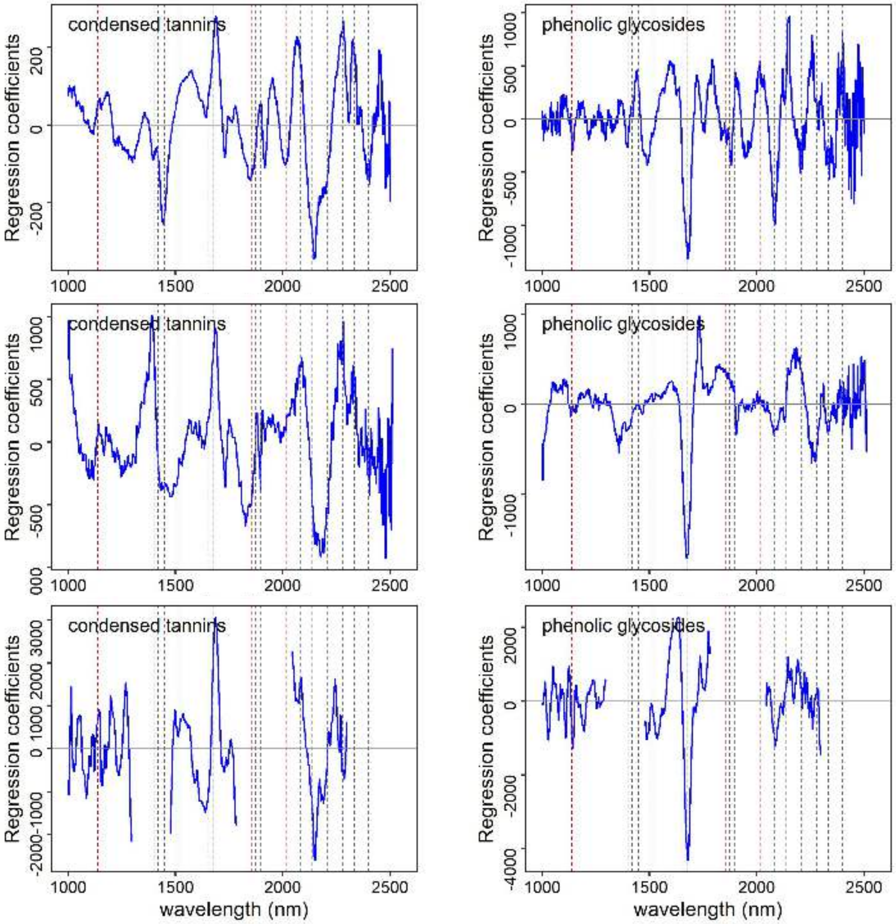
Partial least squares regression

- Chemometric method designed to handle high-dimensional, multicollinear data
- Permutational approach (jackknife) to get model uncertainties



Coefficients

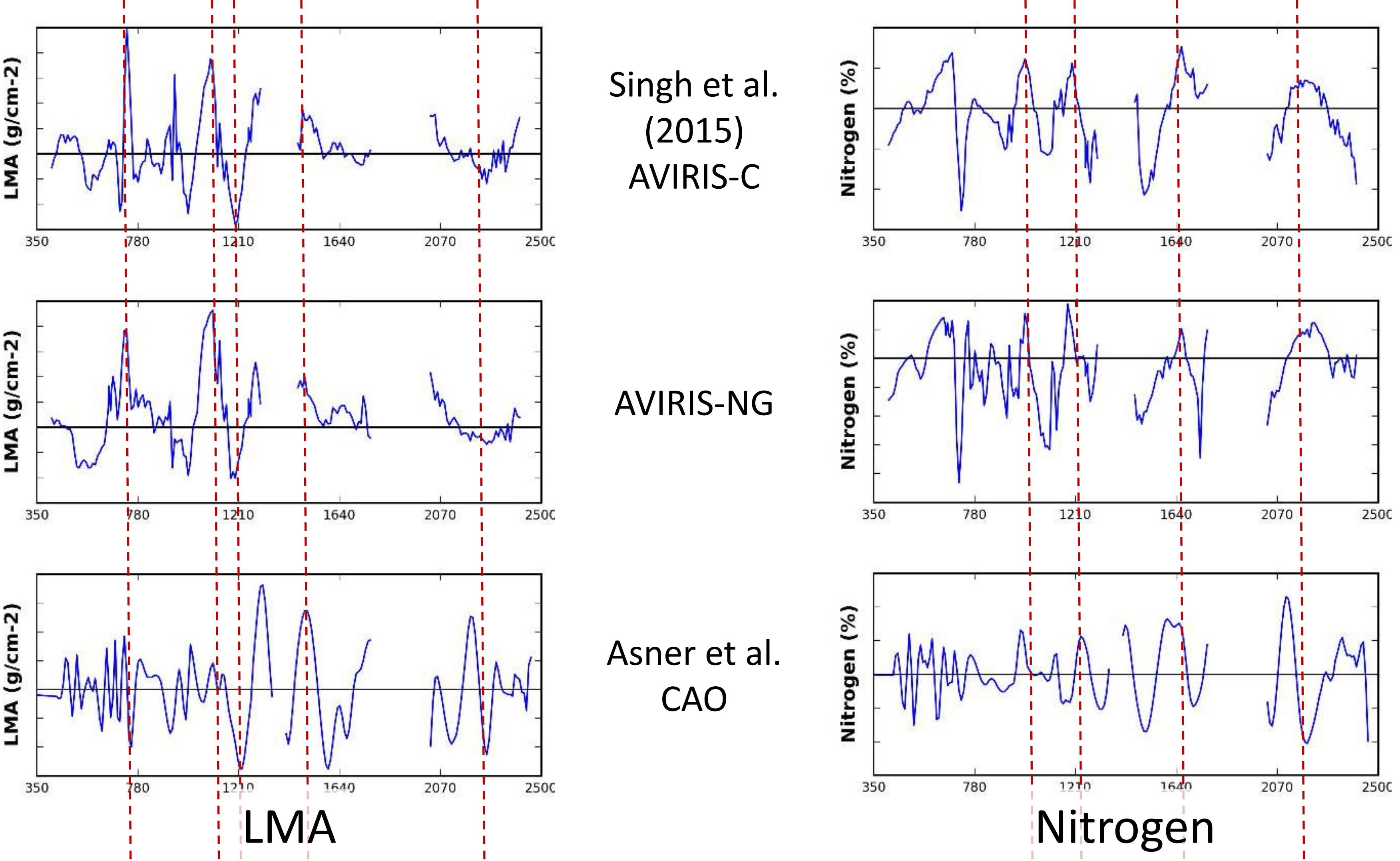
Model comparison



Dry spectra

Fresh spectra

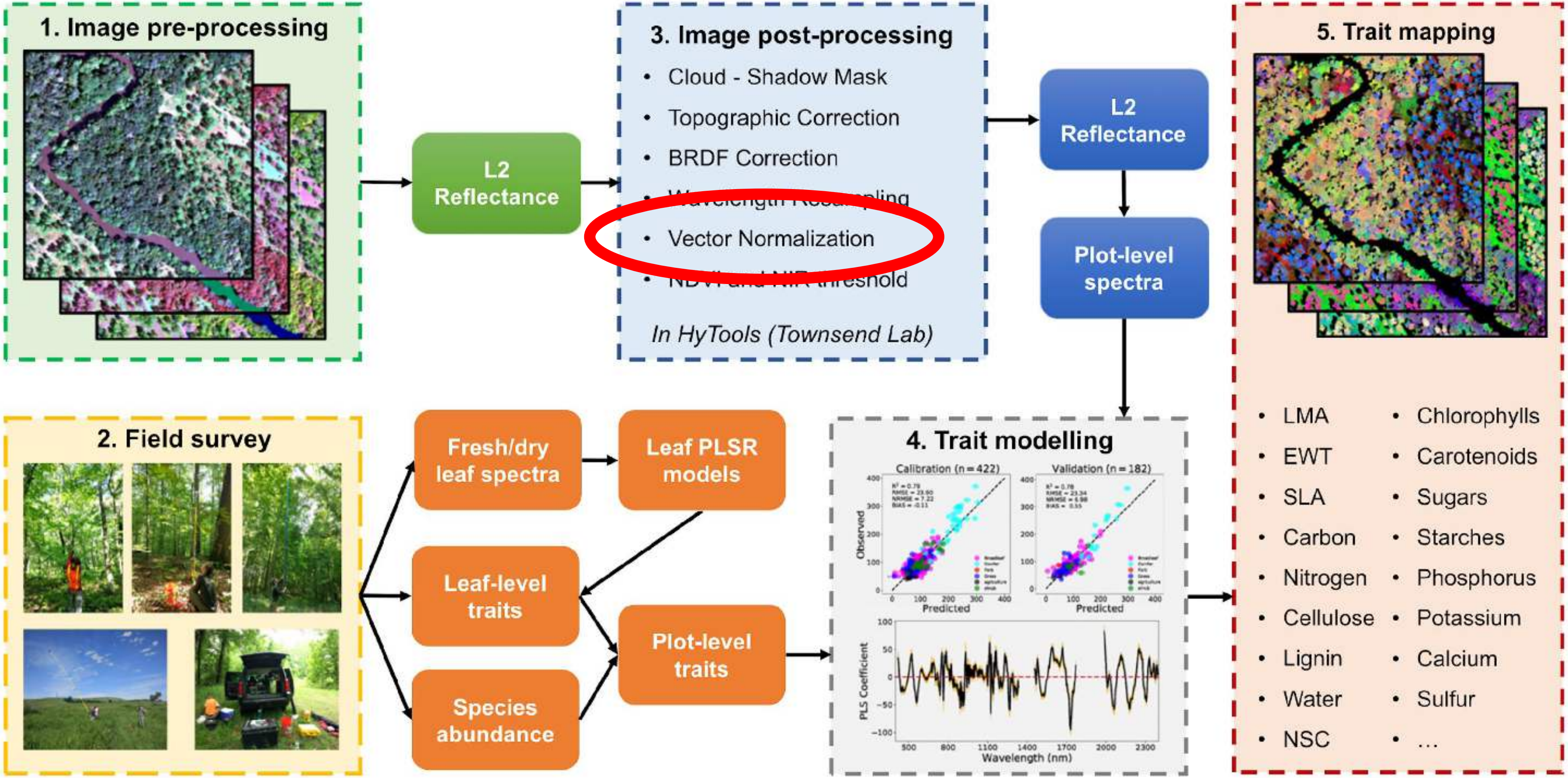
Image spectra


LMA (g/cm-2)
Nitrogen (%)
350
780
1210
1640
2070
2500
Singh et al.
(2015)
AVIRIS-C
AVIRIS-NG
Asner et al.
CAO
LMA
Nitrogen

Analyzing your data - normalization

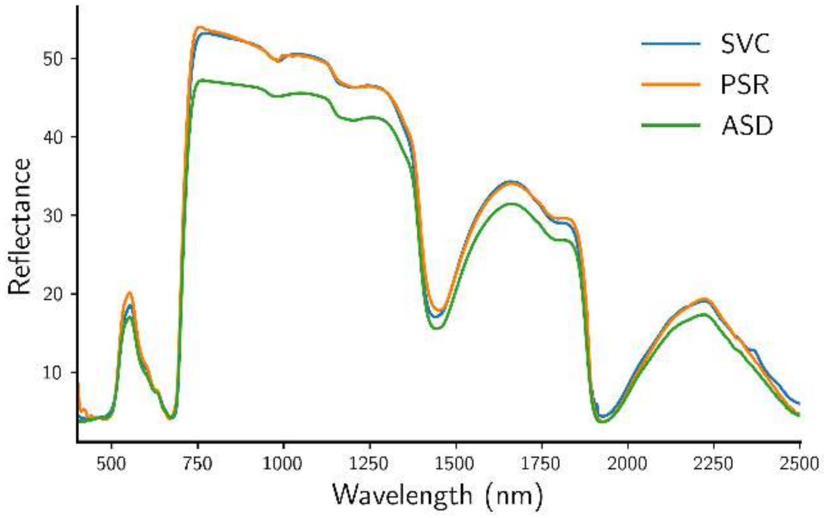
- Many approaches to reduce systematic differences due to differences in lighting
  - Unit-vector normalization
  - Continuum removal

Modeling Overview



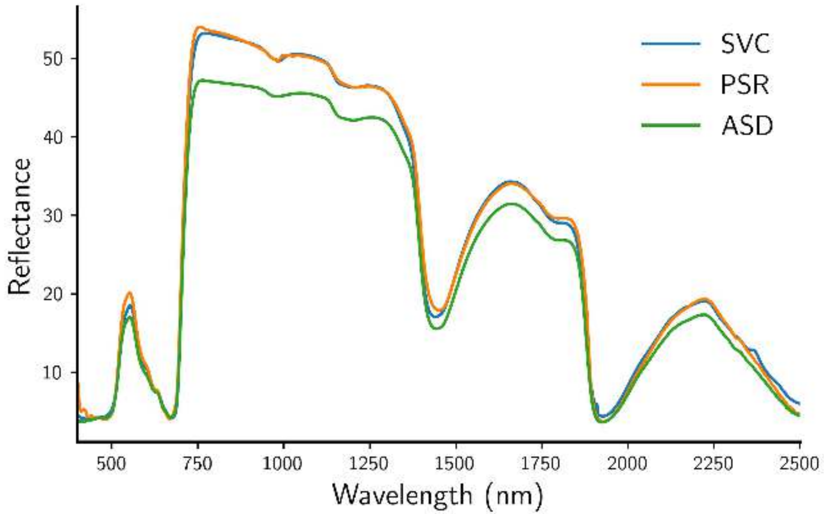
Analyzing your data - normalization

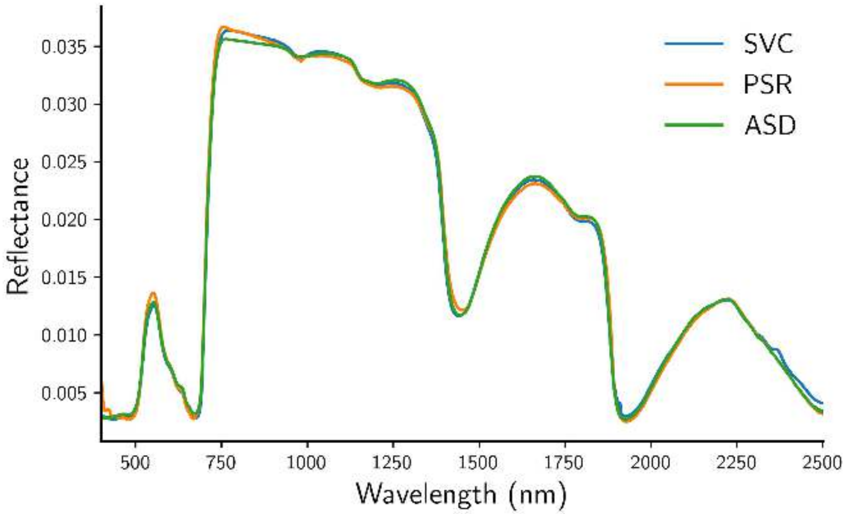
- Many approaches to reduce systematic differences due to differences in lighting
  - Unit-vector normalization
  - Continuum removal



Analyzing your data - normalization

- Many approaches to reduce systematic differences due to differences in lighting
  - Unit-vector normalization
  - Continuum removal





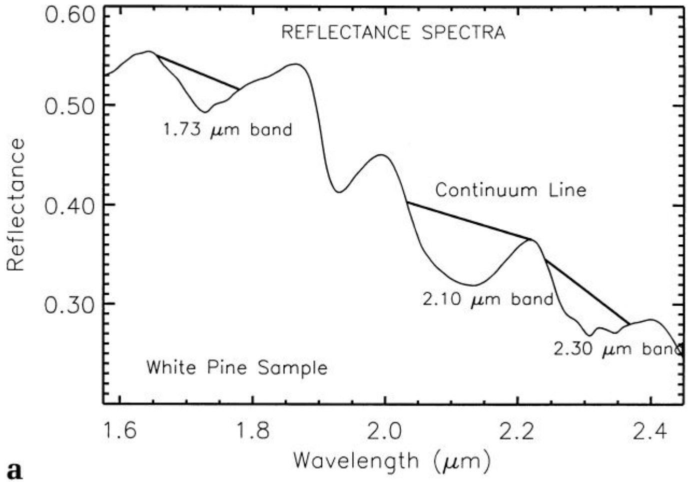
Normalization

- It is a good idea to remove wavelengths you do not plan to use before normalizing

Spectroscopic Determination of Leaf Biochemistry Using Band-Depth Analysis of Absorption Features and Stepwise Multiple Linear Regression

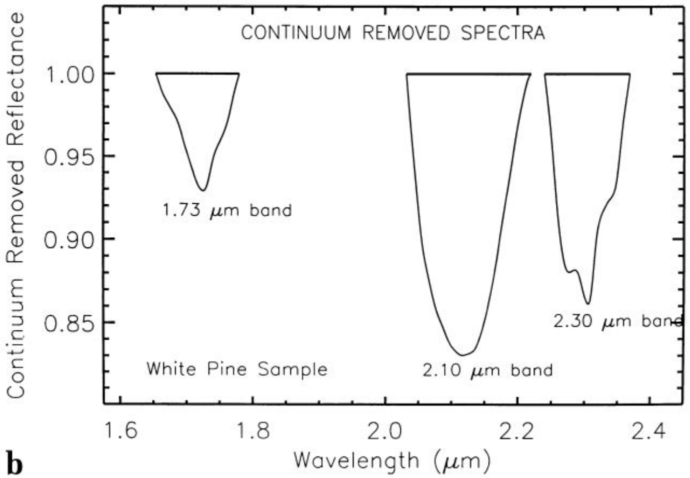
Raymond F. Kokaly and Roger N. Clark**

REMOTE SENS. ENVIRON. 67:267–287 (1999)



a

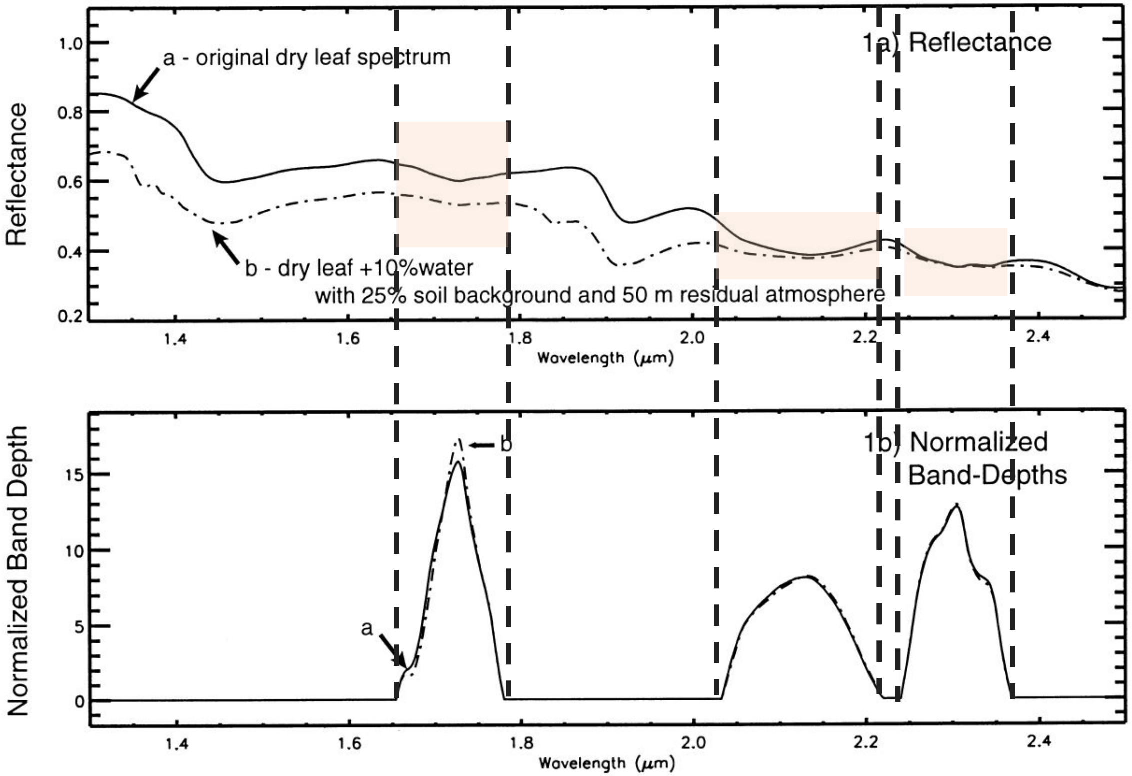
= continuum removal



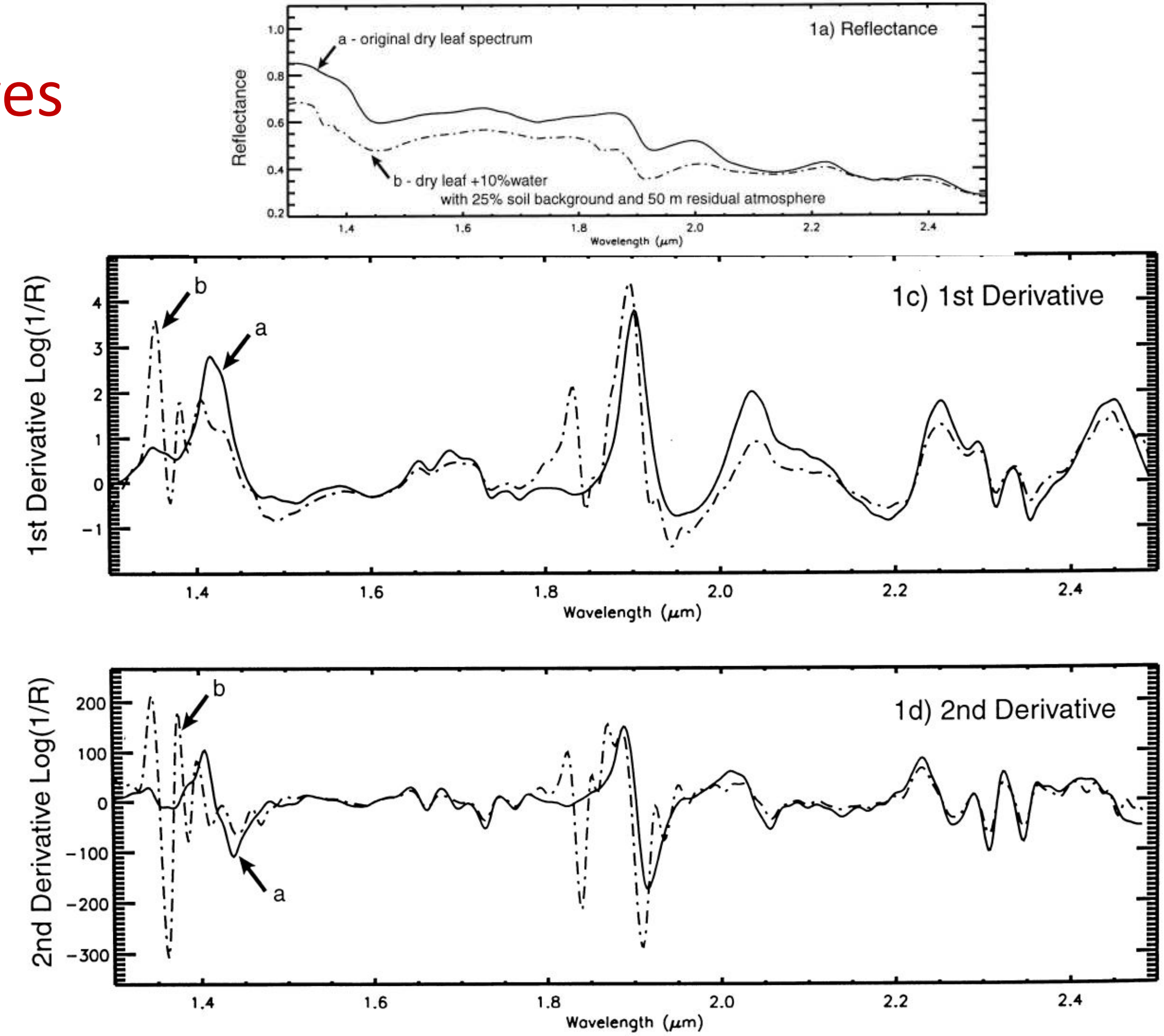
b

Figure 2. Continuum analysis is demonstrated for a white pine sample. Figure 2a shows the continua used to isolate each major absorption feature in dry leaf reflectance spectrum (1.73 μm, 2.10 μm, and 2.30 μm). Figure 2b shows the result of continuum removal for the three features. The continua end points are defined in Table 2.

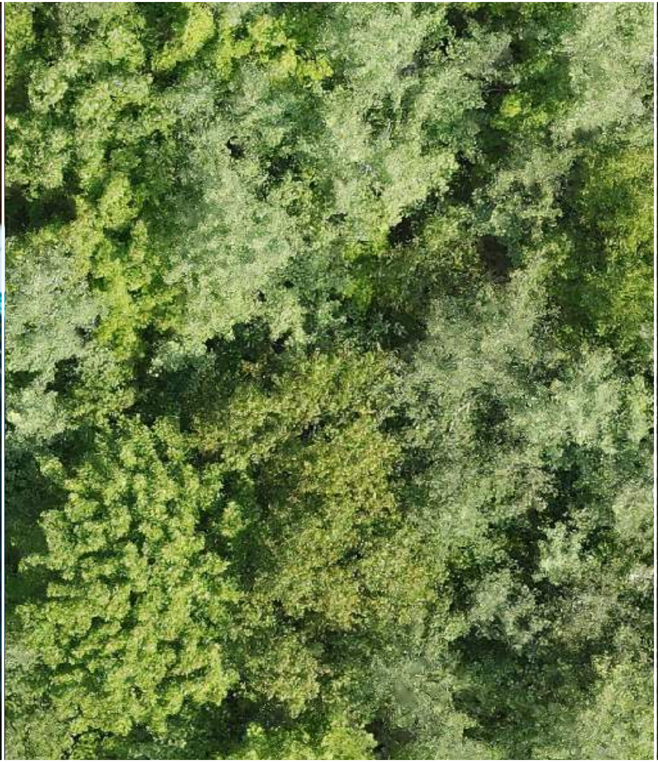
Continuum Removal

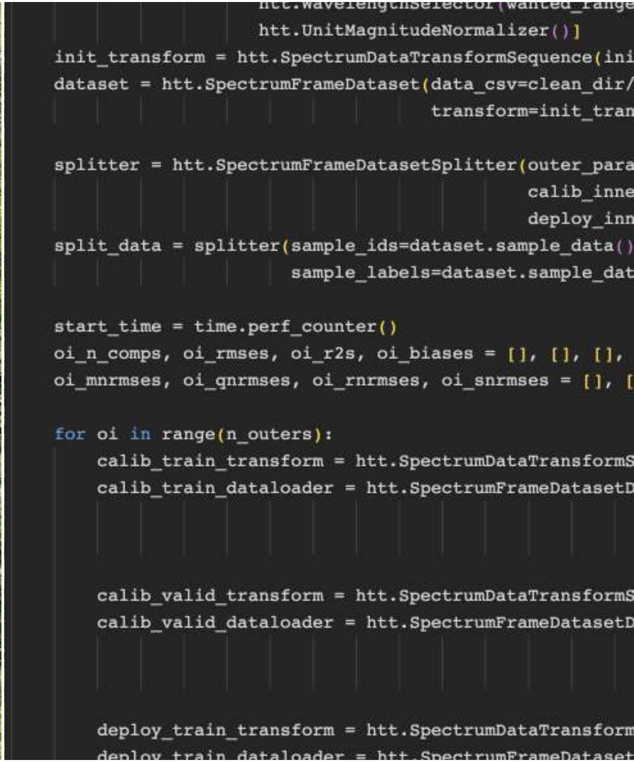


Derivatives



Foliar trait modeling with hyperspectral data – an overview





BIOSCAPE Biodiversity Survey of the Cape

WISCONSIN UNIVERSITY OF WISCONSIN MADISON

EnSpec

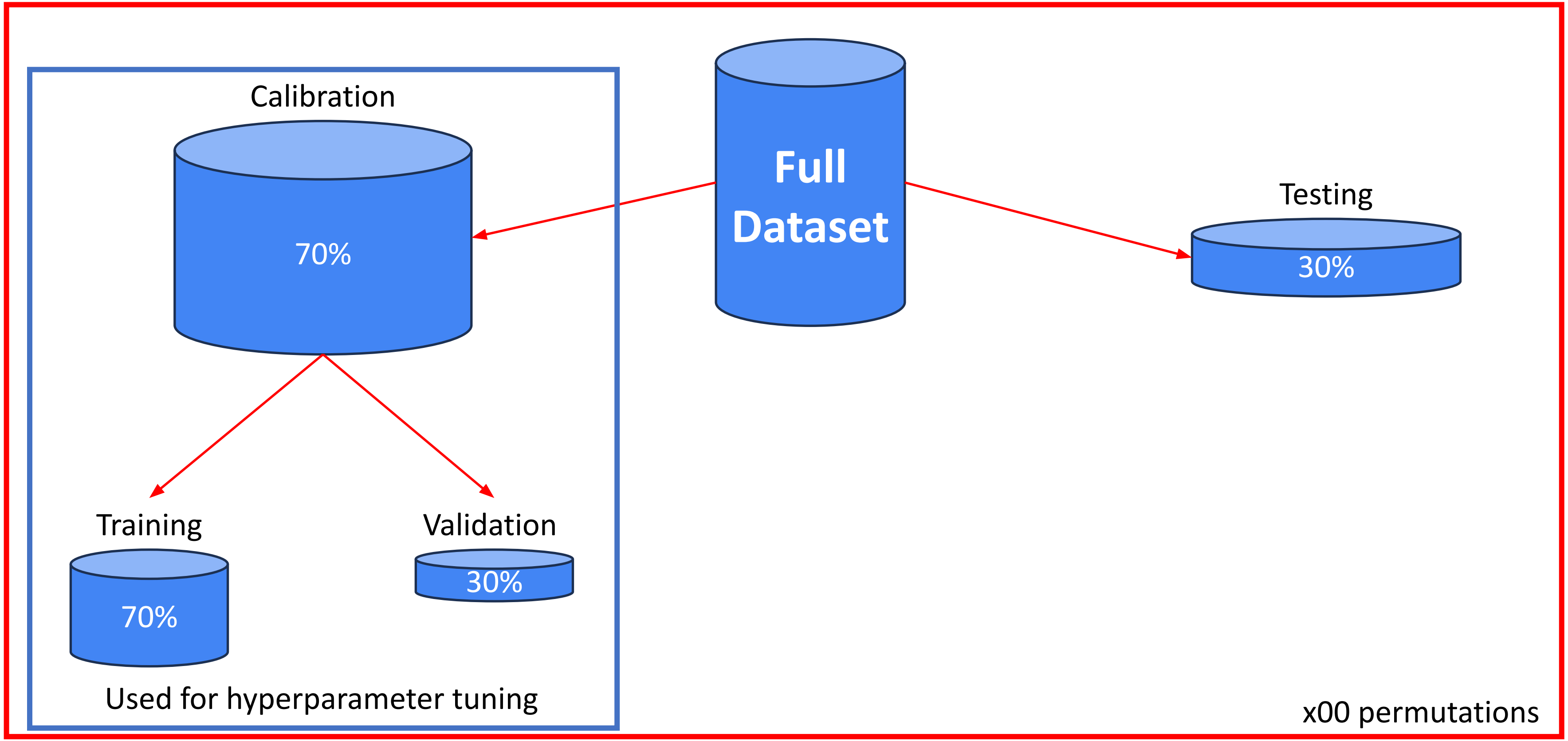
ASCEND spectralbiology.org

NASA

JPL Jet Propulsion Laboratory California Institute of Technology

Split Spectral Data

- Split data into calibration and validation datasets, and also withheld testing set
- Withheld testing set is never used, better if collected separately (such as different year, site, etc.)

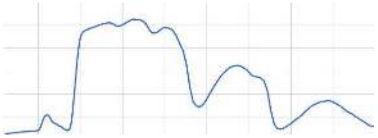


Field to table

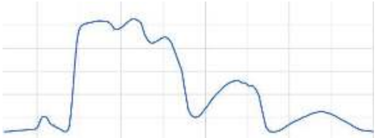
Fieldwork

Consolidated and curated data

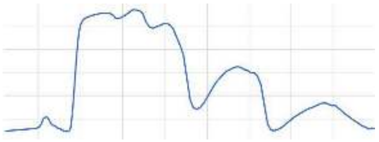
A	B	C	D	E	F	G	H
	0.55033	0.550843	0.550599	0.549292	0.54855	0.548535	2.77
	0.567641	0.568098	0.5683	0.567753	0.565823	0.566529	1.35
	0.62218	0.622497	0.6231	0.621212	0.619791	0.621486	0.88



2.7
7



1.3
5

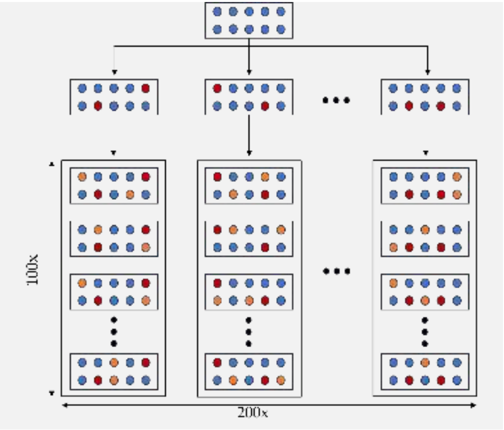


0.8
8

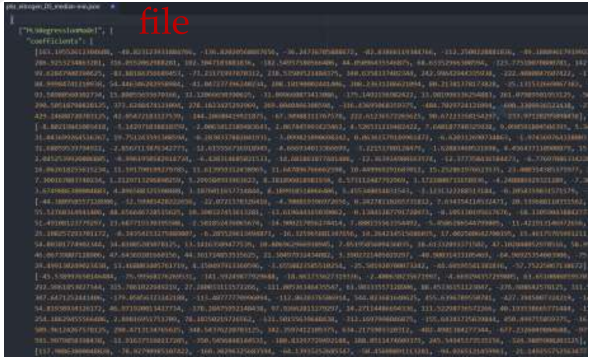
Table to model (HyTraits)

A	B	C	D	E	F	G	H
	0.55033	0.550843	0.550599	0.549292	0.54855	0.548535	2.77
	0.567641	0.568098	0.5683	0.567753	0.565823	0.566529	1.35
	0.62218	0.622497	0.6231	0.621212	0.619791	0.621486	0.88

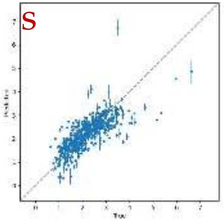
```
import hytraits as H

dataset = H.CSVDataset(spectra_file=<SPECTRA_CSV>,
                       traits_file=<TRAITS_CSV>,
                       trait_column=<TRAIT_NAME>)
splitter = H.CSVDatasetSplitter(split_type='monte_carlo',
                                n_outers=200,
                                n_inners=100,
                                train_valid_test_props=(0.6, 0.2, 0.2))
transform = H.Compose([H.ToDevice(device='cpu'),
                       H.WavelengthsRemoval(remove_ranges=<RANGES>),
                       H.UnitVectorNormalization()])
trainer = H.PLSRegressionTrainer(n_components=<MAX_NUM_COMPONENTS>,
                                 transform=transform)
models = trainer(dataset=dataset,
                 splits=splits)
```

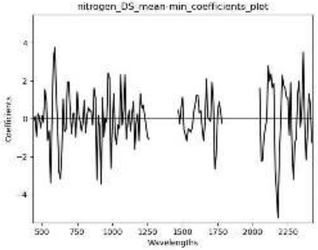


Model JSON file

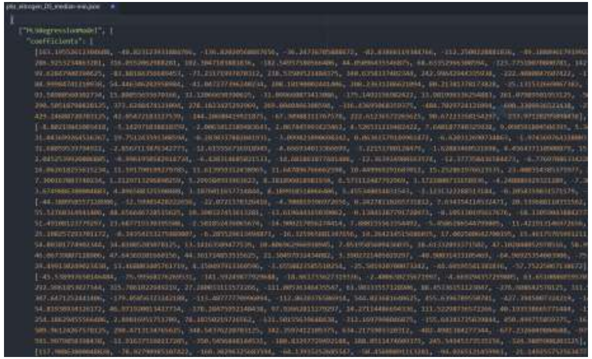


Diagnostics

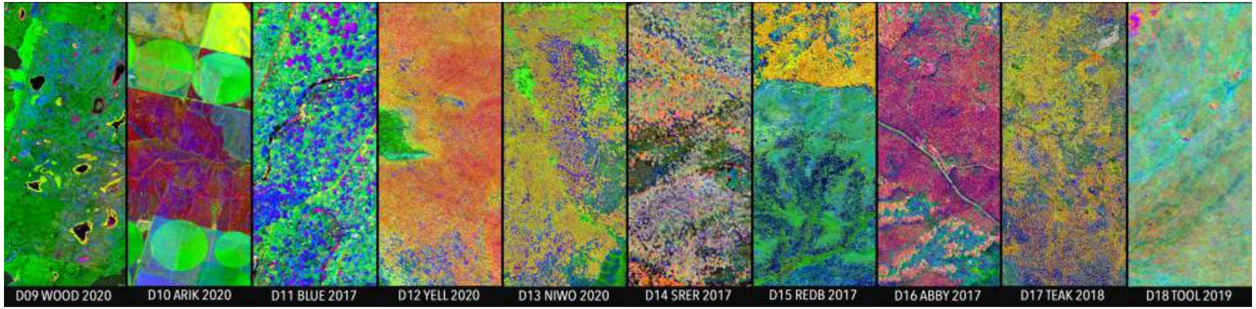




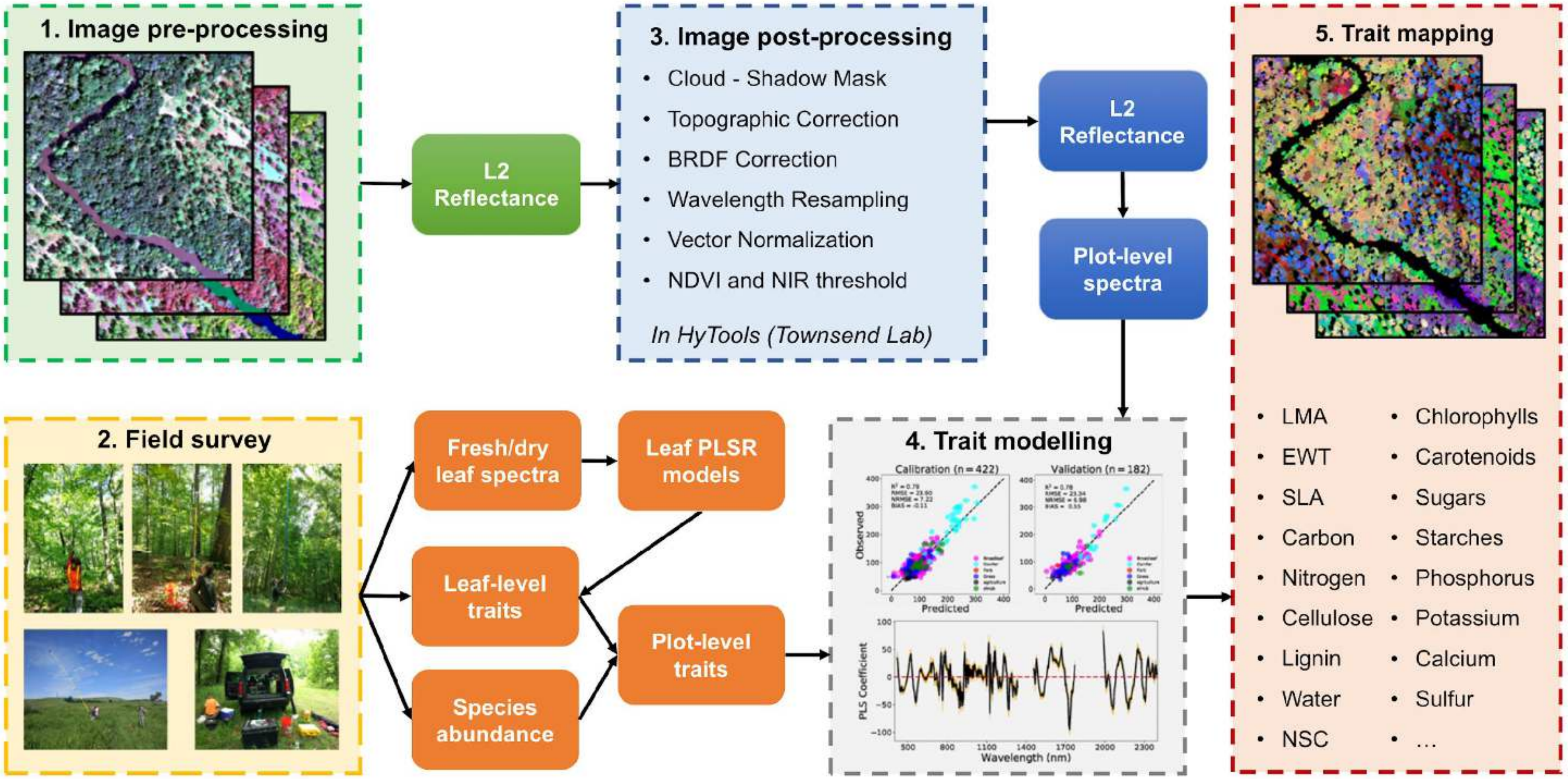
Model to trait maps (HyTools)



HyTools



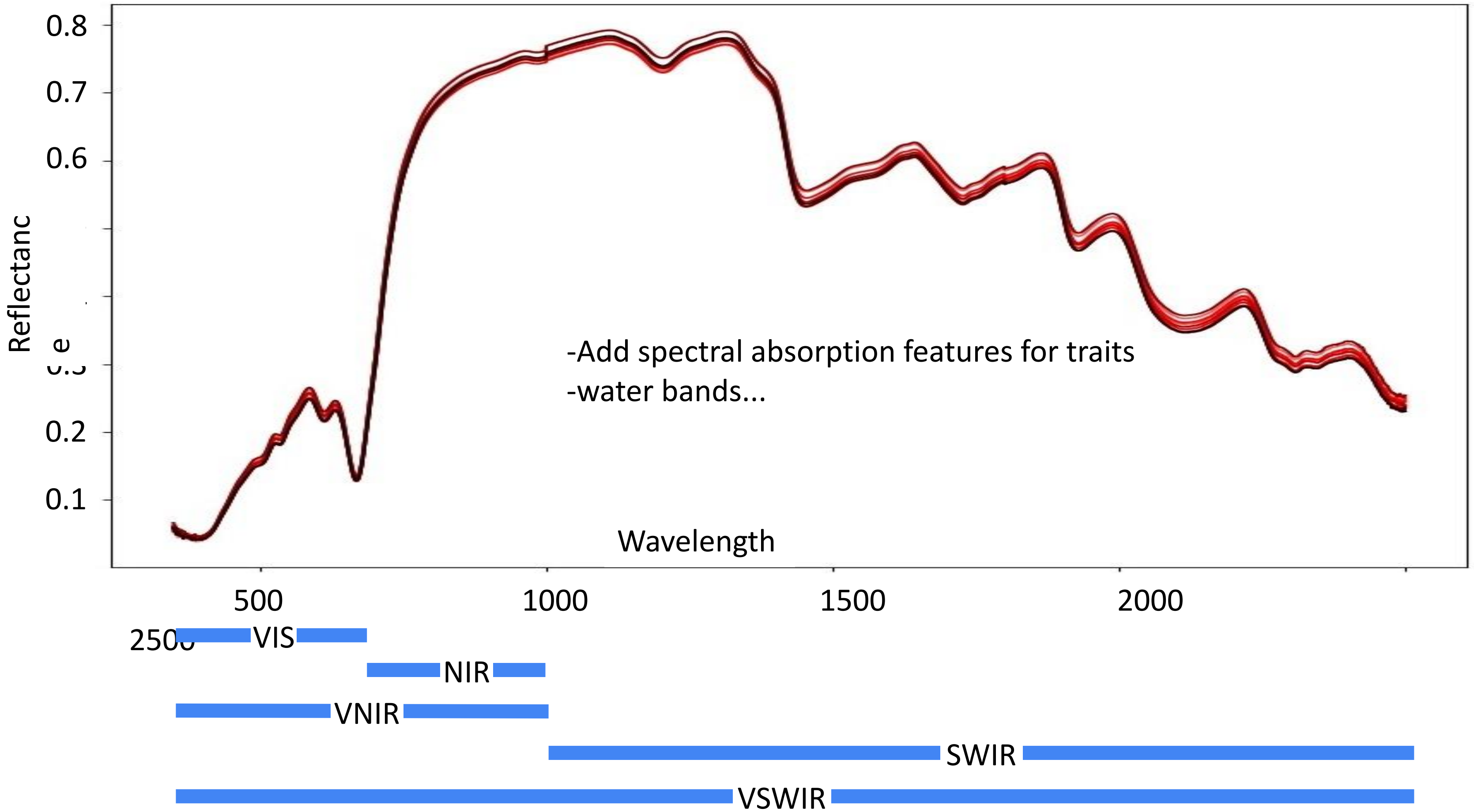
Modeling Overview



PLSR Interpretation

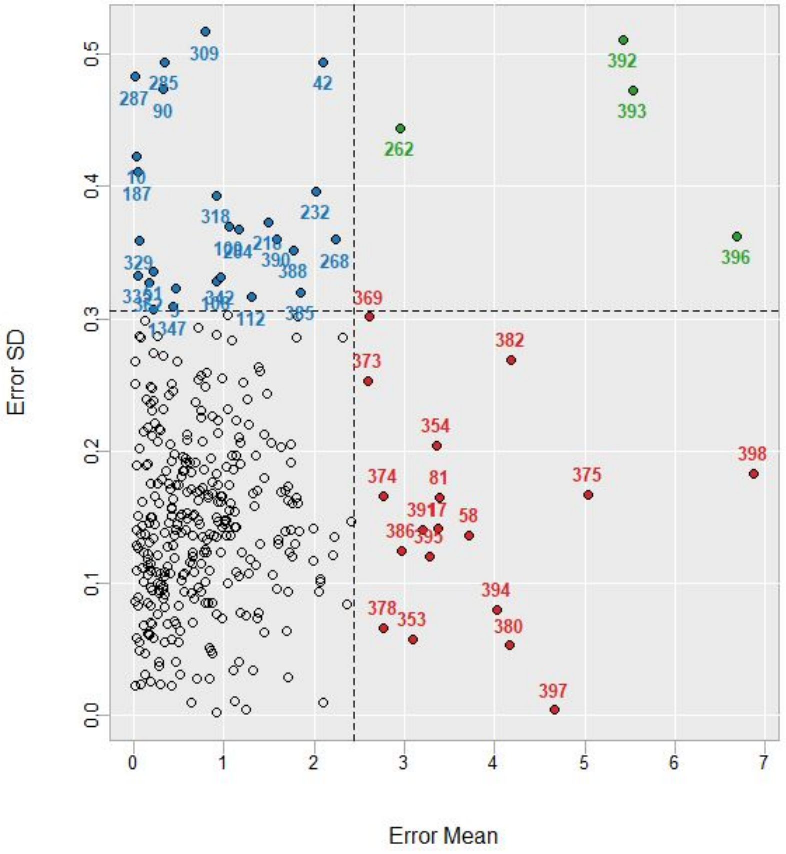
- Coefficients
    - Raw coefficients
    - Standardized coefficients
    - VIP

Model Spectral Range Grouping



Handling Outliers

- Train PLSR model with cross-validation.
- Run PLSR model with a specific number of permutation.
- Use a standard deviation threshold to remove outliers.

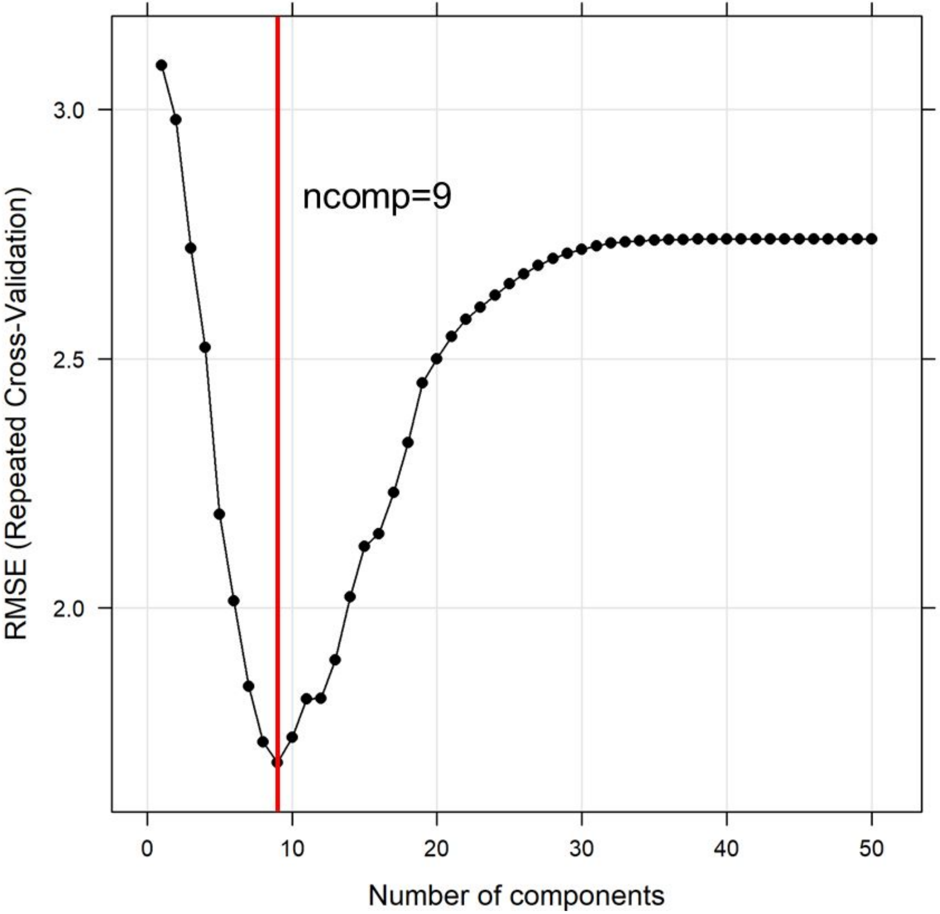


Permutational PLSR

- We quantify uncertainties by running up to 500 models for each trait and using the variation as metric of error

Hyper-parameter Tuning

Determine number of components (latent vectors)

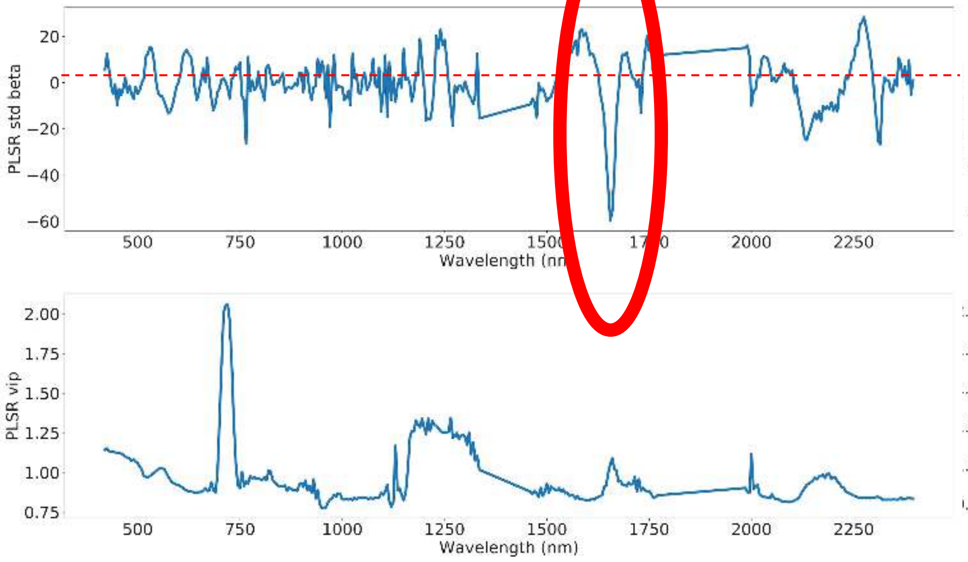


Model Diagnostics

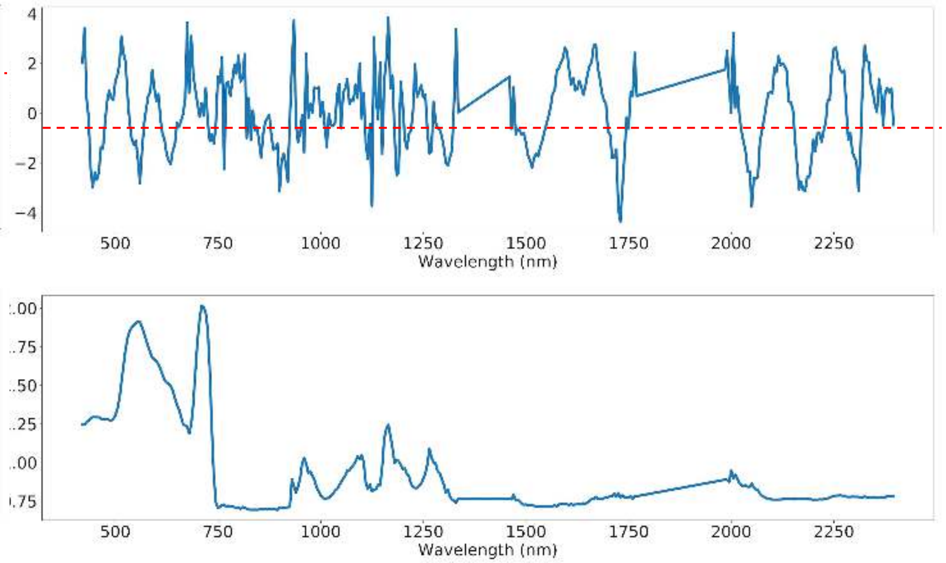
- Diagnostics come from model building, the cross-validation step, and the independent validation step
- R^2
- Root mean squared error
- Normalized root mean squared error (mean, range)

PLS coefficients and model VIPs

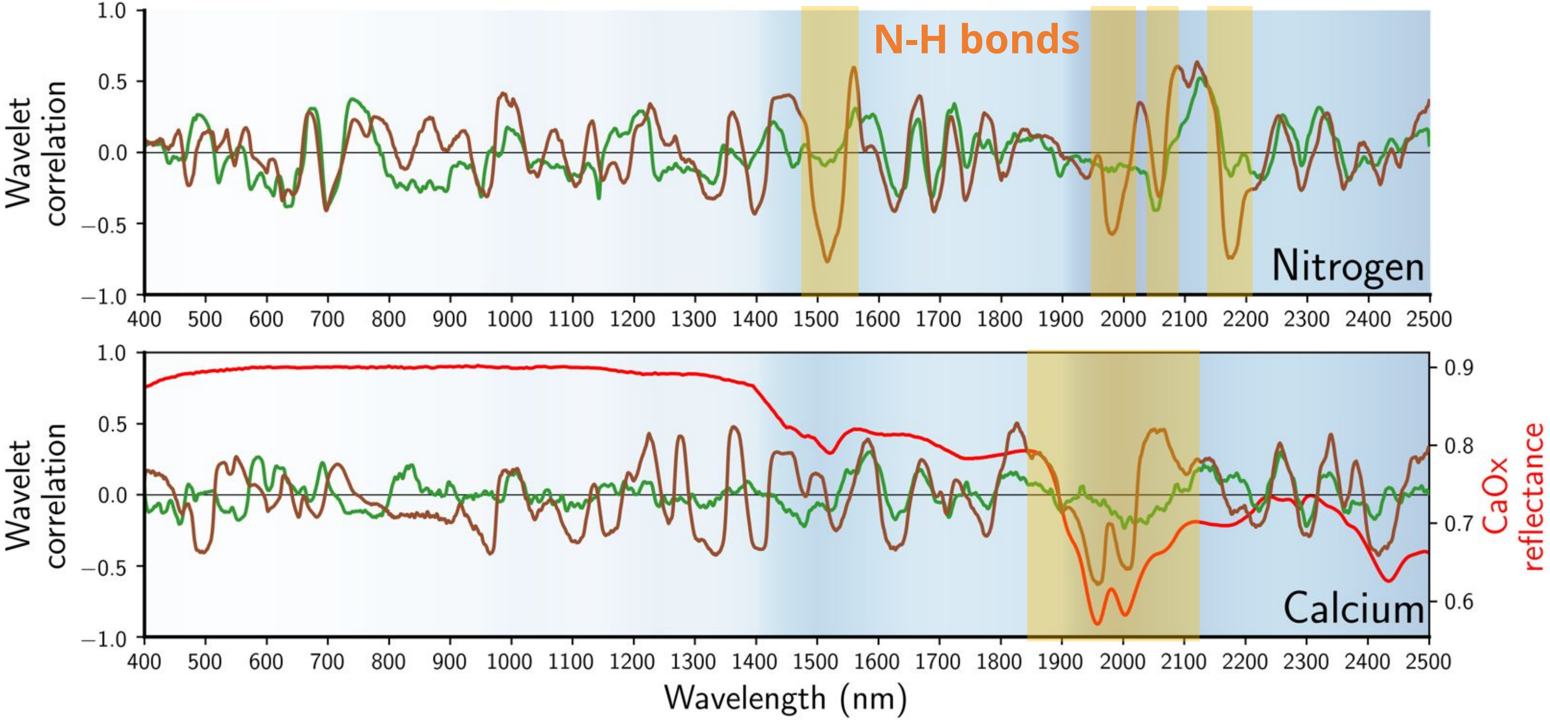
Total Phenolics

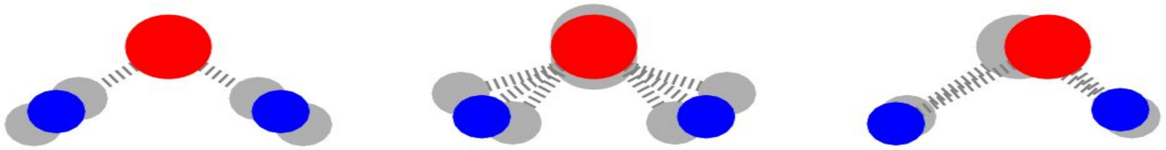


Nitrogen

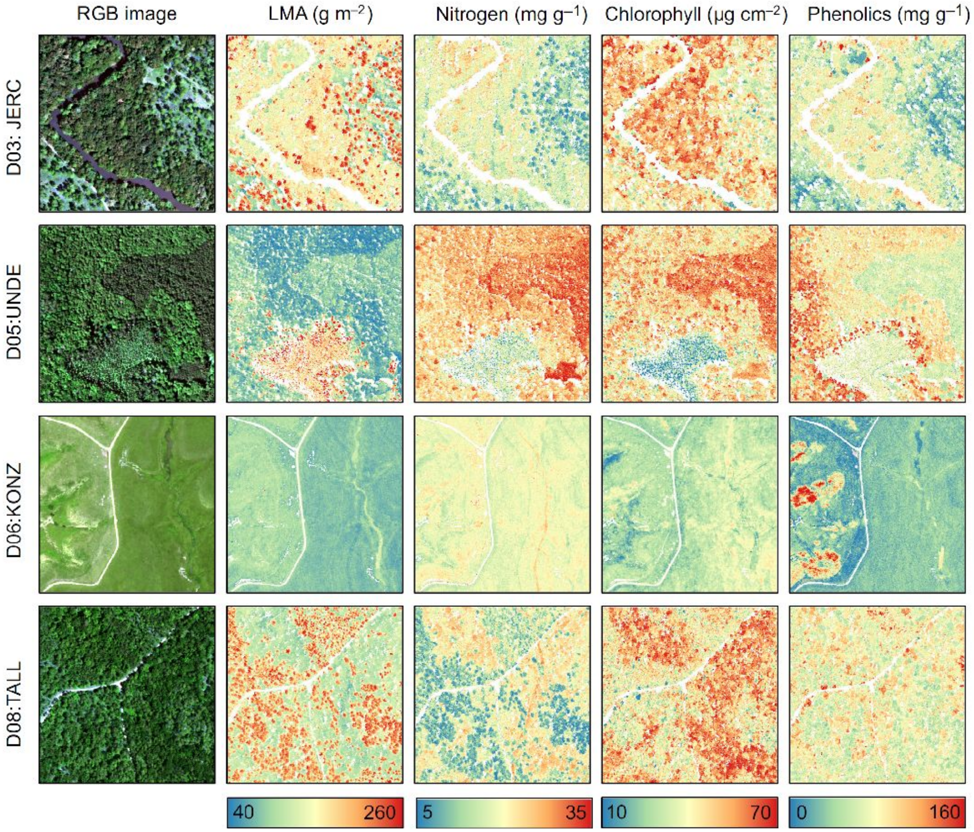


Spectra-trait correlations





Traits	Unit	Field measurements					Statistics of PLSR models for validation dataset							
		Minimum	Maximum	Mean	SD	CV	h	R^2	RMSE	BIAS	%RMSE	$RMSE_s$	$RMSE_u$	d
LMA	$g\ m^{-2}$	33.87717	507.069	105.370	62.990	0.60	17	0.77	24.067	−3.94381	9.2	8.71	22.43	0.94
Nitrogen (N)	$mg\ g^{-1}$	7.40809	58.620	21.338	7.205	0.34	18	0.61	4.392	−0.17317	14.3	1.96	3.93	0.88
ChlAB_area	$\mu g\ cm^{-2}$	11.56937	98.389	44.140	12.060	0.27	19	0.68	6.919	0.60446	11.4	2.97	6.25	0.91
Phenolics	$mg\ g^{-1}$	9.02305	271.966	67.783	36.926	0.55	20	0.82	15.332	−2.60204	9.1	7.90	13.14	0.95



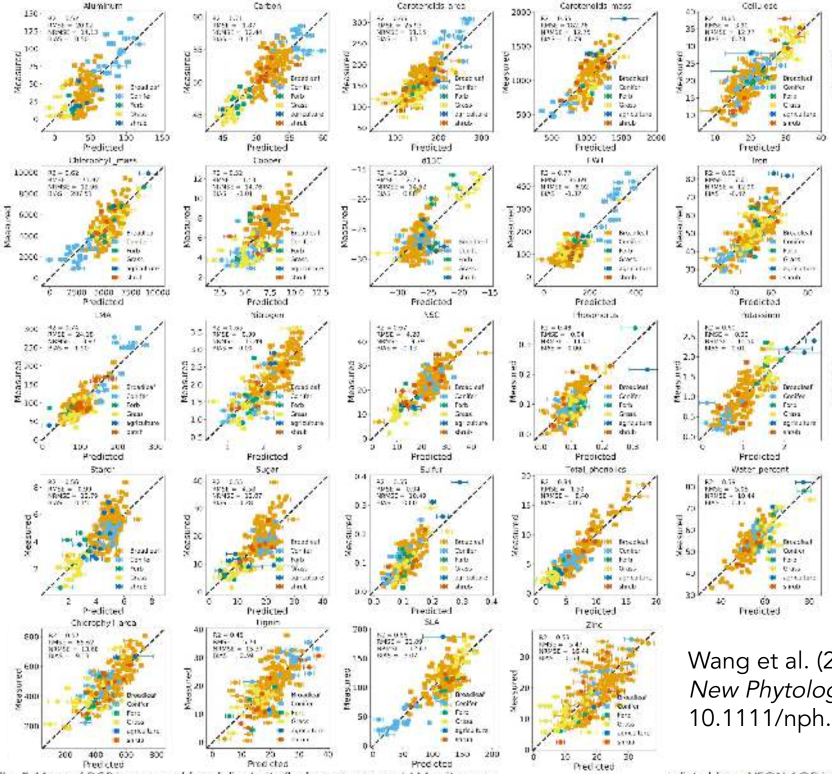
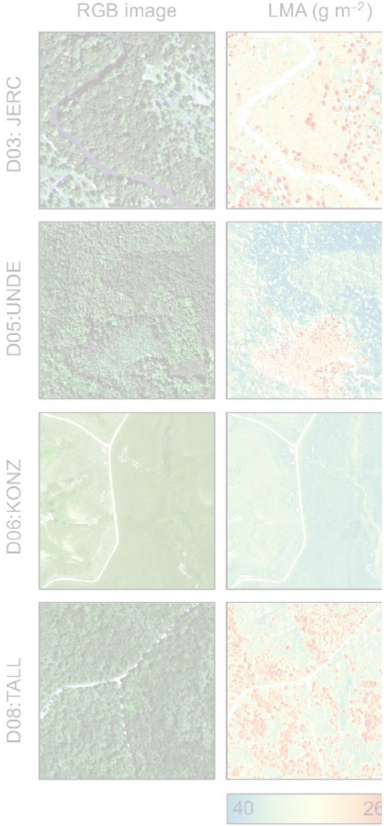
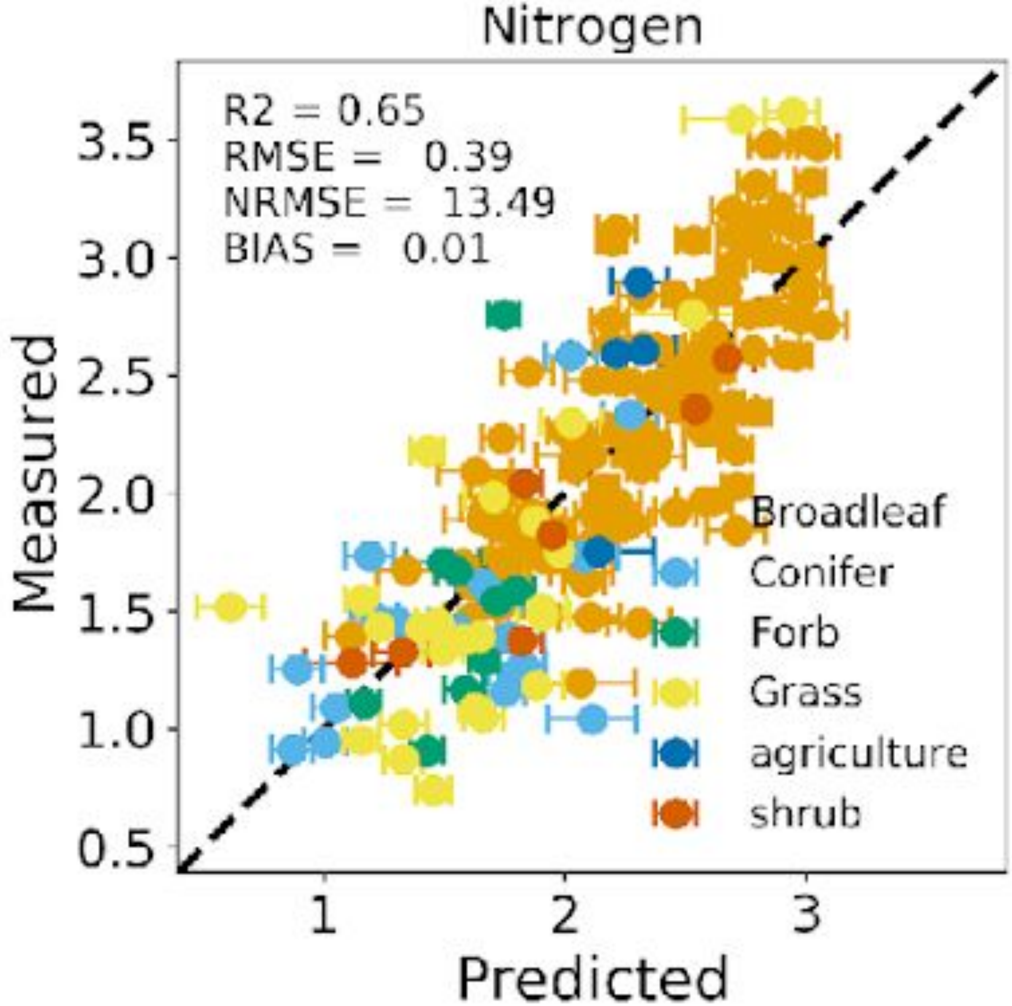


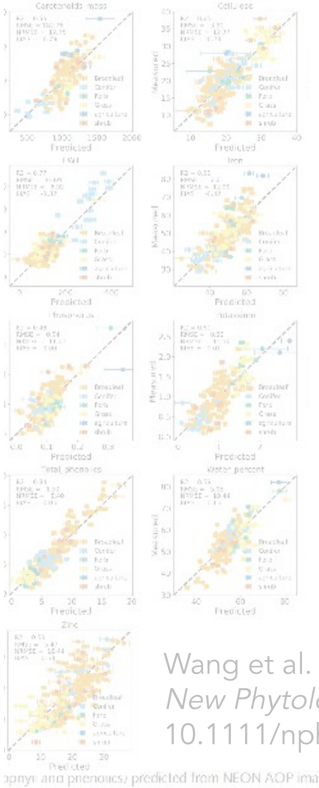
Fig. 5 Maps of RGB images and four foliar traits (leaf mass per area, LMA; nitrogen; chlorophyll and phenolics) predicted from NEON AOP images for four NEON sites. D03:JERC, Jones Ecological Research Center in Southeast; D05:UNDE, Notre Dame Environmental Research Center (UNDERC) in Great Lakes; D06:KONZ, Konza Prairie Biological Station in Prairie Peninsula; D08:TALL, Talladega National Forest in Ozarks Complex.

Wang et al. (2020)
New Phytologist
10.1111/nph.16711

Traits	Unit	M…	…SE	RMSE$_s$	RMSE$_u$	d
LMA	g m^{-2}			8.71	22.43	0.94
Nitrogen (N)	mg g^{-1}			1.96	3.93	0.88
ChlAB_area	µg cm^{-2}			2.97	6.25	0.91
Phenolics	mg g^{-1}			7.90	13.14	0.95







Wang et al. (2020)
New Phytologist
10.1111/nph.1671

…ophyll and phenolics) predicted from NEON AOP images for four
…amie Environmental Research Center (UNDERC) in Great Lakes;
…ional Forest in Ozarks Complex.

True Color

Table Mountain

True-color

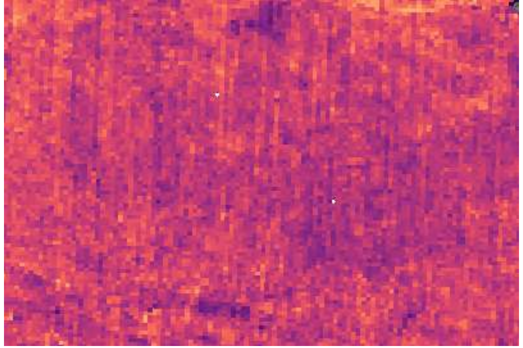
Close-up of two vegetation plots (white dots), T195 and T072 on top of Table Mountain.

T195 is dominated by Anthochortus (Restionaceae >90% cover).

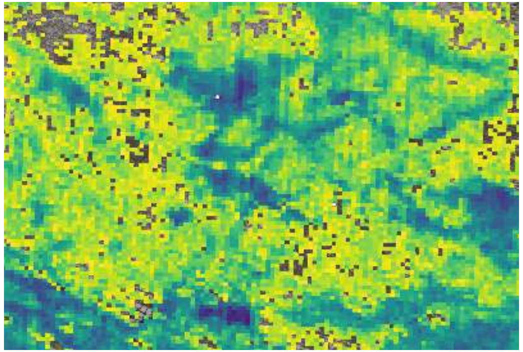
T072 is a mix of Elegia and Anthochortus, likely lower cover.

T195 is dominated by Anthochortus (Restionaceae >90% cover).

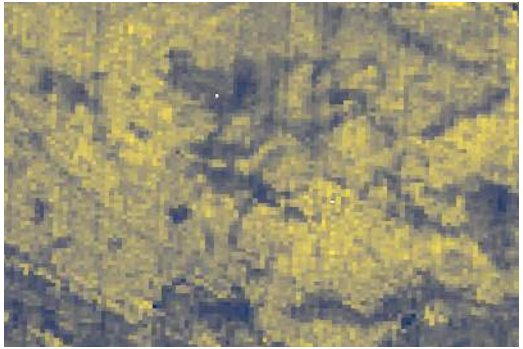
T072 is a mix of Elegia and Anthochortus, likely lower cover.



Nitrogen

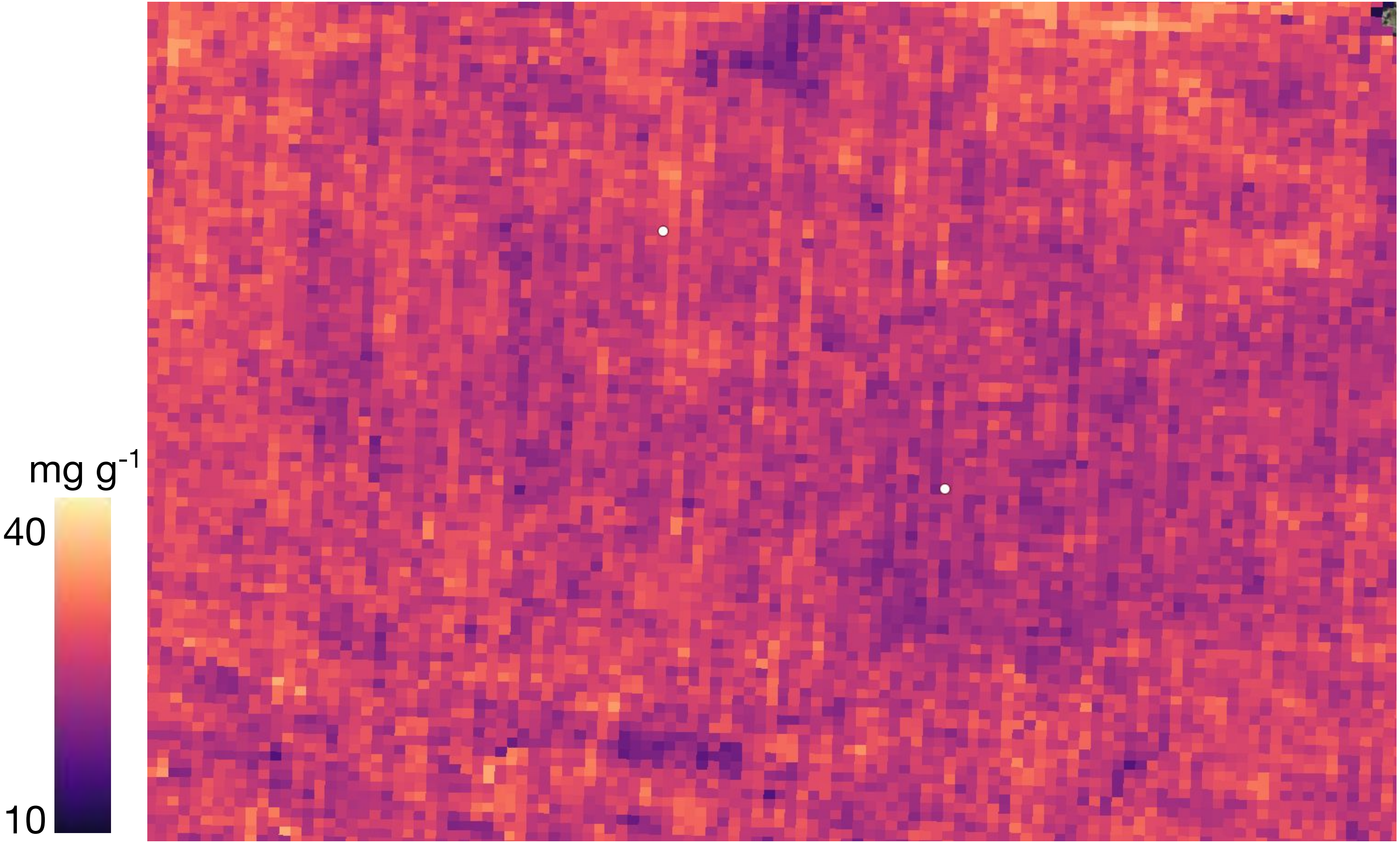


Non-structural carbohydrates

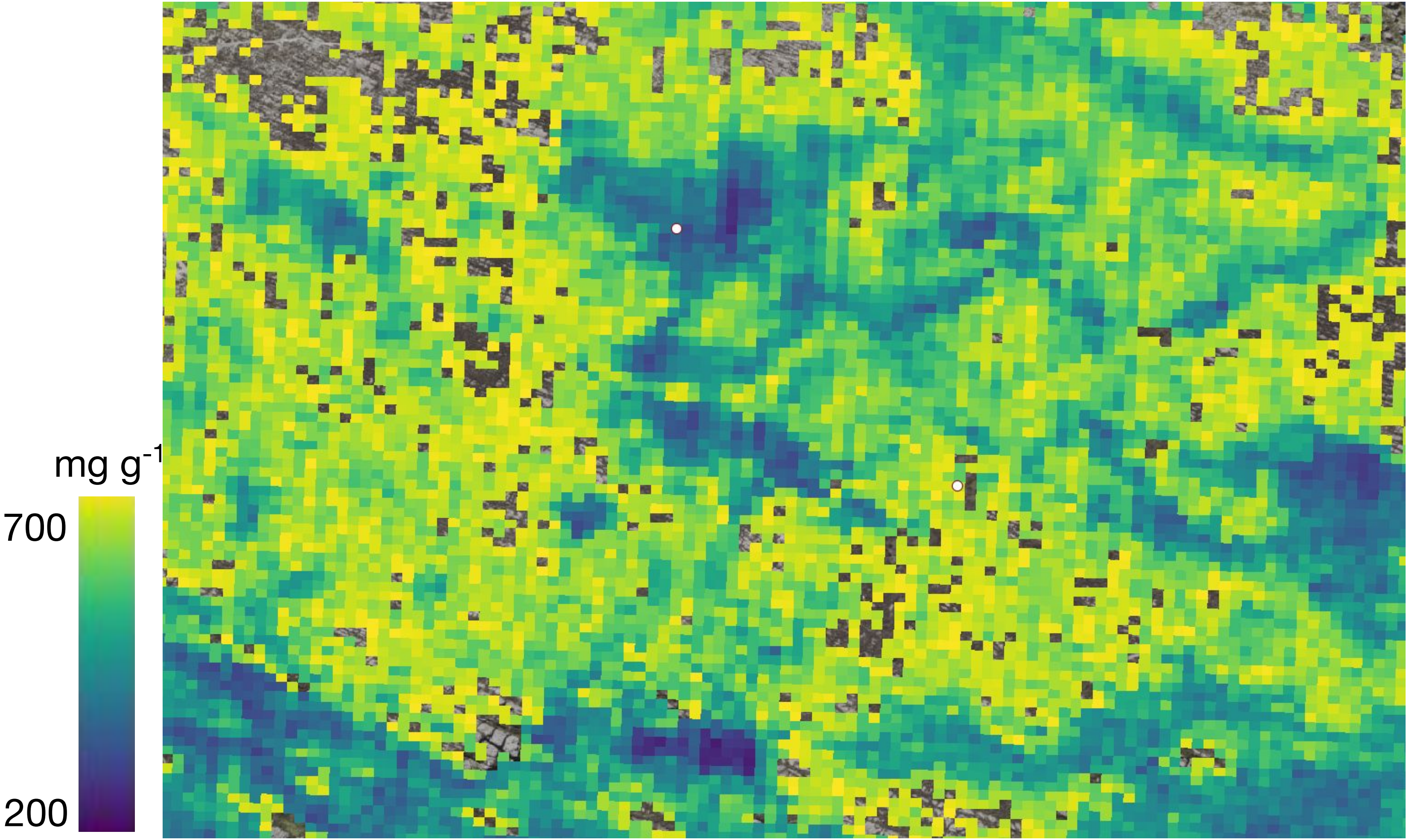


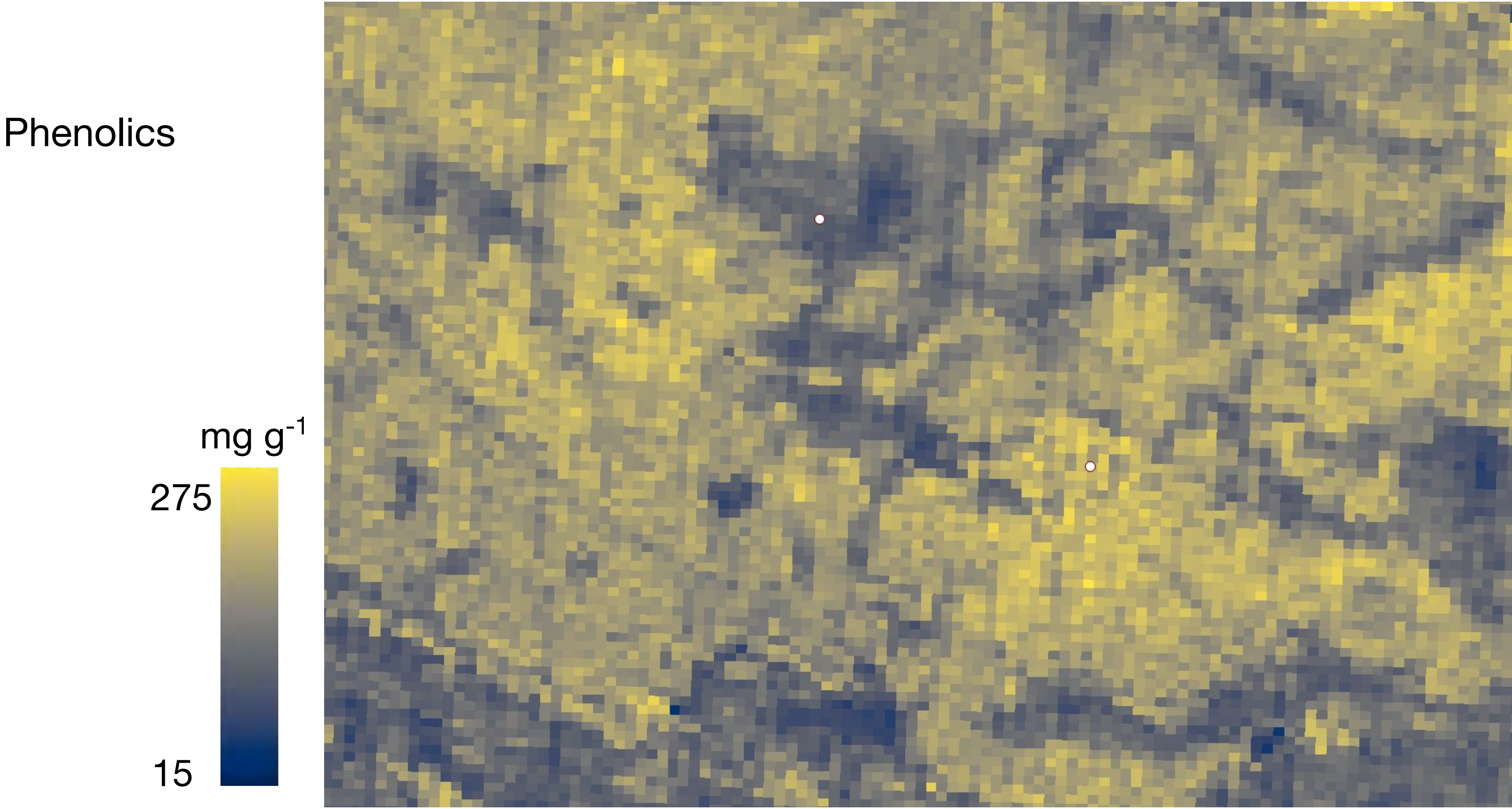
Phenolics

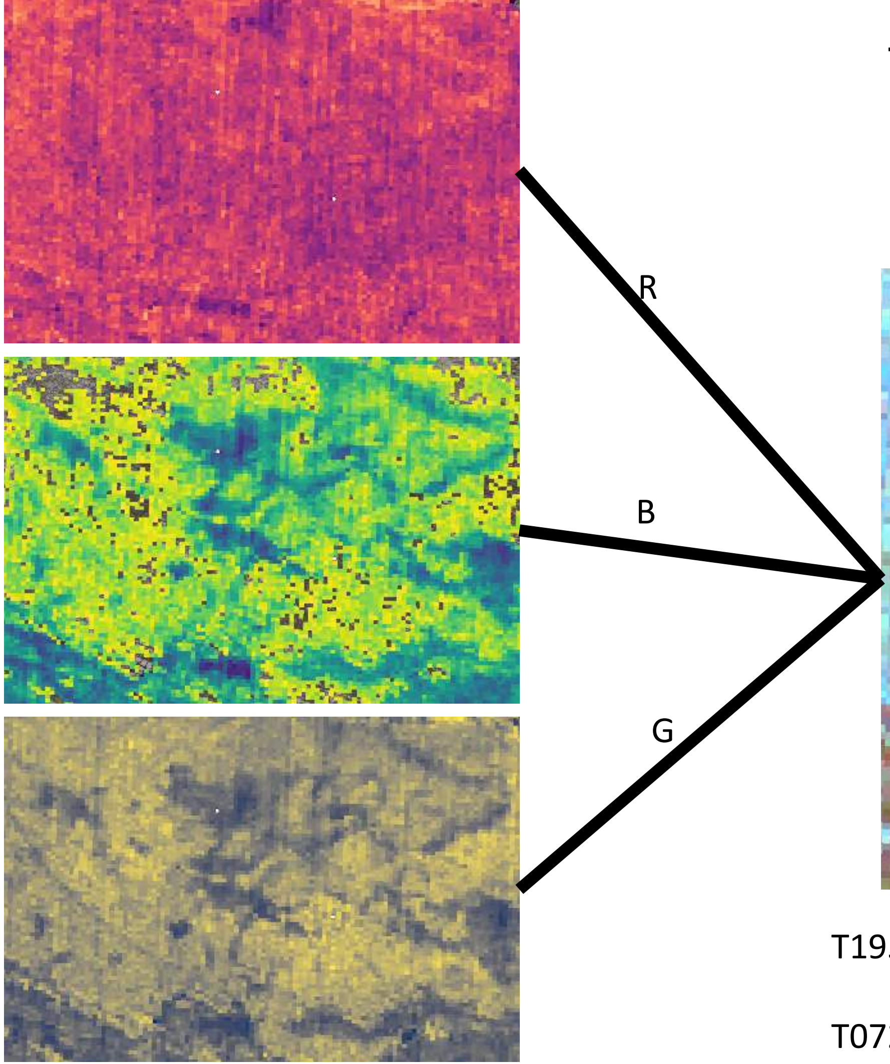
Nitrogen



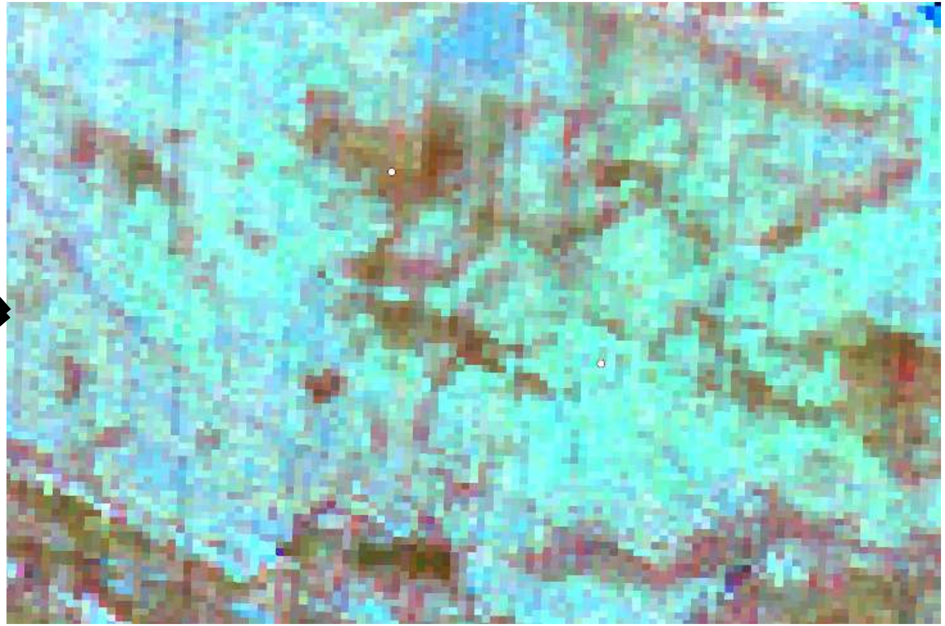
NSCs




Phenolics
mg g-1
275
15



Trait ternary map (Redder pixels = Nitrogen predominates, Greener pixels = Phenolics predominate, bluer pixels = NSCs predominate



T195 is dominated by Anthochortus (Restionaceae >90% cover).

T072 is a mix of Elegia and Anthochortus, likely lower cover.

True Color

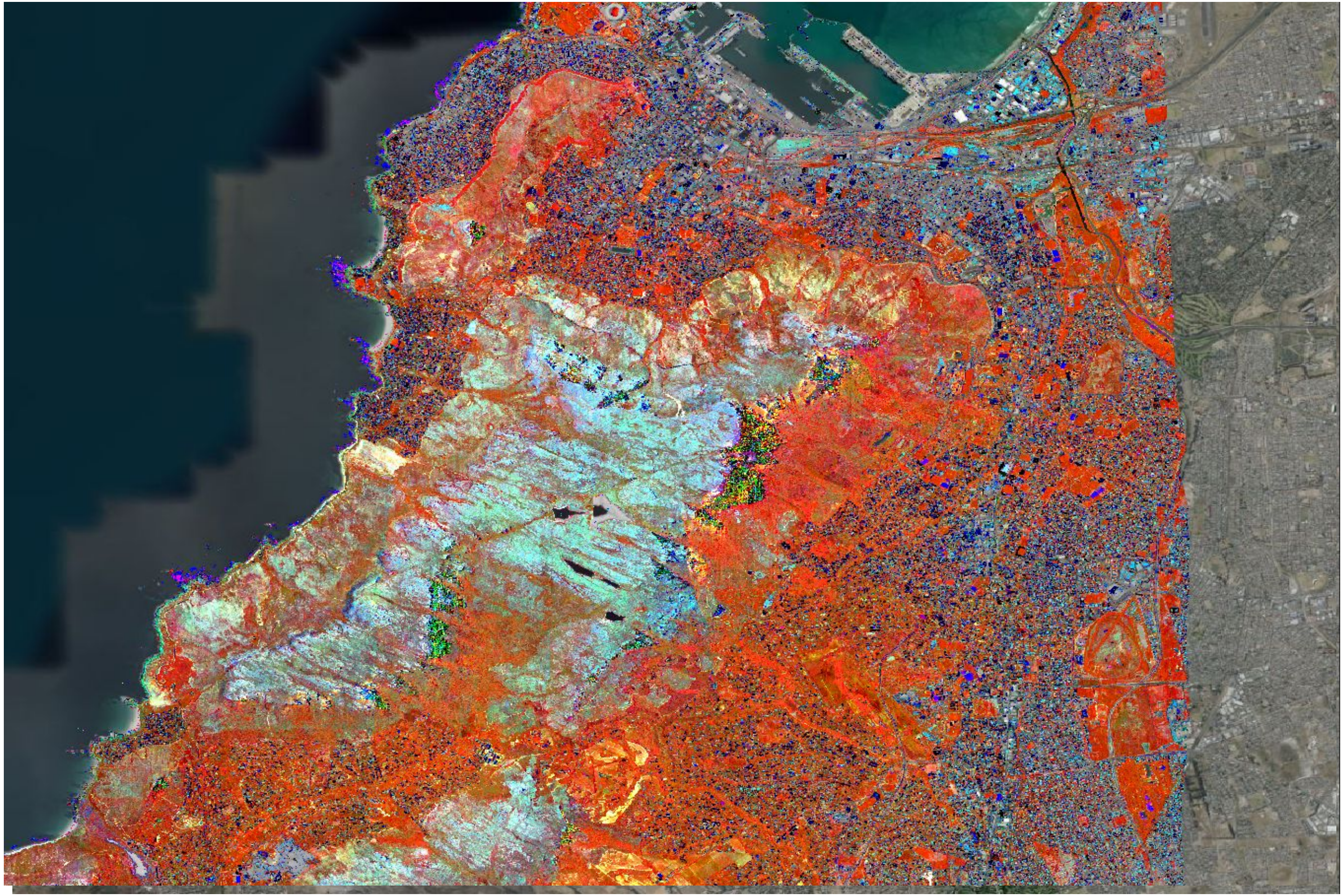
Table Mountain

Ternary map of the same traits as before over Table Mountain

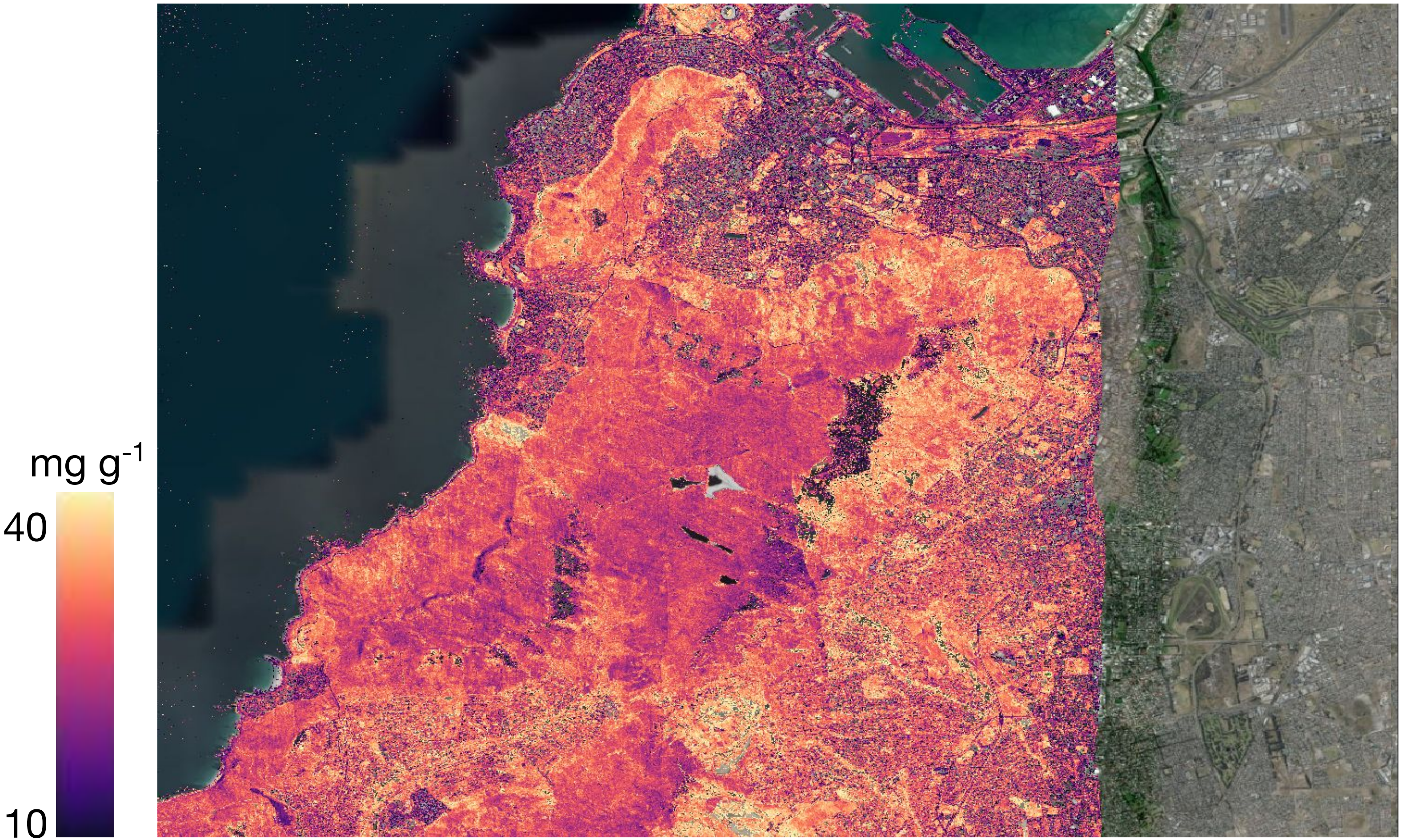
Red = nitrogen

Green = NSCs

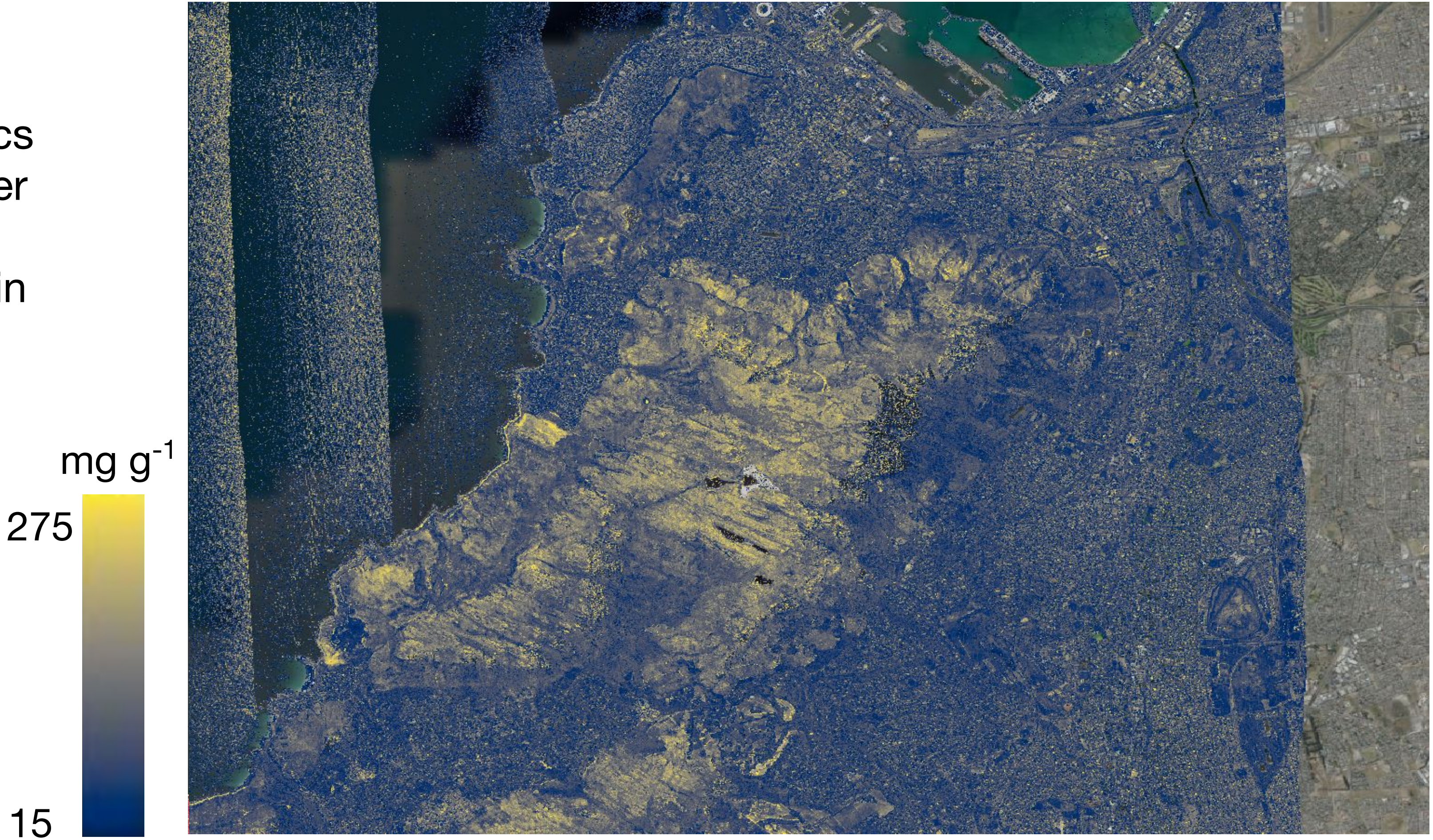
Blue = Phenolics



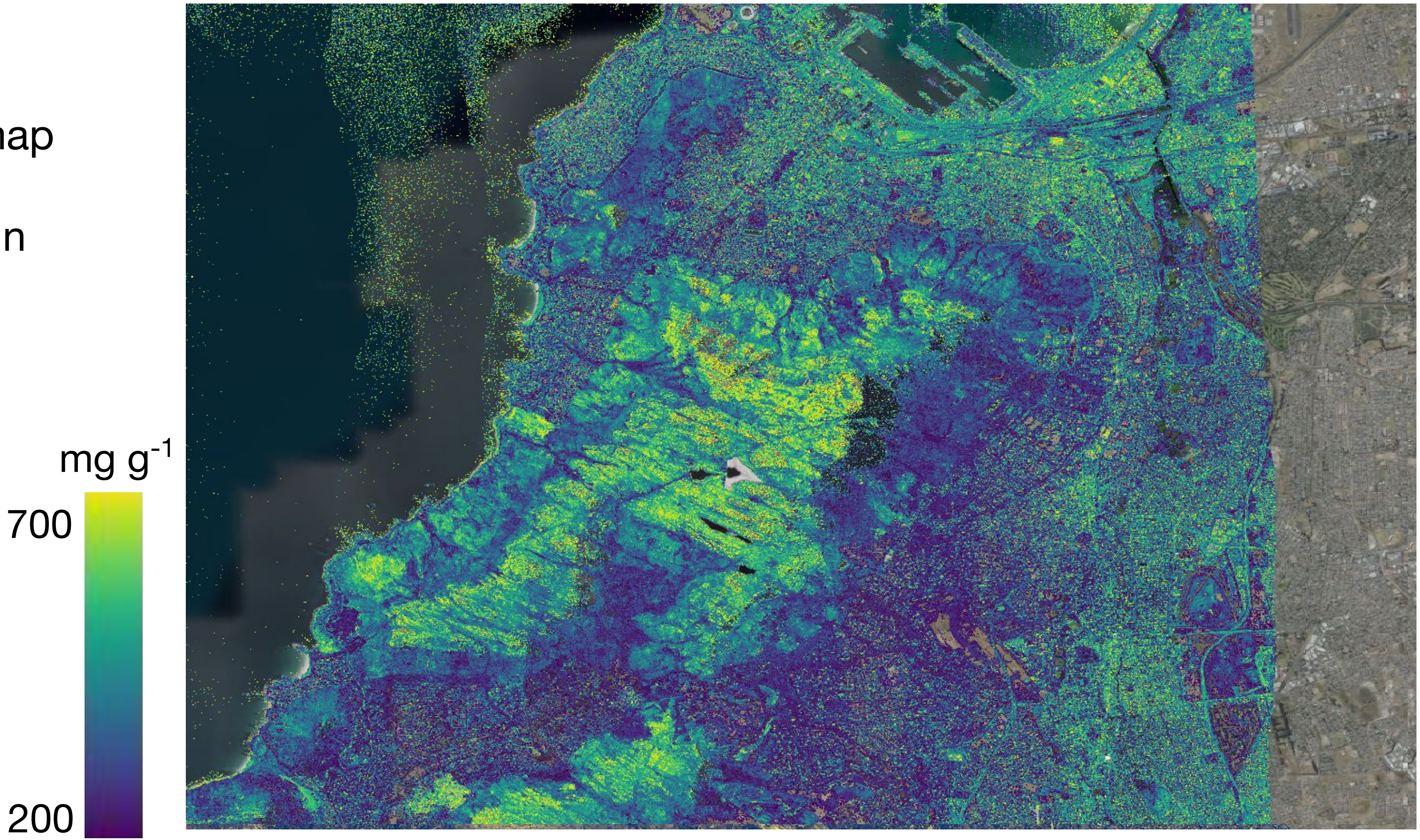
Nitrogen map over Table Mountain



Phenolics map over Table Mountain



NSCs map of Table Mountain



True Color

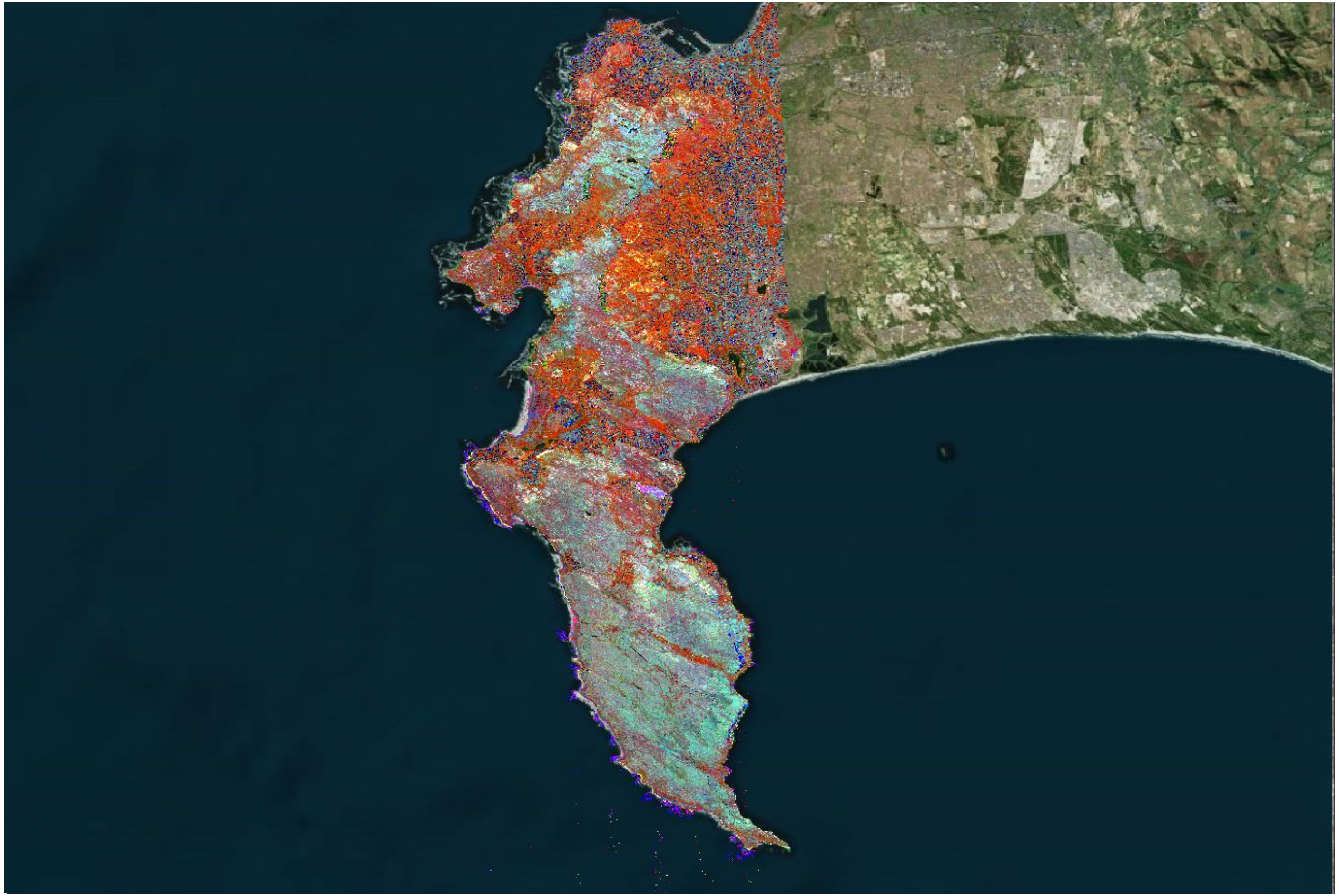
Cape
Peninsula

Ternary map of the Peninsula

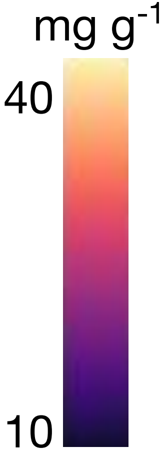
Red = nitrogen

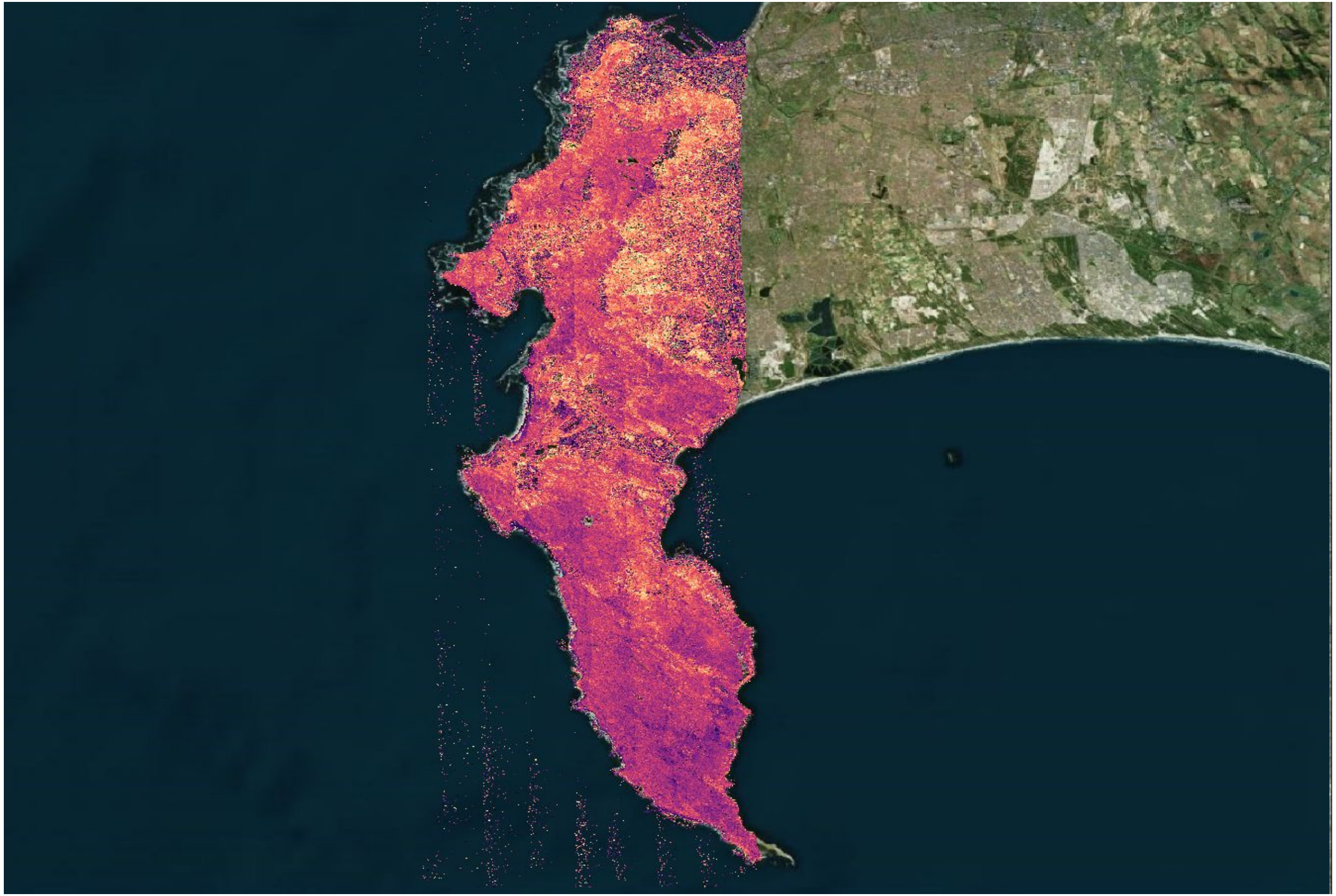
Green = NSCs

Blue = Phenolics

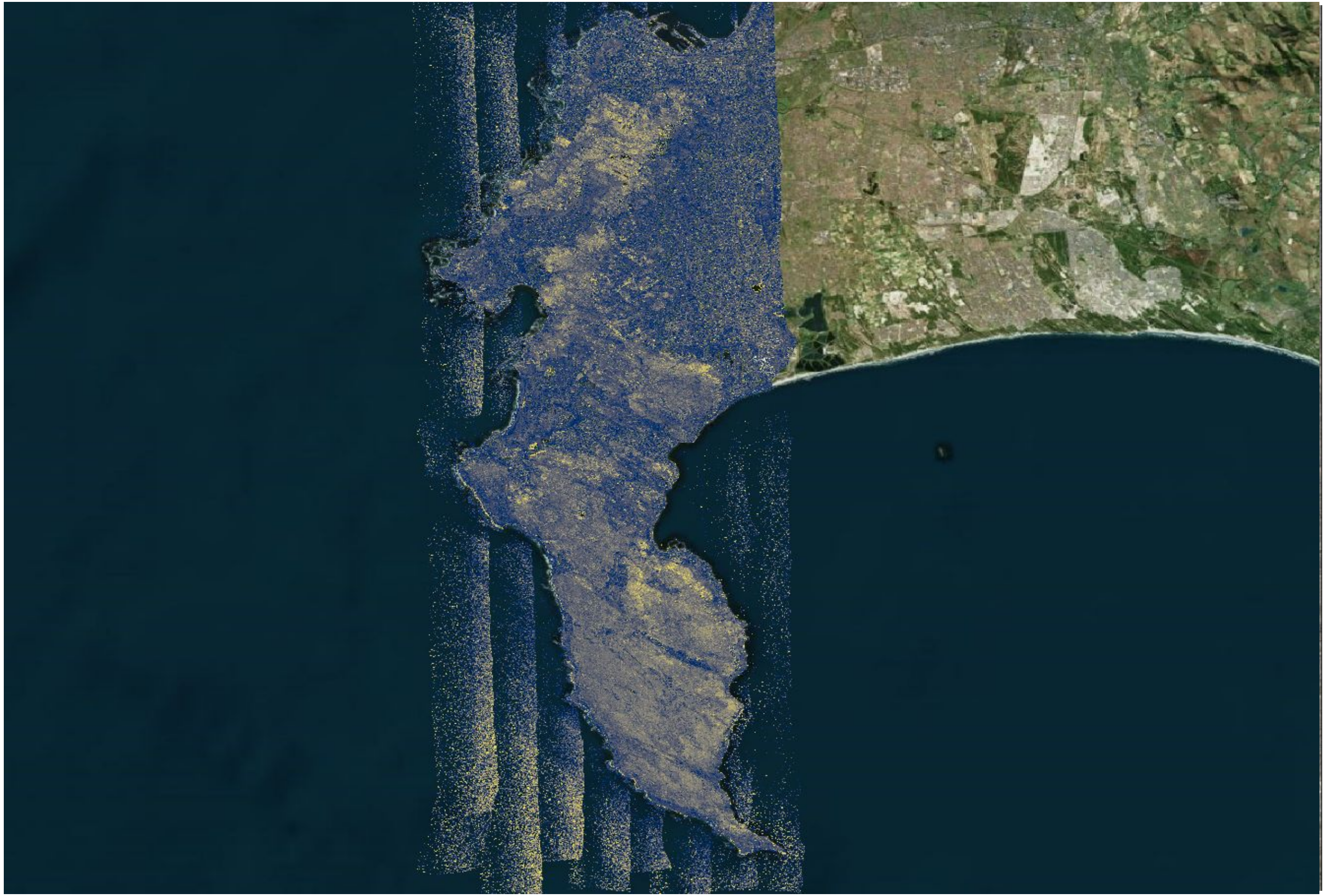


Nitrogen
map of the
peninsula

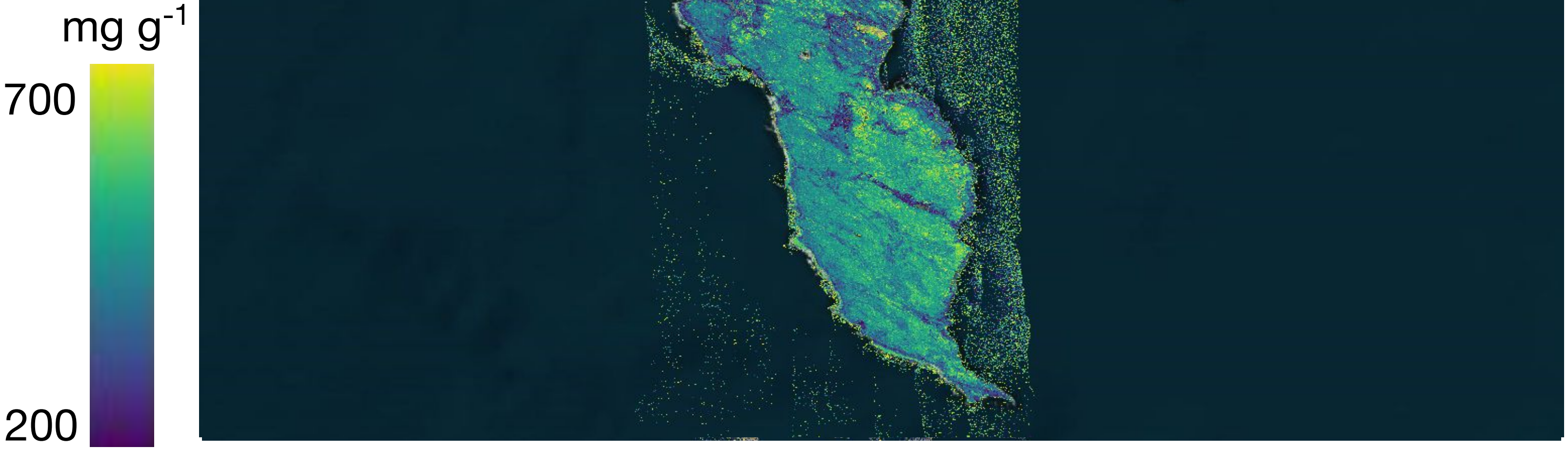




Phenolics map of the peninsula



NSCs map of the peninsula




I HAVE QUESTIONS
LOTS OF QUESTIONS