Erasmus+

COSI Computational Colour and Spectral Imaging

Master in Computational Colour and Spectral Imaging (COSI)

UNIVERSITY OF EASTERN FINLAND · UNIVERSIDAD DE GRANADA · UNIVERSITÉ DE LYON · UNIVERSITÉ JEAN MONNET SAINT-ÉTIENNE · NTNU Norwegian University of Science and Technology

Bryophyte Bioindicator Estimation with SVMs

Master Thesis Report

Presented by

Austin Ryan English

and defended at the

Norwegian University of Science and Technology

September 2022

Academic Supervisor(s): Dr. Semenov, University of Eastern Finland, Finland
Host Supervisor: Dr. Semenov, University of Eastern Finland, Finland
Jury Committee:

1. Dr. Sole, Norwegian University of Science and Technology, Norway
2. Dr. Shrestha, Oslo Metropolitan University, Norway

Submission of the thesis: 10$^{\text{th}}$ August 2023
Day of the oral defense: 4$^{\text{th}}$ September 2015

Abstract

Often viewed unfavorably as inhospitable and complex land, peatlands, in recent decades, have risen in status as significant, long-term carbon reservoirs, acting as vital counterbalances against global atmospheric carbon emissions. Advances are being made in methods for global monitoring of peatland ecosystems using multimodal photogrammetry, supporting peatland protection and restoration efforts by quantifying peatland extend, composition, and relative health. Here, some of the state-of-the-art metrics, models, and methods for assessing peatland status and peatland species health and resilience to climatic changes are reviewed. Furthermore, a new dataset is assessed with some of these methods and novel techniques for predicting the physical bioindicator parameters. Results show great performance of SVR models in predicting maximum photosynthesis values of peatland mesocosm specimens.

Glossary

$\overline{crsc}_{Chl.}$ the peak magnitude from $\overline{cr}_{Chl.}$ 47
$\overline{crsc}_{Chl.}$ the peak magnitude from $\overline{cra}_{Chl.}$ 47
$\overline{crsc}_{NIR}$ the peak magnitude from from $\overline{cr}_{NIR}$ 47
$\overline{crsc}_{NIR}$ the peak magnitude from from $\overline{cra}_{NIR}$ 47
$\overline{crsc}_{RG}$ the peak magnitude from from $\overline{cr}_{RG}$ 47
$\overline{crsc}_{RG}$ the peak magnitude from from $\overline{cra}_{RG}$ 47
$\overline{cr}_{LD}$ the peak magnitude from from $\overline{cr}_{LD}$ 47
$\overline{cr}_{LD}$ the peak magnitude from from $\overline{cra}_{LD}$ 47
$\overline{r}_{MTVI2}$ ((560nm 670nm 800nm)) the MTVI2 value derived from $\overline{r}_{SD}$ 47
$\overline{r}_{MTVI2}$ ((560nm 670nm 800nm)) the MTVI2 value derived from $\overline{a}_{SD}$ 47
r_{MTVI2} ((560nm 670nm 800nm)) the integral of MTVI2 values over $\overline{r}_{SD}$ signals 47
r_{MTVI2} ((560nm 670nm 800nm)) the integral of MTVI2 values over $\overline{a}_{SD}$ signals 47
S. balticum (S_{balt}) *Sphagnum balticum* a generalist bog moss species 61
S. magellanicum (S_{mage}) *Sphagnum magellanicum* a characteristic fen species 61
S. recurvum coll. (S_{recu}) *Sphagnum recurvum coll.* a generalist fen moss species 61

absorbance signals absorbance signals derived in this project
  $\overline{a}_{SD}$ derived absorbance from $\overline{r}_{SD}$ 47
  $\overline{cra}_{Chl.}$ continuum removal of $\overline{r}_{Chl.}$ 47
  $\overline{cra}_{LD}$ continuum removal from $\overline{r}_{LD}$ 47
  $\overline{cra}_{NIR}$ continuum removal from $\overline{r}_{NIR}$ 47
  $\overline{cra}_{RG}$ continuum removal from $\overline{r}_{RG}$ 47
absorbance spectral products spectral products composed of portions from the mesocosm reflectance factor spectra
  SDA 47
acrotelm Upper layer of live plants in a peat bog 2

coefficient of determination (r^2) the proportion of variation in a dependent variable that is predictable from an independent variable 58, 59, 64, 71
continuum removal Normalization processing technique where a ratio is taken of the signal to its convex hull VI, 50, 52, 53

end member todo 16

index of agreement (r^2) the proportion of variation in a dependent variable that is predictable from an independent variable 59, 64, 71
index of agreement (d) todo 16

pigments 15, 16
648-719 nm ($\mathbf{cr}_{648-719\ \mathrm{nm}}$) continuum removal of reflectance signal from $\mathbf{r}_{648-719\ \mathrm{nm}}$ 15, 16
648-719 nm ($\mathbf{r}_{648-719\ \mathrm{nm}}$) Reflectance signal from Cab pigments VI, 15, 16
708-848 nm ($\mathbf{r}_{708-848\ \mathrm{nm}}$) Reflectance range exhibiting TWC and LD phenology 15, 16
708-782 nm ($\mathbf{r}_{708-782\ \mathrm{nm}}$) Reflectance range exhibiting TWC/LD phenology but excluding the NIR plateau VI, 15, 16
711-778 nm ($\mathbf{cr}_{711-778\ \mathrm{nm}}$) Continuum removal of $\mathbf{r}_{708-782\ \mathrm{nm}}$ 15, 16

Acronyms

fPAR fraction of photosyntetically active radiation absorbed in photosynthesis 12

AIC Akaike's Information Criterion 18

Cab Chlorophyll a and b 14, 15, 16

CCRF conical-conical reflectance factor 19, 41, 52, 80

CDR Carbon Dioxide Removal 1

CIm Modified Chlorophyll Index 20

CMF color matching function 24

CWT Continuous Wavelet Transform 17, 18, 79

EVI Enhanced Vegetation Index 20

GHG Greenhouse Gas 1, 13

GPP gross primary product 12, 18, 20, 48

HDRY history of low water table conditions 5

HIST Peat Site History 15

HVS human visual system 24, 25

HWET history of high water table conditions 5

IPCC Intergovernmental Panel on Climate Change 11, 14

IR Infrared light spectrum 44

LAI Leaf Area Index 4, 48, 49, 51, 64, 80

LD Leaf Density 15, 16

LUE Light Use Efficiency 12

MTVI2 Modified Triangular Vegetation Index 2 4, 16, 49, 50, 51, 64, 71, 80

NDVI Normalized Difference Vegetation Index 20, 48

PAR photosynthetically active radiation 12

PRI Photochemical Reflectance Index 20

Ra autotrophic respiration 13

RBF Radial Basis Function 54, 62, 64

Re total ecosystem respiration 12

Rh heterotrophic respiration 13

RMSPE root mean squared percentage error 27

RVI Relative Vigour Indicator 14, 15, 16

S1 image capture session 1 6

S2 image capture session 2 6

S3 image capture session 3 6

S4 image capture session 4 6

SIPI Structurally Insensitive Pigment Index 20

SITE Peat Site Location (fen, bog, hummock, etc 15

SNR signal to noise ratio 8, 32, 41

SPD spectral power distribution 31, 38, 80

SPEC Moss specie 15, 17

SVI Spectral Vegetation Index 16, 17

SVM Support Vector Machine 15, 54

SVR Support Vector Regression 15, 54, 57, 61, 62, 64, 71

TVI triangular vegetation index 50

TWC Turf Water Content 15, 16

UAV Unmanned Aerial Vehicle 16

UV Ultraviolet light spectrum 44

UV-VIS Ultraviolet and visible light spectrum 44

VI vegetation index 4, 20, 48, 49, 72

VIS Visible light spectrum 44

VNIR Visible and Near-Infrared 15, 17

WLD water-level-drawdown 5, 62

WTL water table level 12, 19

X

Contents

1 Introduction

Peat is a common term for organic soils in the histosol soil classification. Peatlands are generally considered to be any region of wetland containing this organic soil, usually with actively growing vegetation (i.e., grasses, mosses, shrubs) contributing to the peat formation. The conditions under which peat develops are unique. Still, the soil can be found in many climates and regions across the world, from the Amazon Rainforest to northern Europe and Australia, with estimates putting the terrestrial land share of peatland at 2.84% Xu et al. (2018).

Although commonly viewed unfavorably as untraversable and complex land, peatlands have been found useful, historically, due to the fuel resource they provide with their rich organic soil; once the land is drained of water, the ground can be mined out and used as a primitive fuel or as an organic ingredient in various products and soil amendments. Additionally, peatlands and other wetlands give way to rich agricultural farmland after being drained; the high organic content in the waterlogged soil is rich in nitrogen and used to grow root and salad crops.

With an international focus on climate change mitigation, peatlands are receiving special attention for their pivotal roles in carbon emissions. When peatlands are drained of water, they become substantial Greenhouse Gas (GHG) emitters, but in their natural waterlogged state, they are agents of terrestrial-based biological Carbon Dioxide Removal (CDR). This dual nature of the peatland biome being unloosed by human intervention is not uncommon among vegetative systems (Hicks Pries et al., 2013, 2015) and simply results from an imbalance in the growth and decay of vegetation biomass. Still, uniquely in peatlands, human interference through land-use changes has exacted an overarching toll on global peatlands. These changes though are more straightforward to reverse when compared with other artificial GHG sources.

After having been dried, mined, and abandoned, the rewetting of peatlands has been found effective in reestablishing the primary peat-producing genus, *Sphagnum* González et al. (2014). This is crucial, as *Sphagnum* species are uniquely responsible for maintaining the peatland's hydrological system to function as a terrestrial carbon sink van Breemen (1995). And so, via rewetting, peatland restoration has become a focus, and the need for unified, large-scale peatland monitoring is growing.

Source: P199, CC BY-SA 3.0, via Wikimedia Commons

Figure 1.1: *Typical undisturbed peatland bogs in Canada with a dense Sphagnum acrotelm.*

1.1 Peatland mapping

As part of international carbon budget modeling and resource and initiative planning, monitoring peatland cover and carbon stock on a global scale has become a crucial necessity. Considering that peatlands store such vast reserves of carbon, preserving still-pristine peatlands and restoring currently damaged and drained peatlands may underpin climate change mitigation initiatives. Monitoring peatlands on a global scale requires scalable methods. Well-established practices in botany and biology exist for thoroughly and extensively studying and monitoring peatland behavior, composition, and status; however, these traditional methods require much manpower and expertise, leaving them somewhat unfeasible for the time being for continuous global monitoring. Therefore, monitoring methods which utilize remotely sensed data are likely to form the most effective basis for global peatland monitoring (Lees et al., 2018; Minasny et al., 2019).

Peatland must first be identified and mapped through some standardized classification as a prerequisite to monitoring. Most global peatland maps are based on disparate local soil classification systems or formed as composite maps from several land use and soil classification maps (Minasny et al., 2019). This is somewhat problematic for a few reasons:

1. Soil classification systems are typically organized around agronomy rather than ecology; they usually group all organic soils into a single classification (histosols), sometimes with subgroups for differentiating on soil acidity and permafrost content. Peatlands would be a subgroup, ill-defined in such a

system where topology and hydrology are not considered. These organic soil classes, however, are commonly used to estimate peatland coverage.

2. Soil classification systems are disparate, with many countries having their own system suited to their gown geology; unifying data from multiple soil system databases can be challenging and can produce erroneous estimates. The prevailing international system, the World Reference Base for soil resources, is a good start to globally unified soil classification and hence more accurate peatland classification. Yet, it is still not well suited for the detailed classification of peat soils and the hydrological states influencing their carbon flux.

Global peat area estimates vary between 3.3 and 4.6 million km^2, with the corresponding estimates for crucial properties such as carbon stock, peat thickness, and carbon density ranging much as well (Minasny et al., 2019). Though several global peatland mapping estimates exist, there is significant variance among the estimates; this is likely due to the use of the source mentioned above data which is not explicitly intended for peatland classification.

Several classification systems are tailored to peatlands and their critical characteristics; these vary in scope and cover domains such as topology, morphology, chemical composition, botanical origin, and physical characteristics (Andriesse, 1988). These classification systems provide a more accurate accounting of peat soil decomposition, thickness, and water availability, each of which would provide more detailed input to peatland mapping estimates; however, these systems are not readily available in global datasets like the more fundamental soil classification systems. And so, detailed peatland mapping is left in a somewhat fragmented state, with each country producing peatland estimates of varying levels of detail, quality, and uncertainty.

According to Minasny et al. (2019), peatland mapping studies should focus on quantifying the uncertainty in peatland predictions, validating the results, and using more than one covariate for making the predictions. Additionally, Minasny et al. (2019) raised a few research questions, one of which pertains to this project: "How do current global peatlands respond to climate change?" As will be explained in the sections below, this project deals with a moss mesocosm dataset which contains peatland vegetation species reflectance data as a dependent variable of drought-induced stress.

1.2 Mosses and bryophytes: crucial species

Mosses are members of the bryophyte plant group: non-vascular, seedless plants that lack water distribution regulation capability within their tissues and systems.

While they are not the sole constituent of the peatland ecosystem, mosses are a vital component nevertheless; they are so crucial that wetland classification systems based on moss abundance have been proposed after having found that mosses served as essential predictors of historical peatland environmental conditions (Vitt and House, 2021). Additionally, bryophytes are known to be key predictors of biodiversity (Zechmeister et al., 2003; Pardow et al., 2012) This brings up interesting possibilities for detailed global wetland classification based on moss or bryophyte abundance. Vegetation mapping via remote sensing tools is thoroughly explored. However, classifying wetlands via moss abundance would come with new challenges as mosses exist not only in open peatland fields but also under forest canopies where remote sensing methods do not have direct views of moss turf. Furthermore, the interference of fluctuating water levels with vegetation reflectance signals can pose issues with continuously monitoring even open peat accumulating turfs. Furthermore, existing vegetation indexs (VIs) require modifiers to accurately quantify standard metrics for vegetation ecosystems like Leaf Area Index (LAI) for the non-vascular species of peatlands.

1.3 Research Questions

In this project, a new hyperspectral dataset of moss mesocosms collected from peatland areas of differing hydrological histories is available to work with. Considering the previously discussed status of global peatland mapping and the connections between peatland bryophyte species and wetland ecosystem states, this project sets out to answer, in part, the questions:

1. How do peatland bryophyte species respond to climate stresses?

2. Do hyperspectral data and spectral products derived thereof to serve as accurate covariate predictors of health during climate stress?

3. Are absorbance signals advantageous to reflectance when modeling biochemical parameters from spectral data?

4. How do vegetation indices such as Modified Triangular Vegetation Index 2 (MTVI2) compare to spectral data in predictive ability?

5. Can moss bio-resilience be predicted from hyperspectral data?

1.4 Dataset Background

This project is based on a hyperspectral image dataset captured during a controlled moss experimental study that began in 2018 as an extension to field studies of the same species under similar conditions (Kokkonen et al., 2019). The dataset was captured by faculty collaborators (Semenov, 2019), and has had some exploratory processing performed on it including:

1. thresholding segmentation of reflectance images
2. image annotation with experiment variable modifiers
3. image annotation with reference-patch and specimen locations
4. image organization and naming according to experimental design

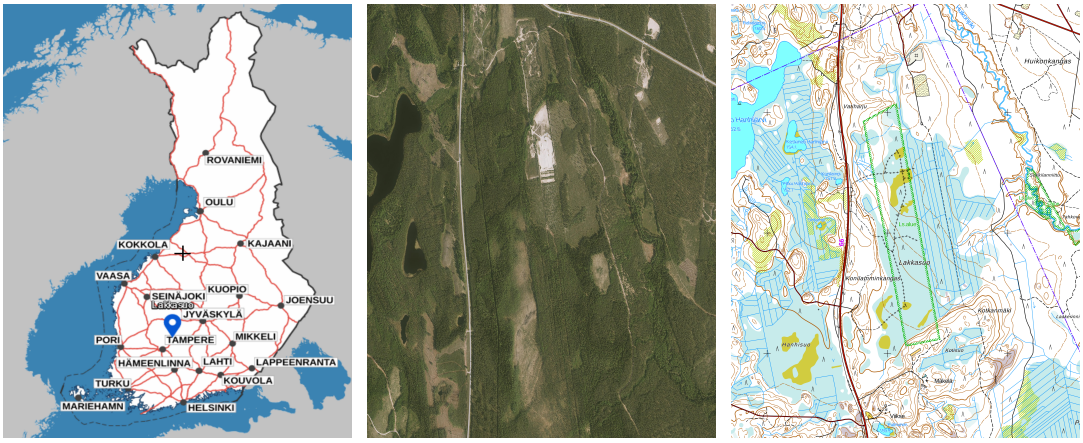
From this initial work on the dataset, what is utilized in this project includes items 2, 3, and 4.

In this section, the experimental background is reviewed to explain the purpose and potential applications of the dataset derived thereof, and the dataset, as used in this project, is further described in detail.

1.4.1 Mesocosm drought experiment

In the original experiment from which this dataset was produced, researchers studied the effect of peatland water-table history on dominant moss species' acute drought response. It was hypothesized that the moss vegetation cover from peatlands of long-term (17+ years) dry conditions and deep water tables would have adapted to drier conditions and therefore be more resilient to severe short-term drought than would moss vegetation from peatlands with higher water tables or a shorter history of dry conditions ($<$ 17 years) (Kokkonen et al., 2023). Researchers collected 96 representative moss mesocosm samples from 3 areas in the Lakkasuo peatland reserve (Fig. 1.2): a site with water-level-drawdown (WLD); a site with history of low water table conditions (HDRY); and a wet bog with a history of high water table conditions (HWET). The Lakkasuo peatland reserve where these samples were collected is located in southern Finland; the peatland is large, containing fens, bogs, a tree-cover area, and several experimental sites where controlled WLD has been applied over the years.

After collection in May 2018, due to the hot summer conditions, the samples were stored at 6°C until the start of the experiment in August 2018. In August, all the mesocosm samples were trimmed to even turf heights and acclimated



Source: National Land Survey of Finland Topographic Database 05/2023

Figure 1.2: *Lakkasuo peatland site where experimental specimens were extracted.*

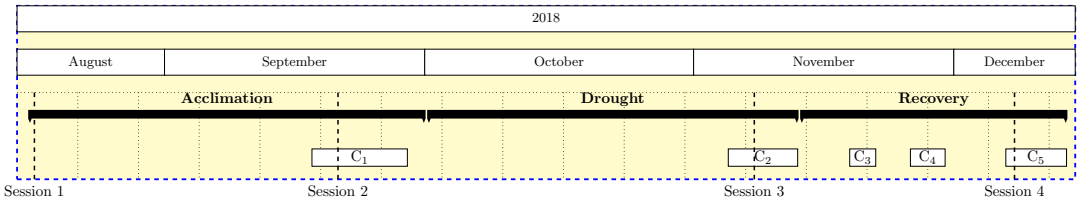
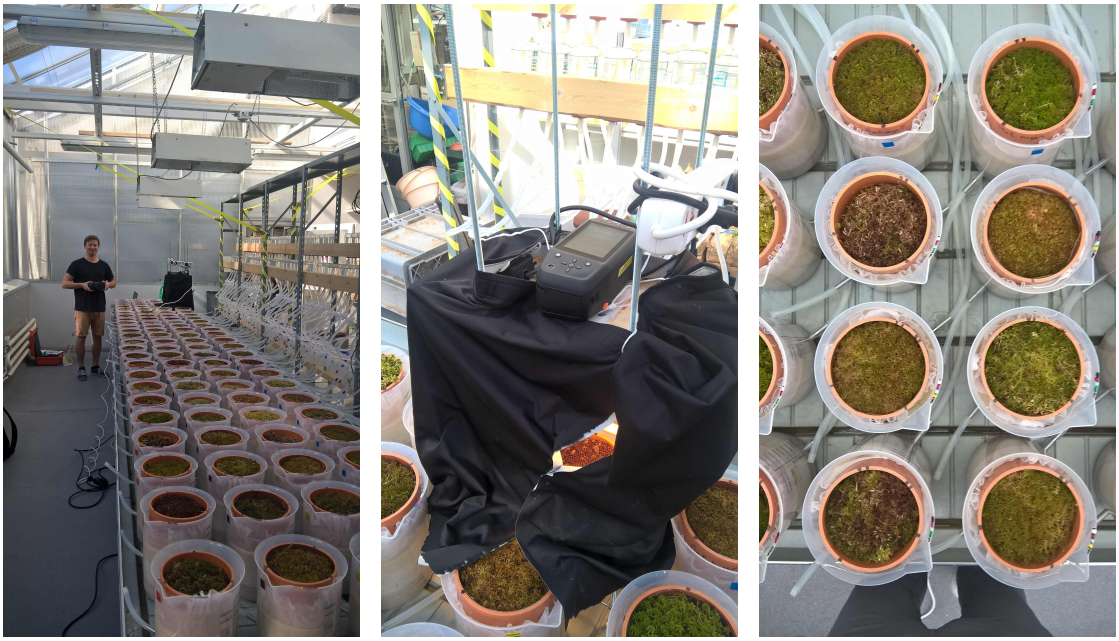


Figure 1.3: *Data capture timeline*

to standardized experimental conditions with plentiful irrigation (15 cm water table) using water profiles corresponding to the mesocosms individual collection sites. This period corresponds to image capture session 1 (S1). Similarly to the experimental setup in Malenovský et al. (2015), successive stages of treatment were then applied to the moss samples: drought/irrigation and recovery (Fig. 1.3), with the recovery phase being assessed at three intervals throughout. In total, there were five measurement campaigns (denoted by C_n in Fig. 1.3); C_1, C_2, and C_5 correspond to the successive phase's peaks while C_3 and C_4 were intermediate measurements during the recovery phase. Images were successively captured at measurement campaign C_1 , corresponding to image capture session 2 (S2), C_2, corresponding to image capture session 3 (S3), and C_5, corresponding to (image capture session 4 (S4)).

In the study, ninety-six wild moss mesocosms were collected, potted, and kept in a controlled environment (see Fig 1.4a), where a variable treatment could be applied to each sample. Treatment consisted of applying a varied water level to the moss samples: 10 cm deep for control specimens and 30 cm deep for variable specimens undergoing a simulated, short-term drought. Each moss sample was



(a) *Ninety-six mesocosms* **(b)** *Imaging enclosure* **(c)** *Varied mesocosms*

Figure 1.4: *Physical setup of moss mesocosms.*

assessed and imaged at intervals throughout the treatment stages. Researchers measured several biological properties and outputs from the specimens, and the hyperspectral images were captured for post-analysis and further studies. In total, there are 396 hyperspectral images recording the chemometric phenotyping of these samples as they progress from their initial state through drought and on through recovery at the end of the experiment. By the end of the experiment, some moss samples appear completely desiccated (although still alive), others increased in vitality, while the rest displayed a range of relative health in between depending on the species or other experiment modifiers.

As seen in the timeline, the hyperspectral image capture sessions 2, 3, and 4 roughly correspond to the peak phase measurement campaigns C_1, C_2, and $C5$. In contrast, image capture session 1 was taken at the beginning of the experiment shortly after collecting and placing all the specimens.

1.4.2 Hyperspectral Images

The hyperspectral images in this dataset were captured with the VNIR Specim IQ camera (see full manufacturer specifications (Viitakoski, 2023)). No spatial or spectral binning was used during image capture; therefore, each image has the camera's full spatial resolution of 512×512 pixels and full spectral resolution

consisting of 204 bands ranging from approximately 400 - 1000 nm. In all images, three diffuse gray tiles were placed around the perimeter of the mesocosm specimens to serve as ideal diffuse reflector references for performing corrections. Also, in each image, each specimen's modifier combination marks can be seen painted on the ceramic pot, classifying the specimen into appropriate site history, site location, treatment, and species categories.

Hyperspectral images were captured at four stages throughout the experiment (see Fig. 1.3) at specimen setup, peak acclimation, peak drought, and peak recovery. For the 96 specimens, there are a total of 396 hyperspectral images covering the experimental stages. The sessions are summarized below:

Session 1 (S1) : This is at the start of the experiment; wild specimens are recovered from storage, trimmed, and supplied with sufficient irrigation.

Session 2 (S2) : Captured at the peak of the acclimation phase, where all specimens consistently underwent favorable conditions.

Session 3 (S3) : Captured at the peak of the drought phase, half of the specimens underwent severe drought treatment while the other half remained in favorable irrigation conditions.

Session 4 (S4) : Captured at the peak of the recovery phase, regular irrigation was resumed for drought samples and continued for the control samples.

Due to the nature of hyperspectral image capture, the imaging scene configuration is essential. Perhaps the most critical is illumination. To ensure sufficient signal to noise ratio (SNR), samples must be illuminated with enough light taking care to not overexpose the samples at any point along the spectral domain. Additionally, when in-scene white or gray references are used, care must be taken not to overexpose the reference; in these images, gray references were used to allow for a longer exposure integration time, 20 time units (tu) (time unit is unknown, but assumed to be seconds), to pick up more signal from the specimens. If a white reference is used, it is sometimes necessary, as in this case, to capture it in a separate picture at a faster exposure time. Each session's white reference images in this dataset were captured with a 10 tu integration time.

To ensure a consistent lighting configuration and camera placement, researchers constructed an imaging scaffold to mount the lights and camera onto (see Fig. 1.4b). Using the scaffold ensured that the nonuniformities in light dispersion would be consistent from image to image, making calibration easier. It also ensured that the camera focal distance and physical distance to the specimen were constant. Around the scaffold, a diffuse shroud was attached; this was used to block out external light and provide a surface to scatter light inside the scaffold diffusely.

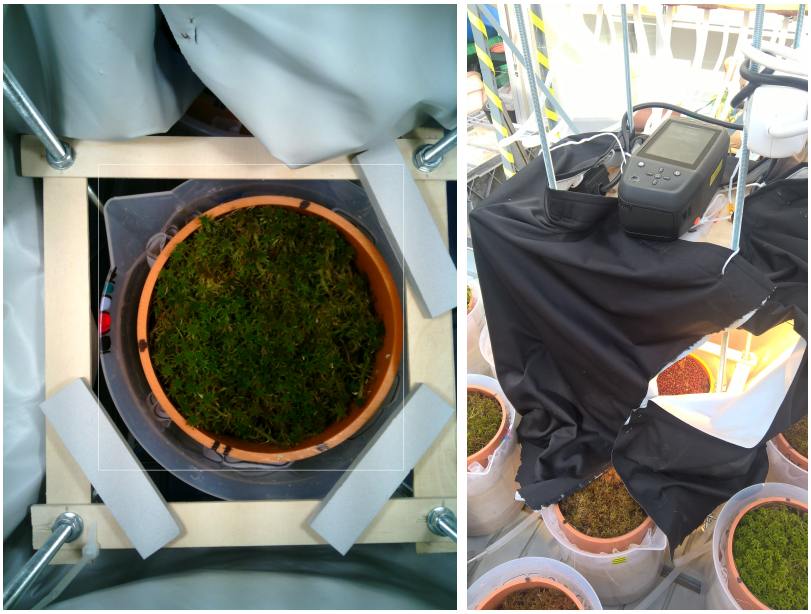


Figure 1.5: *Scaffolding structure and interior*

Inside the platform, the gray reference tiles were placed onto a wooden frame (see Fig. 1.5) under which the specimens were situated.

1.5 Research Goals

With this hyperspectral dataset and the corresponding mesocosm trait measurements and estimates provided from (Kokkonen et al., 2023), the primary goals of this project are to answer the research questions posed and contribute new insights and data to the peatland ecology research community. A brief overview of what work was performed in this project is provided below:

1. A versatile mesocsom dataset processing pipeline was built. This includes functions for performing all operations used in this project and described below with extensibility built in.
2. A basic hyperspectral image web inspector was built to support viewing of hyperspectral TIFF files,
3. The raw dataset was converted to spectral TIFFs for ease of handling.
4. All images were calibrated to a standardized unit.
5. Image noise was characterized and corrected.

6. A total of 22 spectral products were derived from the dataset and used to optimize 264 mesocosm trait prediction models, over two dozen or so with very high prediction performance.

7. Moss spectral features corresponding to climatic drought stress were identified and used in predicting spatial maps of moss drought response.

8. Absorbance was found to produce more performant bioindicator prediction models in certain cases than did reflectance signal.

2 State of the art

Numerous studies and reviews in ecology, remote sensing, and other fields have covered continuous advancements in peatland monitoring over the recent decades. Motivations for understanding peatland status and inventory are far-ranging from local agricultural resource planning up to national conservation efforts and beyond into intergovernmental mandates as part of Intergovernmental Panel on Climate Change (IPCC) carbon budget reform.

Studies on the use of remote sensing data for peatland diagnostics via moss species reflectance signals have been in the undertaking for decades (Bubier et al., 1997); yet recent reviews on the topic express frustration at not having reliable means within the peatland community for assessing global peatland status (Minasny et al., 2019) to have unified, global peatland carbon flux reports as inputs into decision-making policies and international carbon budget planning. Literature reviews on peatland mapping and assessment have highlighted the need for several improvements in the existing body of work to achieve these goals (Minasny et al., 2019; Lees et al., 2018):

1. reporting model uncertainty
2. using multiple covariants, if possible, for predictions
3. exploring the relationship of the peatland carbon cycle to climate variations
4. modeling broader sets of peatland variants rather than focusing only on specific hyper-localized peatlands
5. including split-parameterization for ecological models, which allow for different parameters to be applied within existing ecosystem models for vascular and non-vascular species
6. scaling micro-scale models into repeat-mosaic models, which can potentially serve as accurate predictors for peatlands considering their high heterogeneity at small scales and homogeneity at larger scales

Peatlands are challenging to model, particularly with remote-sensing methods, for a few reasons:

- There is a high variation in peatland ecosystem composition and behavior. Peatland ecosystems are commonly considered a single group, but many subclasses affect the biome response to climatic factors.
- Water table level, perhaps the single most influential driver of peatland ecology, is not directly accessible through remote sensing data. It has been shown, however, that water table level (WTL) can be inferred from moss reflectance data and other proxies (Burdun et al., 2020).
- In gross primary product (GPP) modeling of vegetation ecosystems, estimating the total ecosystem respiration (Re) is difficult and biased (up to 25% (Keenan et al., 2019)). Peatlands undergo drastic changes in Re due to degradation. Therefore, it is vital, yet difficult, to quantify this uncertain term globally for peatland areas. Mapping peatland health would aid in this.
- Peatlands have high heterogeneity at a microscale within their ecosystems; they are, however, homogenous at larger scales, known as repeat mosaics.

Due to these imposing challenges, primarily the inability to transfer existing ecosystem and vegetation carbon flux models directly to peatlands, estimations for peatland contribution to national and international carbon flux models are thought to be very roughly over or under-estimated. This is concerning for countries with underestimated carbon sinks from their peatland ecosystems and those with underestimated carbon emissions from their drained and degraded peatlands.

Global carbon fixation is primarily carried out through photosynthesizing organisms during photosynthesis; for this reason, carbon flux models are broadly defined around GPP, the measure of total radiant energy converted to chemical energy within an ecosystem (global or local). With photosynthesis being the primary driver of terrestrial GPP, carbon fixation as a process within photosynthesis is very closely related to GPP estimation. Convenient for remote sensing capabilities, the Light Use Efficiency (LUE) model is most often used to describe the relationship between GPP and photosynthetic processes within ecosystems:

$$GPP = fPAR \cdot PAR \cdot \epsilon \tag{2.1}$$

Of the total radiant energy available for photosynthetic activity, photosynthetically active radiation (PAR), some will be lost due to variables such as cloud cover, leaf area, and topology, reducing the PAR down to a fraction of photosyntetically active radiation absorbed in photosynthesis (fPAR). Of the total photosynthetic

radiant energy that finally impinges on photosynthesizing organisms, a fraction of that amount, ϵ, will be converted to chemical energy; the rest is re-emitted due to inefficiencies, limitations around light conversion, and ecosystem respiration. Accurately quantifying peatland and wetland soil respiration is particularly important as ecosystems release chemical energy through respiration, both autotrophic respiration (Ra) directly from vegetation and heterotrophic respiration (Rh) from other sources such as soil decomposition.

Peat composition specifications and peatland inventories are often derived from national soil data and taxonomy systems developed within national scopes by each country's respective efforts in soil science. At the global scale, soil science is very diverse, and due to how it typically develops around national agricultural initiatives, the resulting soil taxonomies, classification methods, and experimental methodologies are just as varied and disparate as the agricultural applications which the systems and processes are developed around.

The varied nature of peatland vegetation introduces challenges as well. Northern boreal peatlands are dominated by *Sphagnum* mosses, exhibiting as quickly recognizable open fens and bogs. In contrast, other peatland regions may show sedge, reed, and forest species characteristic of soil types different than peat. This makes peatland mapping through vegetation classification not so straightforward, as the vegetative cover is not necessarily a direct indicator of the presence of peat. Instead, the underlying hydrological disposition of the land drives peat formation (Rydin and Jeglum, 2013); key moss species, however, do serve as indicators of the underlying hydrological state of a wetland and could be used as predictors of peatland class and, therefore, an indicator of peat formation (Vitt and House, 2021). Mapping peatland species cover comes with additional challenges, such as dealing with the interference of surface water with vegetation signals and the location of bryophyte species underneath taller vegetation canopies; nevertheless, accurate monitoring across all peatland regions is vital to climate change mitigation initiatives, as some of the peatlands about which least is known and studied, such as those in the Congo basin (Dargie et al., 2017) may contain some of the highest amounts of carbon deposits. Peatlands are predicted to become major GHG emitters shortly (Leifeld and Menichetti, 2018; Harris et al., 2022); therefore, a focus on framing local peatland studies into the global context is vital for developing consistent methods for tracking and predicting international peatland contributions to carbon flux.

2.1 Key Research Directions

The following sections briefly review key studies addressing some of the current gaps in peatland and wetland monitoring. Based on recommendations from the

most recent reviews of peatland mapping and carbon flux monitoring, the following research directions are highlighted here:

1. Mapping peatland vegetation bioindicators, particular vital health indicators as they relate to climate response

2. Peatland characterization or classification via vegetation proxies

3. Quantifying peatland diversity

2.2 Mapping Peatland Bioindicators

Estimating peatland carbon flux and stock provides the knowledge needed for carbon budgeting; part of estimating ecosystem-wide carbon flux is assessing the biological status of dominant species in the ecosystem. Changes in vegetation chemical composition provide insight into their photosynthesizing ability; these chemical changes often significantly alter a plant's absorption of light, evident in reflectance measurements of vegetation cover, both from remote sensing data and field spectroscopy. Tracking changes in this informs restoration efforts regarding the effectiveness of deployed tactics and enables prioritization of targeted areas. However, peatland health is a vague measure characterized by many metrics. Considering Sphagnum species and their critical role in peat production, the following studies focus on quantifying Sphagnum health responses to climatic stresses.

2.2.1 Field-scale moss vigor with VNIR

In response to studies and reports of changes in temperature, wind speed, and ozone across Antarctica, the IPCC recommended in 2009 to establish ongoing datasets for the continent. Prompted by this and in anticipation of plans from the Scientific Committee for Antarctic Research to establish research around ecosystem health in Antarctica, Malenovský et al. (2015) performed a study on Antarctic mosses, developing models for predicting moss bioindicators with high-resolution spectral signals, and produced a Relative Vigour Indicator (RVI) with the results.

Although Antarctica is not known for expansive vegetation, green coastal ecosystems do exist around the continent (Fenton, 1980; Loisel et al., 2017), and they are particularly fragile due to the extreme, niche conditions to which they have adapted. This study focused on two moss beds on rocky soil areas in a coastal region near the Australian Casey Station. The authors hypothesized that changes in moss bioindicators due to drought stress (Chlorophyll a and b (Cab),

Table 2.1: *Experiment overview*

Datasets	Primary (2013), 32 specimens (3 species, two collection sites) Auxiliary (1999), 73 specimens (1 specie)
Independent var.	watering regimen
Dependent var.	lab-measured Cab, LD, TWC
Covariates	HIST, SPEC, SITE
Correlates	VNIR ($\mathbf{r}_{496-719\text{ nm}}$, $\mathbf{r}_{648-719\text{ nm}}$, $\mathbf{cr}_{648-719\text{ nm}}$, $\mathbf{r}_{708-848\text{ nm}}$, $\mathbf{r}_{708-782\text{ nm}}$, $\mathbf{cr}_{711-778\text{ nm}}$)
Methods	SVR modeling
Scales	leaf, field

Leaf Density (LD), and Turf Water Content (TWC)) would induce changes in vegetation reflectance along the Visible and Near-Infrared (VNIR) range. With enough correlation between the bioindicators and respective regions in VNIR, linear models could be trained off of collected samples, and the results combined into a composite RVI with which to identify and track wild moss bed health in the Antarctic.

2.2.1.1 Methodology

An experimental overview is summarized in Table 2.1. The primary dataset used in this experiment was built from 32 moss specimens collected onsite.

In the laboratory, specimens were divided into a control and an experimental group. The control group received the regular irrigation regimen throughout the experiment, while the experimental group underwent two stages: a six-day drought and a twenty-five-day recovery in which the specimens were irrigated the same as the control group.

On the last day of each stage, measurements were taken, including the ground truths (Cab, LD, TWC) as well as spectroscopic readings covering the respective spectral ranges (*Correlates* in Table 2.1), used later for modeling correlations between spatial signals from field imaging to the specimen attributes.

After running the treatment applications over 27 days (6 days of drought and 25 days of recovery), SVR prediction models were set up for each of the three specimen bioindicators, using the spectroscopy-derived reflectance signals as predictor variables, and the measured bioindicator parameters as observed responses. For each bioindicator, Malenovský et al. (2015) used three Support Vector Machine (SVM) variations. For the Cab bioindicator, separate models were

set up for $\mathbf{r}_{496-719\text{ nm}}$, $\mathbf{r}_{648-719\text{ nm}}$, and $\mathbf{cr}_{648-719\text{ nm}}$. For TWC and LD, separate models were set up for $\mathbf{r}_{708-848\text{ nm}}$, $\mathbf{r}_{708-782\text{ nm}}$, and $\mathbf{cr}_{711-778\text{ nm}}$ each.

2.2.1.2 Results

After training on data collected in their experiment as well as on the data contained in Robinson et al. (2005) and performing model tests, Malenovský et al. (2015) observed that the following prediction variables produced the best bioindicator predictions: $\mathbf{cr}_{648-719\text{ nm}}$ for Cab, $\mathbf{r}_{708-782\text{ nm}}$ for LD, and $\mathbf{cr}_{711-778\text{ nm}}$ for TWC. The best model for TWC, however, had abysmal prediction accuracy with an index of agreement, d, of .43, and the authors concluded that this was most likely due to lack of full radiometric signal related to TWC in their spectroscopic measurements.

The LD and Cab models were then used to evaluate hyperspectral images of the same moss beds where the samples were collected. Bioindicators were predicted for every moss bed pixel after masking out non-vegetation pixels using MTVI2 thresholds. A resulting RVI map was then created with the mean of each normalized bioindicator.

2.2.2 Landscape-scale chlorophyll and nitrogen prediction with VNIR

One of the most challenging aspects of accurately monitoring peatland carbon cycle flux is the heterogeneity of the vegetation cover. Satellite imagery is typically captured at a scale where dozens or more species will be included in a single pixel. Spectral unmixing can help extract spectral end members. However, little work has been done on this for peatland vegetation species. Therefore, at landscape scales, the high heterogeneity in the typical peatland renders the standard Spectral Vegetation Indexs (SVIs) used in many ecological landscape studies inaccurate, eliminating SVIs as a convenient measure of peatland vegetation bioindicators without special modification for the many variations of vegetation response. Even at finer scales, such as imagery captured with Unmanned Aerial Vehicles (UAVs), the species-specific SVIs for peatland species is inconsistent across the seasons as water table levels and peatland canopy reflectances change.

Kalacska et al. (2015) took a comprehensive approach to this problem. In the Mer Bleu peatland bog in Ottowa, Canada, they collected data from 20 assorted peatland vegetation species from which neural network prediction models were trained to evaluate chlorophyll and nitrogen content. Data included physical chlorophyll and nitrogen content estimations based on field sample chemistry and reflectance measurements, $\mathbf{r}_{325-1075\text{ nm}}$.

Table 2.2: *Kalacska et al. (2015) experiment overview*

Dataset	890 data pairs (19 species, 190 specimens, temporally repeated over six months)
Independent var.	seasonal climate factors
Dependent var.	chlorophyll and nitrogen content
Covariates	SPEC
Correlates	VNIR (various SVI and CWT)
Methods	nonlinear neural network
Scales	leaf, field, aerial
Accuracy	(Chl., nr.) field: .87, .73; aerial: .86, .89

2.2.2.1 Methodology

Table 2.2 summarizes a brief experimental overview. Given the challenge of predicting common bioindicators across a vast collection of peatland species with a single metric, spectral reflectance, Kalacska et al. (2015) chose to employ neural networks and Continuous Wavelet Transform (CWT). Neural networks are non-linear machine learning systems beneficial for dealing with very diverse and sufficiently sampled datasets; in this case, this approach works well for the number of samples collected (up to 60 per species) and the number of primary covariates affecting the prediction (19 species and multiple seasons). The CWT provides a method of decomposing signals into base component coefficients. Various base wavelets can be used for the transform, and certain ones work better than others depending on the application (Cocchi et al., 2003).

Applying the CWT to a one-dimensional signal yields a two-dimensional result (scalogram) of n wavelet coefficients for every input sample, with the value of the coefficient corresponding to the scaled base wavelet's correlation to the original signal at the sample location. This method can be used to parameterize spectral signals into a fingerprint identifier for various materials or vegetation species Kalacska et al. (2007). Figure 2.1 demonstrates the CWT transform of a vegetation specimen's reflectance signal. The resulting scalograms could be analyzed for correlation between coefficients and specimen traits if the transform was applied across a training set of specimen reflectances. The most highly correlated coefficients across the scalograms could be used as prediction variables for other data samples.

Kalacska et al. (2015) performed this same analysis with all the specimen reflectance CWTs, finding which coefficients correlated the most to chlorophyll and nitrogen levels for each month that the samples were collected. Coefficients selected by correlation were further reduced, preserving only ones present in at least five of the six trial months. The remaining coefficients could proxy for

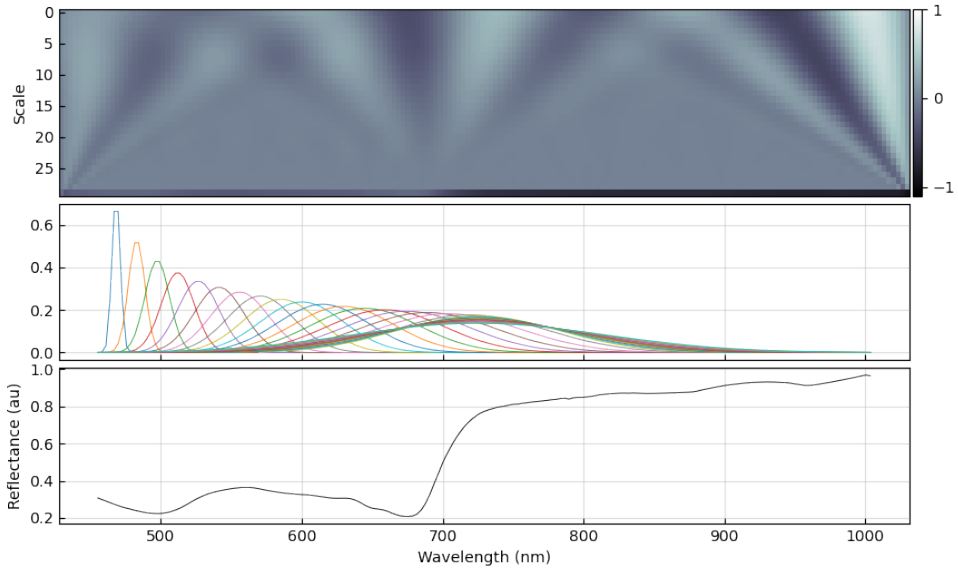


Figure 2.1: *CWT of a moss vegetation sample's reflectance signal. Top row: scalogram of CWT coefficients; middle row: wavelet bank with 31 scales; bottom row: moss reflectance.*

underlying chlorophyll/nitrogen phenology standards across seasons and species. These remaining coefficients were finally filtered further with Akaike's Information Criterion (AIC) to preserve the best modeling coefficients and used as prediction variable inputs into the neural network.

2.3 Peatland Characterization via Vegetation Proxies

This section reviews various works exploring the characterization of certain peatland aspects like habitat and GPP based on radiometric data from peatland vegetation.

2.3.1 Species, habitat, and moisture estimation from VISNIR and SWIR

Peatlands can be classified based on their hydrologic disposition, with water nutrient availability being the primary differentiating factor:

Table 2.3: *Salko et al. (2023) experiment overview*

Dataset	360 samples (90 specimens, repeated at four intervals over drought period)
Independent var.	length of drought
Dependent var.	moisture content
Covariates	species, habitat
Correlates	VIS-NIR, SWIR
Methods	PCA, linear-regression, bicanonical reflectance indices
Scales	leaf
Results	+90% reflectance explainability from species, $R^2 > .8$ moisture prediction with SWIR indices

1. minerotrophic: access to nutrient-rich groundwater, surface water, and precipitation
2. ombrotrophic: access to only precipitation; usually have a layer of peat so thick that access to groundwater is inhibited
3. eu-,meso-, oligotrophic: subclasses representing decreasing levels of mineral availability

These classifications would be very useful for improving global peatland carbon estimates as they provide insight into the nature of vegetation growth and water availability in a peatland; inferring these classes from remotely sensed data may be possible, as it has been shown that WTL can be estimated from peatland vegetation reflectance data (Burdun et al., 2020). WTL for a given peatland area drives nutrient availability, species composition, and vegetation response to radiation. In a recent study, researchers examined the possibility of inferring the peatland site, among other properties, from SWIR and NIR vegetation reflectance data (Salko et al., 2023).

2.3.1.1 Methodology

They collected 90 samples from several peatland sites: minerotrophic, ombrotrophic, and intermediate areas with a range of nutrient levels between the first two classes. For each sample, laboratory spectroscopic readings were taken at four intervals throughout the experiment: 0 hours, 24 hours, 48 hours, and seven days from the start of the experiment. Mean species conical-conical reflectance factor (CCRF) was computed, and results were analyzed for intra- and interspecific differentiation in signal between species and peatland site. This data was analyzed to assess the

variance explainability of species and habitat and to compute optimal bicanonical reflectance indices for predicting moisture content.

2.3.1.2 Results

In this study, results showed that peatland vegetation species explained more of the variance in reflectance data (over 90%) while peatland habitat, or hydrological classification, explained under 65% of the variance at the beginning when moss samples were freshly collected and under 50% after one week of drying. In contrast, the species explained more of the variance as drying progresses, with interspecies reflectance curves differentiating more in NIR and SWIR than in VIS bands. Furthermore, they found that in the ombotrophic samples, intraspecies variation in reflectance was minimal, while mesotrophic and intermediate habitat areas had high interspecies reflectance variance. They concluded that results indicate the importance of moss species spectral libraries in classifying peatlands from remote sensing data. Peatland vegetation species respond differently to radiative signals than vascular species, and the intraspecies responses also vary.

2.3.2 Peatland GPP and water content estimation

With similar aims as the researchers of the experiment from which this project's dataset was taken, Lees et al. (2020) attempted to estimate peatland vegetation GPP and water content using lab and field captured hyperspectral data. They compared the effectiveness of multiple existing water content indices, finding them correlated more to the lab measurements but suffering in accurate predicting the lower water levels in field measurements.

For GPP estimation, they studied the correlation between existing VIs and GPP estimates from carbon dioxide flux measurements. They found that the VIs, Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), Photochemical Reflectance Index (PRI), Structurally Insensitive Pigment Index (SIPI), and Modified Chlorophyll Index (CIm), were strongly correlated with peatland vegetation GPP.

3 Methodology

This chapter details the methods and processes used in developing this project. While not all methods are discussed, the primary stages of the project's processing chronology are covered.

3.1 File Preparation

Before processing the dataset images and extracting spectral features, the dataset files were first prepared with several goals in mind: convenience of inspection, digital storage efficiency, and format and signal standardization.

3.1.1 Data file format

The Specim IQ camera saves its spatial radiance data to raw image files accompanied by ENVI-compatible metadata header files. Using the ENVI-compatible format for the images is common across remote sensing applications. As this project was performed with no advanced hyperspectral imaging visualization workbench tooling (e.g., ENVI, ArcGIS), the ENVI-compatible images were converted to another format to make visualization and sharing of the photos convenient and cross-platform. The TIFF image format, created for storing and transmitting scanned documents, has come to be used in scientific applications around remote sensing and geographical mapping, with NASA developing the GeoTIFF extension format for use with geographically annotated raster images. TIFF files support "subdocument" or "multi-page" structures making them particularly useful for scientific imaging where human-vision inspired, three-channel RGB structure is not always used.

At the start of this project, the ENVI-compatible raw image files were converted to multi-page TIFF format files. Doing this accomplished several things: spectral band metadata could be consolidated into the image file; an RGB image preview could be stored in the TIFF for easy previewing on most operating systems; there is added convenience for processing the image data as TIFF parsers exist

in most software languages and image processing libraries; any dark frames or white reference images for a given specimen capture could also be consolidated into the same image file with the specimen capture. This reduces several related files for each hyperspectral capture to one and reduces the complexity of file handling scripting code (offloading the complexity to TIFF parsers). Also, masks and other related image processing outputs can be stored in the multi-page TIFF. This format was advocated by professors in the Computational Spectral Imaging department at UEF for some of the reasons above, and for the expanded compatibility it provides with existing image management systems in medical imaging and other fields.

Many hyperspectral images were converted, processed, inspected, and managed daily (not just the 384 specimen hyperspectral images but also their derived images). Some free software exists for basic visualization and pixel "Z profile" inspection of hyperspectral images, though they are often designed to support the ENVI formatted images or sometimes GeoTIFF images.

Most operating systems will preview the first page of a TIFF image in their included image viewer. Still, more capable image viewers are usually needed to page through a multi-page TIFF file, where each page stores a single slice of the hyperspectral image cube. Some free TIFF viewers that sufficed for this and were used in this project are IrfanView and FastStone Image Viewer.

To the author's knowledge, no image viewers are designed to support hyperspectral image inspection on hyperspectral TIFFs such as these. Some allow viewing a single pixel value at the cursor's coordinates; hyperspectral image viewers offering inspection tools do not support the TIFF format. Open-source GeoTIFF viewers either did not support basic hyperspectral Z profiling or depended on additional system tools to be installed.

A MATLAB app, Hyperspectral Viewer, is available for essential viewing and inspection. Still, it would need some custom connector code to read the TIFF and provide the hyperspectral cube to the app to work with these hyperspectral TIFFS. Additionally, MATLAB is not free to use.

Javascript is not typically used for intensive computing and image inspection; however, a good Javascript GeoTIFF parsing and manipulation library exists, which can read these hyperspectral TIFFS as well. With this library, a basic web app was created to support quick Z-profile inspections of hyperspectral images throughout the project. It may be just as easy to write a Python script to read the TIFF file, preview it with *matplotlib* and plot the spectra from a selected point in the image, but having a viewer available as a web app removes many system dependencies. Only a modern web browser is needed to run the application.

The main view of the application is shown in Figure 3.1. Here, a single channel of the chosen hyperspectral image is selected with the band slider and rendered in the Viridis colormap. A pixel from the moss specimen is also selected and

subsequently plotted in the right-hand side plot.

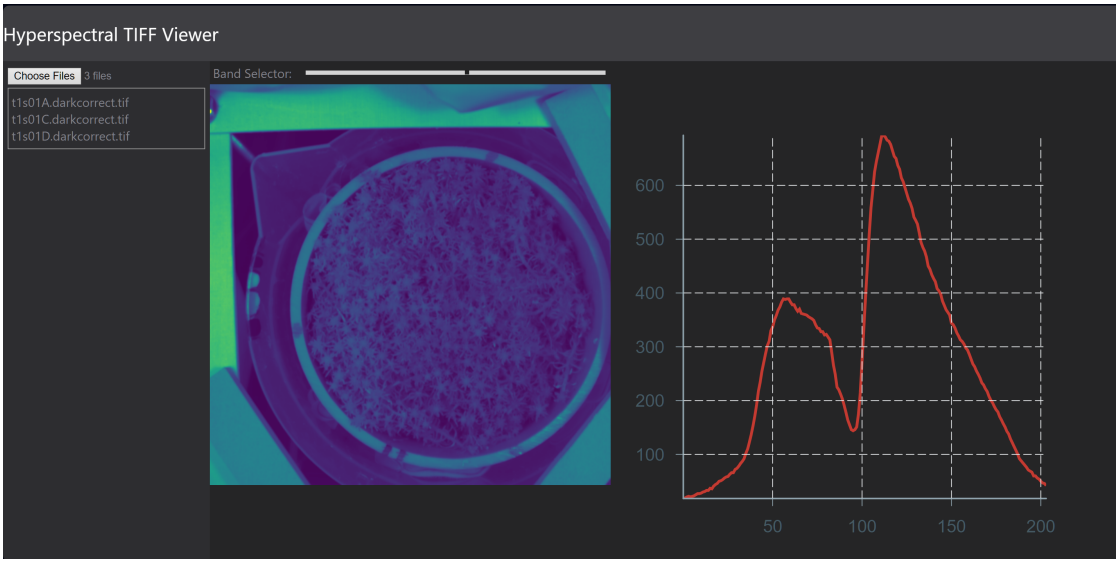


Figure 3.1: *Hyperspectral TIFF Viewer application built in order to inspect spectral TIFF images in this project's dataset conveniently*

3.1.2 Hyperspectral preview rendering

A benefit of using TIFF files to store the hyperspectral images is that the multi-page feature can keep all of the hyperspectral band images independently and a color-rendered preview image for the operating system to use as the file preview. RGB previews are often derived most simply by extracting three channels from the hyperspectral cube that somewhat represent the primary colors (e.g., 650 nm for red, 550 nm for green, and 450 nm for blue) and normalizing them from 0 to 1. Any band most convenient for the application can be used (even just a single band for grayscale). However, for this project, sRGB previews were rendered for all dataset images (some derivative spectral images used other custom color rendering schemes). Complete sRGB rendering details can be found in the original specification (Anderson et al., 1996); the three steps to convert hyperspectral reflectance images into an sRGB rendered preview are shown briefly here.

The XYZ tristimulus values for D65 illumination are first computed by integrating the respective color-matching function with the light source and derived reflectance signal from the hyperspectral images for each pixel from 360 nm to 830

nm as in (3.1).

$$\mathrm{K} = \int_{360\mathrm{nm}}^{830\mathrm{nm}} \mathrm{CMF_K}(\lambda)\, l_{\mathrm{D65}}(\lambda)\, r(\lambda)\, d\lambda \tag{3.1}$$

$$\tag{3.2}$$

where:

K : tristimulus value (X, Y, or Z)

$\mathrm{CMF_K}$: color matching function for tristimulus value K

$l_{\mathrm{D65}}(\lambda)$: D65 illuminant

$r(\lambda)$: reflectance signal

$$\tag{3.3}$$

Once the tristimulus values for a pixel are known, the linear sRGB values are computed with the standard sRGB transformation matrix as shown in (3.4).

$$\begin{bmatrix} R_{lin.} \\ G_{lin.} \\ B_{lin.} \end{bmatrix} = \begin{bmatrix} 3.2406 & -1.5371 & -0.4986 \\ -0.9689 & 1.8758 & 0.0415 \\ 0.0557 & -0.2040 & 1.0570 \end{bmatrix} \begin{bmatrix} X_{D65} \\ Y_{D65} \\ Z_{D65} \end{bmatrix} \tag{3.4}$$

Linear sRGB values must be gamma encoded to appear natural on the typical display monitor. Gamma encoding for sRGB is applied according to (3.5).

$$C_P = \begin{cases} 12.92\, C_{lin.}, & C_{lin.} \leq 0.0031308 \\ 1.055\, C_{lin.}^{-1.24}, & C_{lin.} > 0.0031308 \end{cases} \tag{3.5}$$

$$\tag{3.6}$$

where:

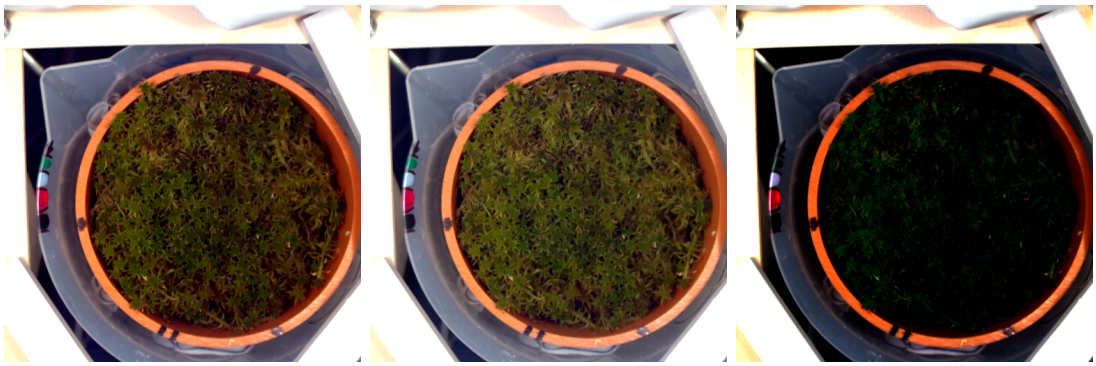
C is R, G, or B

$$\tag{3.7}$$

Finally, the saturation and contrast on the sRGB images are increased slightly to help improve detail visibility.

Rather than using the 1931 CIE color matching functions (CMFs) in (3.1), other kinds of vision systems can render the hyperspectral images. For example, a set of camera sensitivities could be used in place of the CMF to obtain the XYZ values as the camera would render them. However, any other vision system besides the human visual system (HVS) represented in the CMFs will require a separate XYZ to RGB transformation matrix. This can be computed through linear regression between XYZ values from the camera system's results in (3.1) and $RGB_{lin.}$ values from the CMFs results in (3.4).

Though somewhat out of this project's scope, this was tested on some hyperspectral images to compare the rendering results between a camera system and the HVS. Figure 3.2 shows renderings of dataset sample *T1S01A* using the HVS, an iPhone SE sensitivities, and extracted hyperspectral bands, each with an auto-contrast being applied afterwards. Using the same rendering techniques, optimal illuminations could additionally be designed to maximize the specimen features of interest Bartczak et al. (2017); English (2023).



(a) *iPhone SE rendered* **(b)** *HVS rendered* **(c)** *Bands (650, 550, 450) nm*

Figure 3.2: *Alternate previews of the hyperspectral image of specimen T1S05B*

In hyperspectral previews shown throughout the project, some are intended to solely represent the specimen and scene according to how the HVS would perceive the scene, and others represent radiometric values in a spatial image. Unless otherwise noted, dataset previews will represent the specimen and scene with a simulated HVS rather than representing radiometric data.

3.1.3 Data decorrelation and compression

While hyperspectral imaging provides unmatched spectral resolution, it comes at the cost of enormous image sizes and somewhat redundant data. The presence of a single chemical, for instance, chlorophyll *a*, will produce changes across nearly one hundred nanometers or more in the spectral dimension of a hyperspectral image. There is typically a target wavelength at which a peak of absorption or reflection caused by a particular chemical can be found; for many use cases, data at these peak wavelengths can explain most of the variation between highly sampled radiometric signals. This affords the everyday use of "spectral indices," which use a limited set of spectral readings obtained from multispectral cameras to spatially estimate specific chemometric properties of materials or qualitative characteristics of landscapes from remote sensing data. A pitfall of relying on a small set of

spectral readings is that unknown or unexpected covariants may be influencing the lessons at target wavelengths without any indication of this from other spectral data outside the captured sample set; therefore, for multispectral imaging and spectral indices, it is essential to understand the intended application for any particular index as well as the expected environmental constraints.

With an abundance of spectral data in a hyperspectral image, various methods exist for reducing the data redundancy while retaining target signals and potentially unexpected interfering signals from system covariants. One such method in everyday use is principal component analysis (PCA). PCA is the Eigen-value decomposition of the covariance matrix for a normalized dataset. If a dataset spans over N dimensions, then a principal component analysis will find P components through P=N new dimensions, all orthogonal to each other, of which the first component maximizes dataset variance along its axis, the second maximizes the remaining dataset variance along its axis, and so on. In a typical PCA implementation, the number of vector dimensions, P, can be chosen arbitrarily and, following the solution, reduced to the minimum number of vectors needed to explain the desired amount of variance in the system. If sorted in decreasing order by their corresponding eigenvalues, the principal components will be ranked in order of decreasing amount of explained variance, meaning that the first several components will explain most of the dataset variance. In contrast, the last ones will likely represent minor influences in the dataset, such as noise or minority signals.

PCA is frequently used to decorrelate data as a preprocessing step in building machine learning models. Although untypical, PCA can also be used as a means of compressing image data; the images can be reconstructed for rendering by projecting the P components back into the original dimensions by taking the dot product of the transformed image pixel values with the P components (and adding the dataset mean if the dataset was normalized before PCA). In this project, PCA is utilized, among other methods, for creating spectral products from which to train chemometric prediction models. An added benefit of the PCA spectral products in comparison to others is that they contain nearly all of the original dataset's information at a fraction of the size, while other spectral products eliminate much of the presumably extraneous data; with PCA, this allows for both a reduction of dataset size for machine interpretation as well as compression of the original dataset signals for human performance throughout the project.

A series of analyses on one image in the dataset was performed to understand the tradeoffs with using PCA to represent the original data. A PCA with varying components was used to transform the image into new data along a reduced set of principal dimensions. The result is a new image cube for each PCA iteration which can be stored with the same spectral tiff format used for the hyperspectral images; in addition to the transformed data, the principal components, image

data mean, and original hyperspectral sample wavelengths are stored with the new image cube to allow for a reconstruction of the original data, with varying degrees of reconstruction accuracy depending on the number of components used. The resulting PCA data cubes were then used to reconstruct the original image, and the moss specimen signal was visualized to inspect any distortion trends. The results are shown in Fig. 3.3; spectra from one specimen pixel are plotted for each of the reconstructions alongside the original uncompressed pixel signal. File sizes of the resulting PCA cube TIFFs are plotted with the signal root mean squared percentage error (RMSPE) concerning the uncompressed signal, revealing the inverse relationship between the number of components used to represent an image and the resulting error.

Other commonly used image compression algorithms, such as JPEG, could also be used to reduce spatial correlation in the images. Decorrelated data from PCA can be used directly in this project's chemometric predictive models; however, compressed spatial data from the JPEG algorithm can be used for compression only. Nevertheless, JPEG formats of dataset images were created using varying levels of JPEG compression quality (Q_n where n varies from 5 to 100 in increments of 5). The JPEG format does not support floating point data; to use JPEG compression, reflectance factor images were converted from floating point into 12-bit integers and then compressed. In Fig. 3.4, signal distortions caused by compression are visualized along with the RMSPE and file sizes associated with each format. It is evident that the RMSPE and file size are generally inversely related to each other; compression levels down to 90 can remain under 1% RMSPE. However, it is essential to note that these are statistics taken from one pixel within the image's specimen region. Taking the average of as few as 20 pixels within this region brings the RMSPE of the JPEG compression formats down so low that compression quality levels as low as 20 can remain under 1% RMSPE.

In terms of file size reduction, using JPEG compression on the dataset is then comparable to using PCA when taking signal averaging into account, where a JPEG compressed dataset image at a quality level of 20 is about 5% of the full uncompressed size while the PCA data, using six components to explain more than 99% of data variance, is about 3.5% of the uncompressed hyperspectral image file size. Without actually performing the compression, it could be assumed that applying the same JPEG compression level of 20 to one of the PCA images could further reduce the image size down to .175% of the uncompressed image size, about 350 KB, with some additional compounded error.

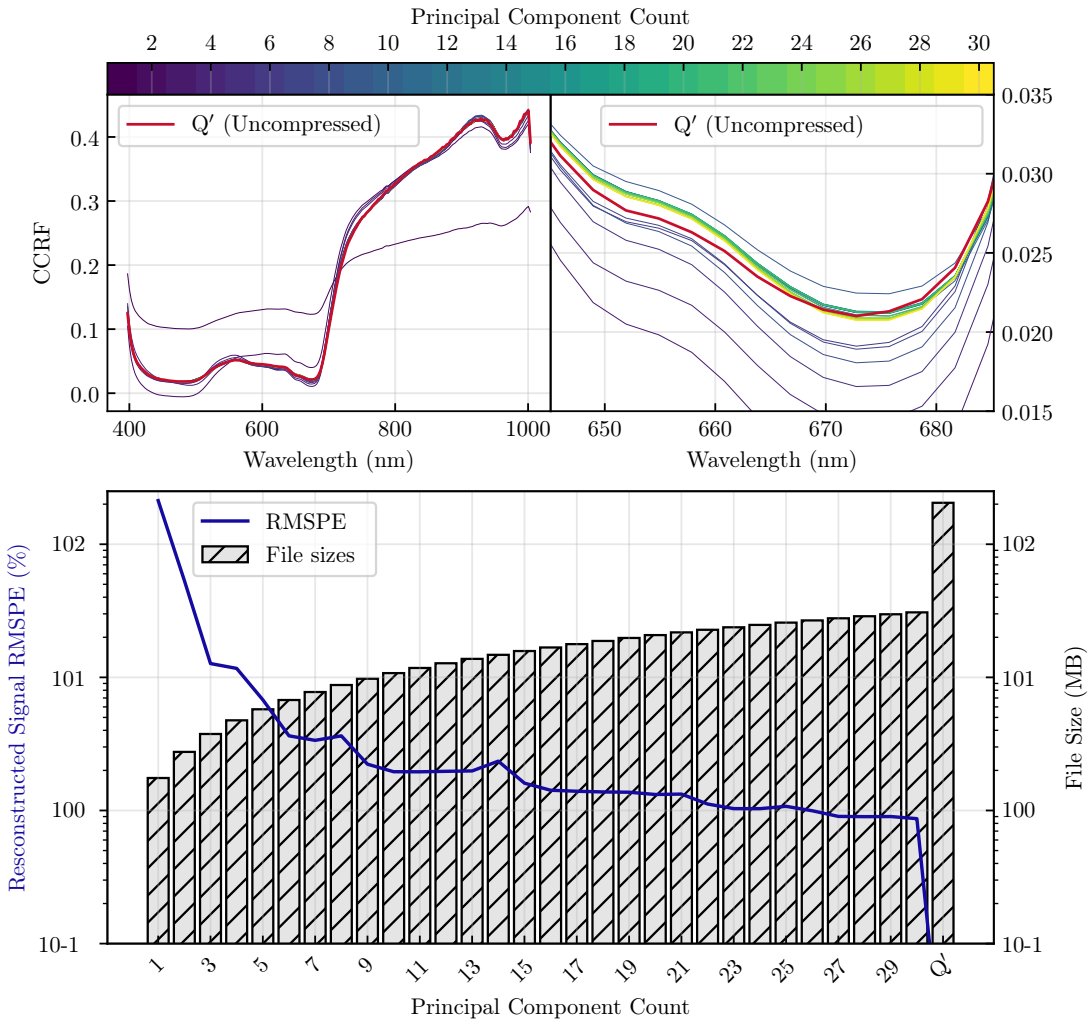


Figure 3.3: *Signal distortion and file size reductions from using PCA as compression. Compressed signals shown are from images reconstructed using 1 through 30 principal components, sorted in order of decreasing variance explained, and the original floating point reflectance image. Data are extracted from 10 pixels located within the specimen region of the image. Top left: reflectance factor plot of each compression level. Top right: crop of the top-left plot. Bottom: RMSPE and file size for each compression level.*

3.2 Hyperspectral Image Calibration

Hyperspectral images are spatial measurements of radiant energy as a function of wavelength. Figure 3.5 illustrates a hyperspectral image as a cube, where the third

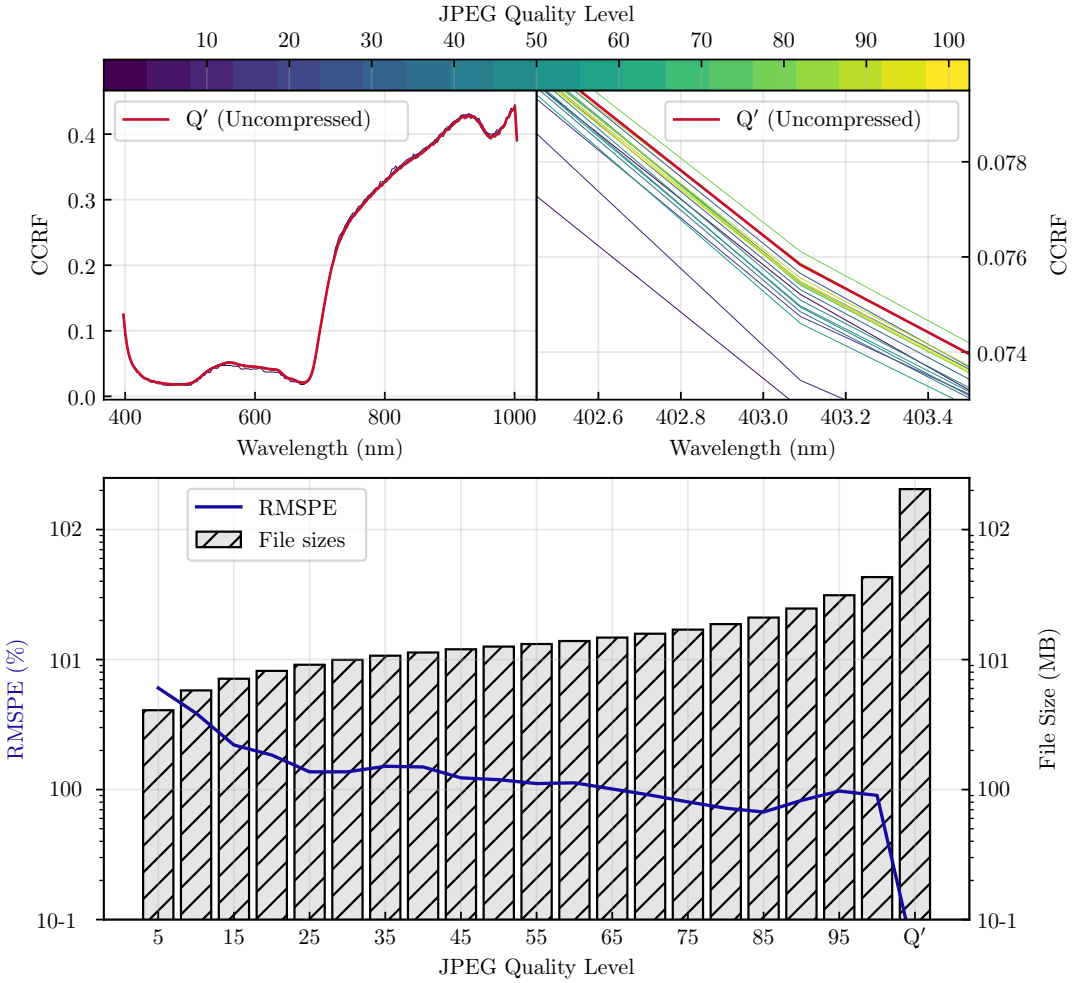


Figure 3.4: *Signal distortion and file size reductions from using JPEG compression on the hyperspectral images. Formats shown are JPEG quality levels 5 through 100 in steps of 5, the original floating point image, and its 12-bit quantization (scaled). Data are extracted from a single pixel located within the specimen region of the image. Top left: reflectance factor plot of each format. Top right: crop of the top-left plot. Bottom: RMSPE and file size for each format.*

dimension is the spectral domain containing pixel measurements as a function of wavelength. A raw hyperspectral cube can be represented as C_{ijk}, denoting the vertical and horizontal spatial axes with i and j, respectively, and the third spectral dimension with k. The hyperspectral image can conveniently be considered a stack of images where each image, C_{ij}, is the camera response at the respective band, k. As shown in Figure 3.5, value samples along the two spatial axes are typically

referred to as *samples* and *lines*. In contrast, according to the ENVI data format, spectral samples are usually referred to as *bands*.

As with any optical measurement instrument, it is beneficial to characterize the measurement device and calibrate measurement outputs into standardized units, considering device noise, optical defects, measurement error, and nonlinear effects. In raw output form, hyperspectral images are given in arbitrary *count* units (or *au*), as is the case with most digital images in raw form. However, images from some cameras may be floating point due to spectral or spatial binning and image averaging. The arbitrary count unit, *au*, represents the camera image sensor's digital quantization of the analog charge accumulated in the photodetector cells. It is not a direct measurement of radiance, reflectance factor, or standardized radiometric value. In fact, the same camera under the same spatial configuration and viewing angle for an invariant imaged material will produce varying raw outputs as the exposure time, scene lighting, and other environmental factors vary. For this reason, basing chemometric prediction models on hyperspectral data requires standardization of the specimen images into some unit of measure independent of the scene lighting and camera properties, solely representing the specimen properties.

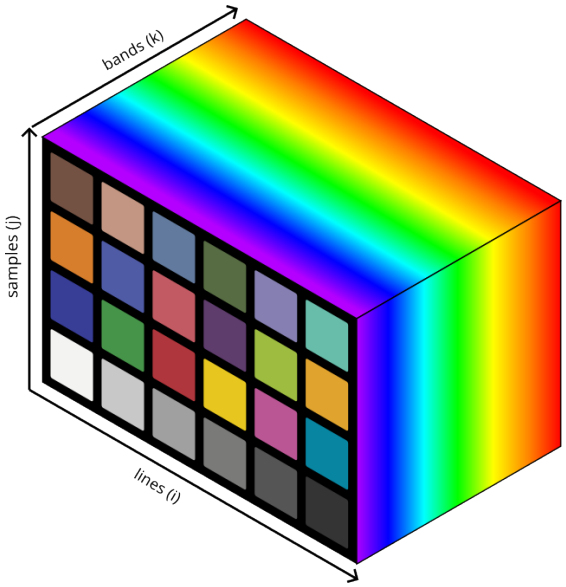


Figure 3.5: *Illustration of a hyperspectral image's three dimensions: two spatial axes and a spectral dimension.*

Hyperspectral image data is often converted into reflectance factor units, often simply denoted as reflectance units. While reflectance of a material is the ratio of its reflected light, r_s, to incident light, $\mathrm{r_n}$ (3.8), reflectance factor of a material

is the ratio of reflected light off of the material sample, r_s, to the reflected light off of an ideal reflector, r_w (3.9). Spectral reflectance factor then is the measure of reflectance factor as a function of the wavelength (3.10). Without additional instruments such as integrating spheres and spectroradiometers, estimating the reflectance of material from a hyperspectral image is challenging without knowing the illumination's spectral power distribution (SPD). It is much easier to standardize the image into spectral reflectance factor by including an ideal diffuse reflector in the imaged scene 3.6 or in a separate scene image under the same lighting and camera configurations. Deriving this reflectance factor, value will be explained further in section 3.2.4.

$$\rho = \frac{r_s}{r_n} \tag{3.8}$$

$$\rho_f = \frac{r_s}{r_w} \tag{3.9}$$

$$\rho_f(\lambda) = \frac{r_s(\lambda)}{r_w(\lambda)} \tag{3.10}$$

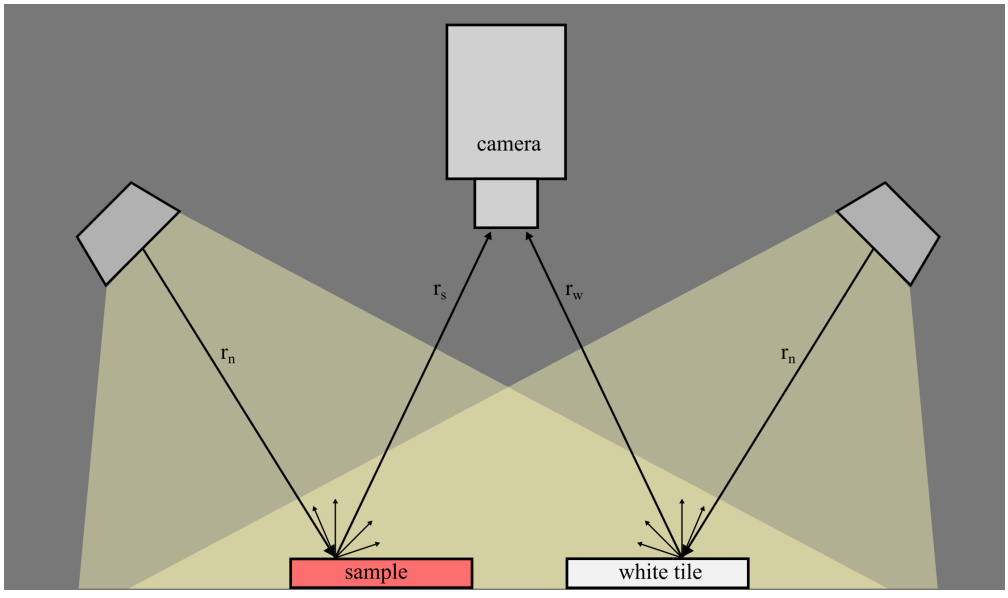


Figure 3.6: *Hyperspectral imaging setup configured with an in-scene calibration reference tile where r_n is the incident light onto the materials and r_s and r_w are the light rays reflected off the sample and white tile, respectively. With no measurements of the light source SPD, equation* (3.10) *can be used to solve for spectral reflectance factor, $\rho_f(\lambda)$ of the sample material.*

3.2.1 Image Noise

During operation, imaging devices are influenced by several noise sources; consequently, image data is sometimes contaminated with noise that can render a dataset severely limited if not corrected. Noise is a distortion of the input radiative signal through various means and produces output with some error. The error may be inconsequential if the input signal is large enough compared to the distortion level. Conversely, suppose the input signal is low in low-light settings or under illumination wavelengths that the sensor is weakly sensitive to. In that case, the noise distortion magnitude can exceed the original signal's.

SNR is a measure of the signal magnitude relative to the distortion magnitude and is typically defined in image processing as the ratio of signal mean to standard deviation (3.11).

$$\text{SNR} = \frac{\overline{s}}{\sigma} \tag{3.11}$$

It can be used as a gauge of signal quality, including image signal quality, but to calculate it, statistical data is needed for the image. Each image sensor pixel produces an image pixel value which is scalar for single-channel images or a vector of n dimensions, $\hat{p}_n$, for multichannel cameras. Defining SNR as a signal mean to standard deviation ratio, multiple signal samples for a given pixel must be captured to calculate the SNR. In this dataset, no two images represent the same scene as an input stimulus for the camera, making strict SNR irrelevant except for comparisons between processed images and originals. This is with two exceptions. Firstly, every hyperspectral image is recorded along with a dark frame image, where the shutter is closed, ideally providing a spatially uniform, constant signal of zero magnitude to the image sensor. These dark frame images ideally represent a constant input to the camera sensor, potentially providing 384 samples for each image sensor pixel to calculate SNR with. In Subsection 3.2.2, the dark frame correction and SNR will be discussed further. Secondly, spatially homogeneous regions in images (areas representing the same material under uniform lighting) could serve as multiple samples of the same input, provided that pixel sensor behavior is uniform and any independent noise at each pixel is random and normal.

Particularly true in CMOS image sensors, where each pixel contains its signal-digitization circuitry; noise can occur on individual pixels independently of other pixels. If the noise is truly random, however, two separate pixels receiving the same input signal could serve as two samples for noise analysis. In the dataset, every image contains three homogenous grey reference tiles. While the light dispersion across the tiles is not uniform, small enough patches from the gray-reference tiles could be extracted where the light uniformity could be assumed to be practically

even. To test this, the pixel values from a small circular patch in the lower left gray-reference tile of sample T1S01A were extracted, treated as separate samples of the same signal, and analyzed (Fig. 3.7). Suppose the signal from the gray-reference tile is assumed to be spatially homogeneous, and the noise source is assumed to be the same across pixels in the gray-reference patches. In that case, the image noise can be characterized by the deviation of each pixel from the mean, as shown in the density plots of Figure 3.7. Looking at the plot of pixel signal deviation against frequency, it can be seen that the deviation closely resembles a normal or Gaussian distribution. Gaussian curves were fit to the deviation distributions of the gray-reference pixels at each channel in the hyperspectral image to obtain the Gaussian curves shown in Fig. 3.7 where the value frequency is plotted on a logarithmic scale. The signal distribution would seem to imply that the hyperspectral image data has the best quality at the extremes where the standard deviation is the lowest, but this is the opposite.

In hyperspectral imaging, this is most evident in channels at the extremes of the spectral range. Most image sensors have very low sensitivity near the extremes of their sensitivity range; as hyperspectral cameras record their image channels over very narrow spectral bands, this results in bands at the extremes of the spectral range having meager signal power compared with the noise present in the system.

3.2.2 Dark-frame Correction

"Dark noise", or thermal noise, is caused when the heat energy of operation or the ambient environment causes additional photon counts to be erroneously registered in a photodetector readout due to thermal excitation and successive electron emission in the semiconductor material. A dark frame image characterizes this type of noise. There is also "readout noise", the inherent error resulting from analog-to-digital conversion at the point where a pixel's analog charge value is quantized into a digital number representing the so-called photon counts. This can also be characterized in the dark frame capture.

The dark frame capture should be taken as soon as possible before or after the scene image, as these various sources of system noise will vary with time as the environment changes and the instrument operates (a hotter room or longer exposure times will cause more thermal noise). Once the dark frame is captured, correction is simple; subtracting the dark frame noise image from the scene image, as in (3.12), removes the noise that was added by the system.

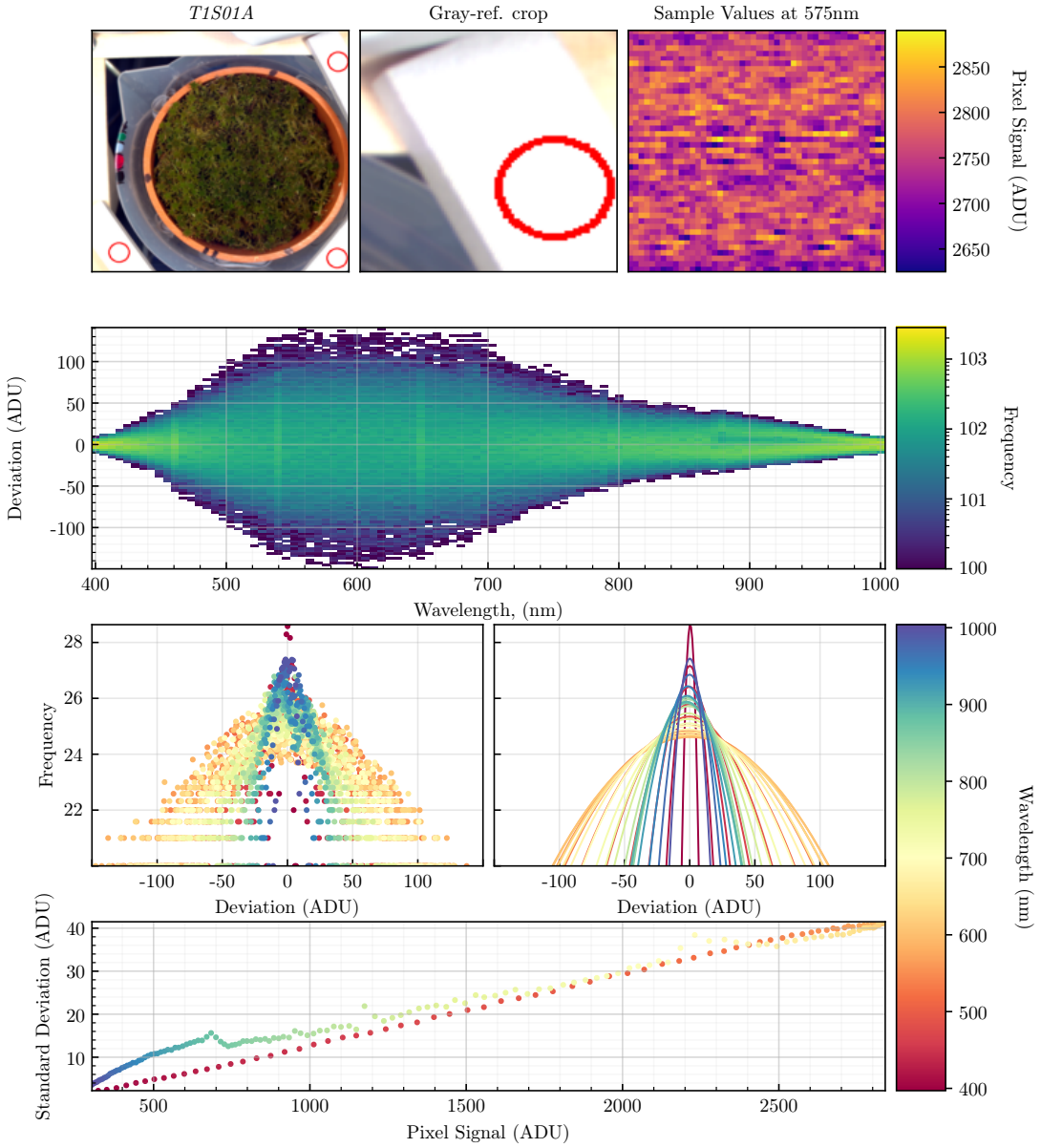


Figure 3.7: *Sample deviation distributions of pixel values in the gray-reference tile patch. Top row: gray-reference patch extraction location, patch zoom-in crop, pixel intensity equalization plot. Upper right: pixel signal intensity raster. Second row: 2D histogram of deviation as a function of wavelength. Third row: pixel deviation as a function of wavelength and their least-squares fit Gaussian models. Bottom row: standard deviation as a function of pixel intensity and wavelength.*

$$\mathbf{R}_{im}^{'} = \mathbf{R}_{im} - \mathbf{R}_{dark} \tag{3.12}$$

where:

$\mathbf{R}_{\text{im}}^{'}$: corrected image

$\mathbf{R}_{\text{im}}$: uncorrected image

$\mathbf{R}_{\text{dark}}$: dark frame image

This dark frame correction was performed on all the hyperspectral radiance images collected in this dataset. This should remove most of the noise covered in subsection 3.2.1. The Specim IQ camera uses the "push-broom" hyperspectral capture technique; a two-dimensional image sensor is used along with light dispersion optics to capture both the spatial and spectral information of a single scan line of the target. The scan line is moved across a target material to obtain a final three-dimensional image that represents the two-dimensional spatial context and the spectral dimension of every pixel. The Specim IQ captures a dark frame automatically for each image by capturing several images with the shutter closed and averaging the results. Although the dark frame image is two-dimensional in this case 3.8, it only represents one spatial axis; the second dimension of the sensor is used for the spectral domain. The dark frame image then corresponds to a single spatial row. This dark frame row can then correct every spatial row in the final three-dimensional datacube.

In Fig. 3.8, a sample dark frame and the mean and variance of dark frames for the entire dataset are shown. Several types of noise can be inferred from the images. The fringing bands appear most obvious as looped lines overlapping across the sensor's image. These are likely etaloning patterns, the result of incident light reflecting within sensor or device coatings before being detected and interfering with successive photons, also known as thin-film interference.

With image sensors, this can be caused by thinly ground, backlit image sensor designs or thinly applied sensor coatings, or thin optical layers in the exterior. When the wavelength of an incident photon exceeds the thickness of some visual material in the camera or sensor, it reflects within the material, constructively and destructively interfering with other incident photons. In this case, it is unclear how the etaloning occurs in the dark frame images; perhaps it is a similar effect produced by some other type of internal thermal noise source. Nevertheless, the noise signal is consistent throughout the images (evident from the variance image); this lends it to being easily removed from the dataset images with the dark frame correction equation (3.12).

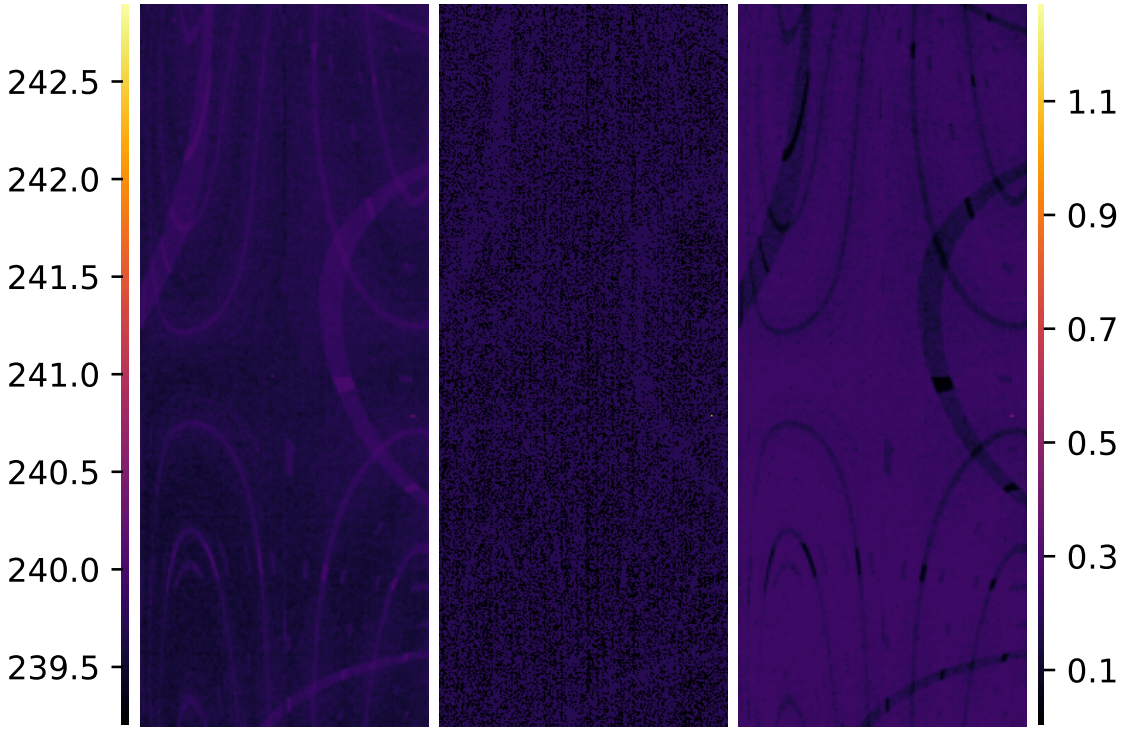


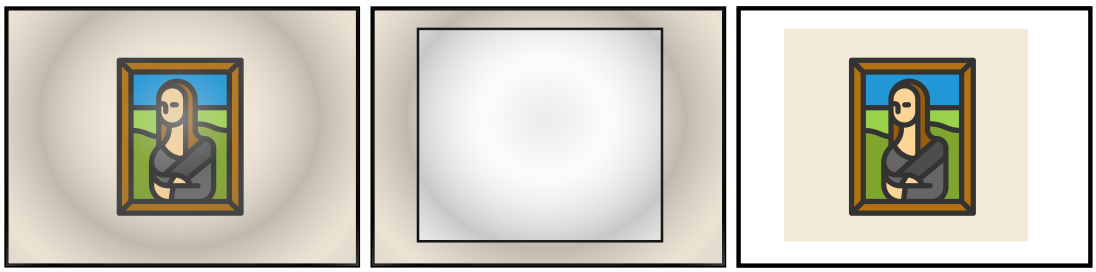
Figure 3.8: *Dataset dark frames (histograms equalized for visibility). Left: mean; middle: individual frame; right: variance*

3.2.3 Flat-field Correction

Flat-field correction aims to correct any non-uniformities in light intensity detected at the sensor. Detected light intensity may be non-uniform because the light source projected light on the image target unevenly or due to non-uniform spatial sensitivity across the image sensor. It can also be caused by optical elements in lens systems producing the vignetting effect; it may also be a combination of multiple causes. Whatever the reason for non-uniformities in detected light intensity, it is essential to correct this to derive a uniform signal from the imaged materials.

Given a planar sample and single image capture, the simplest way to perform flat-field corrections is to take a second image of the scene with a spatially uniform white material in place of the sample. The spatially uniform white material allows the spatial non-uniformity of the light source or image sensor non-uniformities to be recorded.

In this project's dataset, a limited set of white reference images are available; during hyperspectral capture, the primary focus was capturing images with in-scene gray reference tiles consistently placed. For hyperspectral images, the result of computing a flat-fielded result has special meaning; the flat-fielded image is a



(a) *Specimen* $\underline{\boldsymbol{C}}_s$ (b) *White reference* $\underline{\boldsymbol{C}}_w$ (c) *Flat-field result* $\underline{\boldsymbol{C}}_f$

Figure 3.9: *Example of flat-field correction where* $\underline{\boldsymbol{C}}_f = \underline{\boldsymbol{C}}_s / \underline{\boldsymbol{C}}_w$ *Notice that any areas not covered by the white sample become completely white in the flat-fielded result as the pixel values in these areas remain unchanged between (a) and (b) yielding a result of 1 when divided.*

reflectance factor image. This is explained in more detail in Section 3.2.4.

3.2.4 Reflectance Derivation

Camera response is the product of illumination, material reflectance, and imaging system net sensitivity (3.13). Additional environment and system parameters such as atmospheric or optical system transmittance also play a role (3.13). Still, in a laboratory setting, an imaging system's response can be simplified to (3.14) where a single term, ω, represents the entire product of the imaging system's equation terms.

$$c_{\mathbf{i}} = \int_{\lambda_{\min}}^{\lambda_{\max}} l_{\mathbf{i}}(\lambda) r_{\mathbf{i}}(\lambda) o_{\mathbf{i}}(\lambda) \phi_{\mathbf{i}}(\lambda) a_{\mathbf{i}}(\lambda)\, d\lambda + \epsilon_{\mathbf{i}} \tag{3.13}$$

where:

$\mathbf{i}$: vector of $[i, k]$; i specifies image sensor pixel and k specifies sensor channel

$c_{\mathbf{i}}$: camera response

$\phi_{\mathbf{i}}(\lambda)$: color filter transmittance

$l_{\mathbf{i}}(\lambda)$: illuminant power distribution

$r_{\mathbf{i}}(\lambda)$: color sample reflectance

$o_{\mathbf{i}}(\lambda)$: optical system transmittance

$a_{\mathbf{i}}(\lambda)$: sensor sensitivity

$\epsilon_{\mathbf{i}}(\lambda)$: additive noise

$$c_{\mathbf{i}} = \int_{\lambda_{\min}}^{\lambda_{\max}} l_{\mathbf{i}}(\lambda) r_{\mathbf{i}}(\lambda) \omega_{\mathbf{i}}(\lambda)\, d\lambda + \epsilon_{\mathbf{i}} \tag{3.14}$$

where:

$\omega_{\mathbf{i}}(\lambda)$: composite camera sensitivity

Equation (3.14) can be simplified to a discrete form (3.15). This works well with typical discrete output from cameras.

$$c_{\mathbf{i}} = (\mathbf{l_i} \cdot \mathbf{r_i})^t \boldsymbol{\omega}_{\mathbf{i}} \tag{3.15}$$

The raw output from the hyperspectral camera is, therefore, directly dependent on the measurement scene illumination and is, therefore, a poor basis for building generalized machine learning models as complex illumination is difficult to reproduce as a consistent input from experiment to experiment accurately. Instead of using raw hyperspectral output, hyperspectral readings are often converted to relative values of reflectance factor. Reflectance factor is the ratio of reflected radiant fluxes from the target material and some ideal diffuse reflector. Reflectance factor is used as an approximation of a material's true reflectance and ideally serves as a characterization of a material's intrinsic radiometric features; micro-geometry and physical composition influence the material's reflectance, making reflectance factor a good covariant for many phenotypes and material characteristics. As an approximation, however, reflectance factor is not perfect; it will in some way depend on the scene lighting, imaging instrumentation used to capture the radiance, and the reference materials used in the derivation. If unaccounted for, various systems noises can drastically alter a reflectance factor spectra, especially near the extreme ends of the signal where illumination signal and sensor sensitivity are usually very low in magnitude.

Given the camera response equation (3.15) and a hyperspectral imaged radiance dataset, how can reflectance factor be derived? It is straightforward; the process is the same as performing flat-field corrections as explained in Section 3.2.3. Using the camera response equation, however, the derivation will be presented. Generally, to solve for reflectance factor, both the camera system's sensitivities and the light source SPD must be known. However, a perfect reflector in an imaged scene could be used as a proxy for the light source's SPD. If the perfect reflector were instead an ideal diffuse reflector, then the viewing angle for the camera and incident angle from the light source should not influence the sensor's incident light intensity; the diffuse reflector would reflect the light source, however, at some ratio of the source intensity, $\alpha_{\mathbf{i}}$. If a highly reflective diffuse reflector is used, it could be used as an approximation of the full light source intensity.

Dividing the sample image pixel by the white reference image pixel, c_{w}, will allow the light source to be removed from the equation, given some assumptions, and replaced by some constant scaling factor that represents the ratio of light reflected by the reflector (in this case, approximated as 1):

$$\frac{c_i}{c_{\mathrm{w}}} = \frac{(\mathbf{l}_i \cdot \mathbf{r}_i)^t \boldsymbol{\omega}_i}{(\mathbf{l}_{\mathrm{w}} \cdot \mathbf{r}_{\mathrm{w}})^t \boldsymbol{\omega}_{\mathrm{w}}} \tag{3.16}$$

If the following are true or assumed, then $\boldsymbol{\omega}_i = \boldsymbol{\omega}_{\mathrm{w}}$:

- pixel i is the same pixel as pixel w and c_i, c_w are captured in successive images
  - light source and camera system are not temporally fluctuating
- pixel i is NOT the same pixel as pixel w and c_i, c_w are captured in the same image
  - camera system sensitivity is spatially uniform and
  - non-uniform noise has been removed

With camera system sensitivities being equivalent, (3.16) reduces to:

$$\frac{c_i}{c_{\mathrm{w}}} = \frac{(\mathbf{l}_i \cdot \mathbf{r}_i)^t}{(\mathbf{l}_i \cdot \mathbf{r}_{\mathrm{w}})^t} \tag{3.17}$$

Furthermore, if the previous conditions are true or assumed, then $\mathbf{l}_i = \mathbf{l}_{\mathrm{w}}$, with the added condition that the light dispersion must be uniform. Given these conditions, the incident light at pixels i and w are equivalent and cancel out:

$$\frac{c_i}{c_{\mathrm{w}}} = \frac{\mathbf{r}_i^t}{\mathbf{r}_{\mathrm{w}}^t} = \frac{\mathbf{r}_i}{\mathbf{r}_{\mathrm{w}}} \tag{3.18}$$

Finally, if the diffuse reference reflectance spectra, $\mathbf{r}_w$, has a constant reflectance, $\overrightarrow{(r)_{\mathrm{w}}}$, then the target material's reflectance at pixel i is:

$$\frac{c_i}{c_{\mathrm{w}}} = \frac{\mathbf{r}_i}{\overrightarrow{(r)_{\mathrm{w}}}} \tag{3.19}$$

$$\mathbf{r}_i = \frac{\overrightarrow{(r)_{\mathrm{w}}} c_i}{c_{\mathrm{w}}} = \frac{r_{\mathrm{w}} c_i}{c_{\mathrm{w}}} \tag{3.20}$$

Again, if the diffuse reflector is assumed to be reflecting the full light intensity, then:

$$\mathbf{r}_i = \frac{c_i}{c_{\mathrm{w}}} \tag{3.21}$$

Similarly to obtaining a flat-fielded image, dividing the hyperspectral camera response for a target material by the hyperspectral camera response for a white diffuse reflector, given the same imaging configurations, produces the reflectance of the target material. In reality, the diffuse reflector does not reflect the full intensity of the light source, and it does not have a flat spectral reflectance curve; therefore, in most calibration cases, this correction will produce the reflectance factor. Additionally, this equation only represents a single channel for a single pixel of the hyperspectral image; it can be applied to each pixel and each channel of the hyperspectral image in the same way though to obtain a reflectance factor image cube from the entire hyperspectral raw cube.

Considering the hyperspectral camera, each pixel i is a vector of p spectral channels, $\hat{c}_{ip}$, each value representing camera response in the same way described in (3.15), except that the impinging illumination, $\mathbf{l}_i$, is a very narrow band spectra vector. This is because the hyperspectral camera diffracts the full range of impinging light into constituent light waves across the one spatial axis of the camera sensor.

The two primary methods for accomplishing this reflectance factor derivation from hyperspectral images involve placing a diffuse white reference material in the scene. The sample material is either placed next to the reference and captured in the same image, or the reference material can be placed in the exact location and orientation as the sample material and caught in a second image. The advantage to the latter method is that scene's light nonuniformity over the white reference material will match that of the sample image nonuniformities; dividing the raw image by the white reference image then produces a properly flat-fielded reflectance factor image. Using the former method, where the sample and reference sit next to each other in the scene, complicates derivation, as the light intensity is never perfectly uniform, and the light intensity at the reference material will not match the intensity over the sample. Furthermore, the light intensity within each material will not be constant.

Applying (3.21) to a hyperspectral image with some white reference available will yield an approximation of the sample material reflectance; in Figure 3.10, a sample hyperspectral image of a ColorChecker has been captured to demonstrate the transformation from raw arbitrary hyperspectral values to reflectance. In the plot, notice how the radiance values of every sample taper towards the extreme ends of the spectra due to the decrease in camera sensor sensitivity and illumination intensity near the extreme ends of the hyperspectral camera's range. This implies

that, whatever the noise in the system is, the SNR will be much more significant at the ends of the hyperspectral cube. The derived reflectance factor is somewhat misleading in this respect as it appears to show a much greater signal towards the extremes, but this is due to the converging of the signals towards near-zero at the ends of the spectra. With such a low signal value here, the image noise is influential enough to produce likely erroneous values of reflectance factor which exceed 1.

Finally, combining everything along with dark frame correction, and standardizing the reflectance factor into CCRF by multiplying it by the nominal reflectance of the white reflector, the full equation for converting a hyperspectral image, $\mathbf{C}$ into a hyperspectral CCRF cube is shown in (3.22).

$$\mathbf{CCRF}_{im} = r_{nom} \cdot \frac{\mathbf{R}_{im} - \mathbf{R}_{dark}}{\mathbf{R}_{white} - \mathbf{R}_{dark}} \tag{3.22}$$

3.3 Specimen Segmentation

The mesocosm dataset has no spatial mask labels. Although specimen and gray tiles are fairly consistently placed from image to image, placement is not pixel-perfect. A segmentation method is needed to extract pixels containing specimen material and gray reference reflectors. In Figure 3.11, the regions of interest are shown with the specimen outlined in green and the gray tiles outlined in red.

Segmented regions need to be highly pure; small amounts of foreign material, such as the enclosing terra cotta pot, could drastically alter predictive models based on the images. Some options for obtaining a segmentation of the image include:

Fully-manual annotation : for each image in the dataset, draw separate masks to isolate the specimen and gray samples.

Partially-automated annotation : annotate the location of region centroids and extract some uniform geometry around the centroid.

Automated segmentation methods : Several automated methods exist, including supervised and unsupervised methods.

Fully manual annotation could be the most accurate option, but it would be time-consuming and inflexible. A decision later to alter the segmentations somehow would render the first set of segmentations useless. Automated segmentation methods work well in cases like this where the target materials are known and mostly spectrally homogenous; however, as segmenting potted moss samples is neither a common task nor the primary goal of this project, automated methods

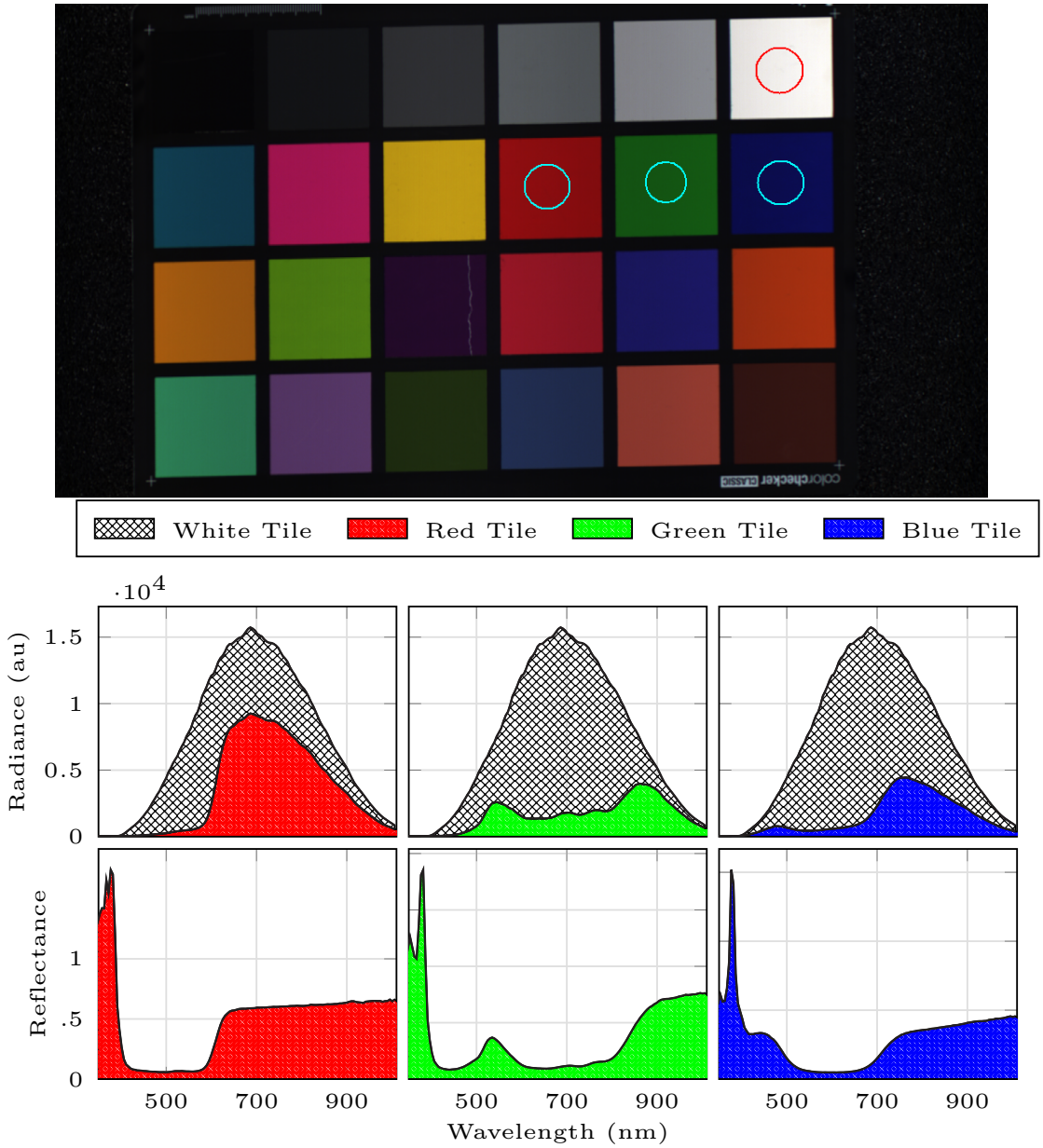


Figure 3.10: *Preview of a ColorChecker hyperspectral image and the extracted P patch radiances of red, green, and blue tiles from the Colorchecker along with their derived reflectance factor using* (3.21)

were decided against as testing multiple ways before obtaining a reliable result would pose a roadblock to subsequent tasks. Partially-automated annotation was used instead of using a set of partial annotations included with the dataset. Along with the dataset are annotations for each image marking the spatial location of the

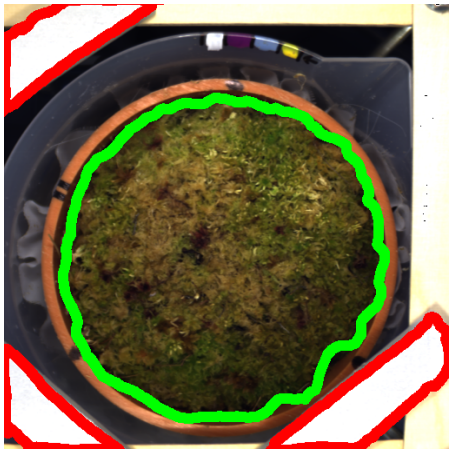


Figure 3.11: *Regions to be segmented: specimen (outlined in green) and gray tiles (outlined in red).*

specimen center and a point on each of the three gray reference tiles. With these available, a consistent segmentation is simple; circular patches were drawn around each location point to represent specimen and gray tile material.

3.4 Chemometric sensing

This section briefly explains the background theories defining the link between chemical composition and light absorption.

3.4.1 Light absorption

Once calibrated, radiometric data obtained from hyperspectral images can be used to correlate the chemical properties of materials in the images. Radiometric data recorded off organic reflective surfaces, however, tend to exhibit a high correlation among spectral samples taken closely together; furthermore, organic materials contain a mixture of chemical compounds, each contributing potentially overlapping absorbance spectra to the total absorbance of the material. The high spectral correlation and mixed material absorbances manifest as broad peaks and plateaus in the absorbance spectra. Ideally, light absorption at the atomic level would produce extremely narrow absorption peaks at selected wavelengths where the light energy matches the band gap energy required to excite an electron in the material to the next energy level. Light energy is described as:

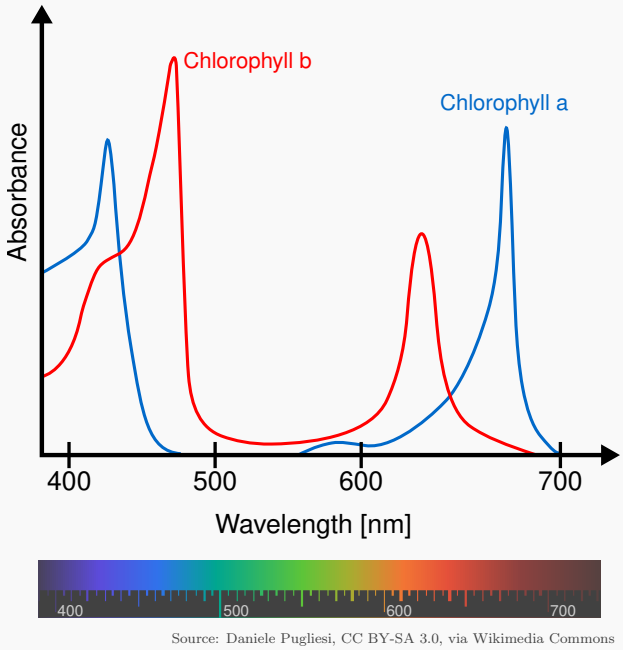
$$E = \frac{hc}{\lambda} \tag{3.23}$$

where h is Planck's constant, c is the speed of light, and λ is the wavelength of the light wave. If the energy of a light photon, E, matches that of the band gap energy for an electron, then the light wave may be absorbed if it strikes the electron. Absorption spectra, however, indicate that light waves of a broad range of energies are absorbed rather than just a single wavelength. Considering the absorbance spectra of the chlorophyll a and b compounds in Fig. 3.12a, three to four absorbance peaks can be observed for each compound, with each peak spanning up to 100 nm or more. Aside from influences from spectral capture instrumentation, several factors cause this broad-band absorbance result.

Besides electronic vibration, there are many more energy states within the atomic and larger molecular structures, each of which can absorb light as energy transitions between their respective energy levels. For a large molecule like chlorophyll a, there are quite a few unique elements and bond configurations within the structure; for every arrangement, a potentially unique band gap energy is required to excite electrons in the outer shells of the atoms. This can result in multiple absorbance peaks in the Visible light spectrum (VIS) and Ultraviolet light spectrum (UV) spectrum from a single molecular compound but does not yet explain the broad-band distribution of absorption. Along with vibrational energy level transitions, electrons also exhibit rotational energy levels, of which the gaps between are much smaller in magnitude than the vibrational energy level gaps (HOLLAS, 1984). These rotational energies act simultaneously with vibrational energies to produce a continuous band gap energy, the variation of which resembles a Gaussian distribution around the vibrational energy gap with the sample variance controlled by the rotational energy gap. VIS and UV photon energies tend to match excitation levels of valence electrons. In contrast, Infrared light spectrum (IR) photon energies tend to reach the vibrational energy levels of the molecular structure (HOLLAS, 1984). Thus, Ultraviolet and visible light spectrum (UV-VIS) absorbance data can generally provide information about valence electron configurations within a molecule, while IR absorbance data contain information about molecular characteristics.

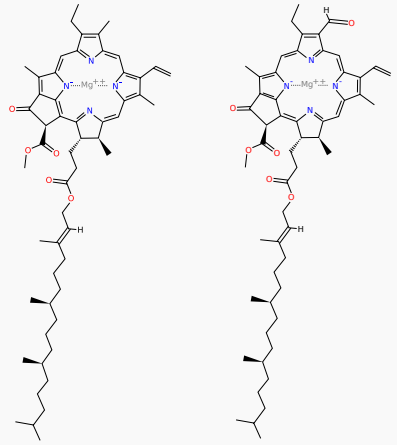
3.4.2 Chemometric correlation

As discussed previously, compounds' atomic and molecular structures affect how light will be absorbed. Beer-Lambert's law further describes the relationship between a chemical's light absorbance intensity and the chemical concentration within the sample. Absorbance, formally defined as the logarithm of the ratio



Source: Daniele Pugliesi, CC BY-SA 3.0, via Wikimedia Commons

(a) *Absorbance spectra of chlorophylls a and b*



(b) *Chlorophyll a* **(c)** *Chlorophyll b*

of incident to transmitted radiant power and given in (3.24), is related to molar concentration by (3.25):

$$A = \log \frac{I_0}{I} \tag{3.24}$$

$$A = \epsilon c l \tag{3.25}$$

Where A is unitless absorbance, I_0 is incident radiant power, I is transmitted radiant power, ϵ is the molar absorption coefficient ($\mathrm{M}^{-1}\mathrm{cm}^{-1}$), c is the molar concentration (M), and l is the length of the light path through the substance (cm). For a constant light path length and molar concentration, the absorbance given in (3.25) has a linear relationship to the chemical concentration.

So far, with the hyperspectral images, only the reflectance has been calculated, taken as the ratio of reflected light off the specimen to the reflected light off of a diffuse reference sample. There is no such measurement in the images of transmitted light from which to calculate absorption; however, for an opaque and diffuse material, the reflected light off the specimen can be substituted for transmitted light as a rough approximation. Making this assumption, an absorbance hyperspectral image,

$\overline{\mathbf{A}}$, could be computed from the reflectance factor hyperspectral image, $\overline{\mathbf{R}}$, with (3.26):

$$\overline{\mathbf{A}} = -\log \overline{\mathbf{R}} \tag{3.26}$$

Assuming system linearities, the absorbance hyperspectral image pixel intensities should be proportional to the constituent chemical concentrations in the imaged specimens.

Table 3.1: *Measured and estimated mesocosm experiment values from (Kokkonen et al., 2023)*

Mesocosm Trait	Range	Unit
P_{max}	160 - 17,200	mg CO_1 m^{-2} day^{-1}
P_{bio}	$1.5e^{-2}$ - $28e^{-2}$	g cm^{-2}
P_{area}	10 - 190	mm^2 cm^{-2}

3.5 Spectral Products

To assess vegetation specimen health via hyperspectral photogrammetry, several spectral products derived from the hyperspectral dataset are compared for efficacy in predicting vital health traits of peatland species. Spectral products are defined in this project as simply the derivatives of the original hyperspectral dataset, which are used for modeling and characterization of the peatland mesocosms and habitat. Table 3.1 overviews the experiment observations from Kokkonen et al. (2023) that are used as modeling outputs, while Table 3.2 overviews all spectral products derived from the hyperspectral dataset as input predictors. Aside from the reflectance and absorbance, the rest will be discussed in further sections.

The three categories of products are described below and will be discussed throughout further sections, later to be used as inputs to machine learning models for predicting the mesocosm bioindicator traits measured by Kokkonen et al. (2023).

raw signal: untransformed signal derived from the hyperspectral images (reflectance factor, absorbance, vegetation index)

normalized signal: signal that has had continuum-removal applied to normalize it, minimizing differences between sample measurement environments

scalar signal: scalar derivations from pure signals (integration of signal)

Table 3.2: *Spectral products derived from the dataset and used in machine learning models for predicting bioindicators from Table 3.1*

Spectral Product	Type	Signal	Range
SD	raw	$\overline{r}_{SD}$	400nm - 1000nm
$CR_{Chl.}$	normalized	$\overline{cr}_{Chl.}$	649nm - 715nm
$CRSC_{Chl.}$	scalar	$\overline{crsc}_{Chl.}$	649nm - 715nm
CR_{LD}	normalized	$\overline{cr}_{LD}$	709nm - 781nm
$CRSC_{LD}$	scalar	$\overline{cr}_{LD}$	709nm - 781nm
CR_{NIR}	normalized	$\overline{cr}_{NIR}$	911nm - 1000nm
$CRSC_{NIR}$	scalar	$\overline{crsc}_{NIR}$	911nm - 1000nm
CR_{RG}	normalized	$\overline{cr}_{RG}$	531nm - 599nm
$CRSC_{RG}$	scalar	$\overline{crsc}_{RG}$	531nm - 599nm
MTVI2	scalar	$\overline{r}_{MTVI2}$	(560nm 670nm 800nm)
$MTVI2_{sum}$	scalar	r_{MTVI2}	(560nm 670nm 800nm)
SDA	raw	$\overline{a}_{SD}$	400nm - 1000nm
$CRA_{Chl.}$	normalized	$\overline{cra}_{Chl.}$	649nm - 715nm
$CRASC_{Chl.}$	scalar	$\overline{crsc}_{Chl.}$	649nm - 715nm
CRA_{LD}	normalized	$\overline{cra}_{LD}$	709nm - 781nm
$CRASC_{LD}$	scalar	$\overline{cr}_{LD}$	709nm - 781nm
CRA_{NIR}	normalized	$\overline{cra}_{NIR}$	911nm - 1000nm
$CRASC_{NIR}$	scalar	$\overline{crsc}_{NIR}$	911nm - 1000nm
CRA_{RG}	normalized	$\overline{cra}_{RG}$	531nm - 599nm
$CRASC_{RG}$	scalar	$\overline{crsc}_{RG}$	531nm - 599nm
MTVI2	scalar	$\overline{r}_{MTVI2}$	(560nm 670nm 800nm)
$MTVI2_{sum}$	scalar	r_{MTVI2}	(560nm 670nm 800nm)

Additionally, ecological response values (eco-response parameters), also used by Kokkonen et al. (2023), were computed for this data (3.27),resistance, recovery, and resilience:

$$\text{Resistance} = \frac{S_{pre} - S_{dst}}{S_{pre}} \tag{3.27}$$

$$\text{Recovery} = \frac{S_{pst} - S_{dst}}{S_{dst}} \tag{3.28}$$

$$\text{Resilience} = \frac{S_{pre} - S_{pst}}{S_{pre}} \tag{3.29}$$

where:

S_{pre} : ecosystem state prior to disturbance

S_{dst} : ecosystem state during peak of disturbance

S_{pst} : ecosystem state at peak of recovery after disturbance

3.5.1 Spectral Indices

Vegetation indices are useful metrics derived from reflectance data representing various properties ranging from chlorophyll to moisture content. As the chemical composition of a plant varies with its response to climatic changes, its reflectance signal often changes in specific spectral ranges, which can be correlated to the chemical change. Vegetation indices combine two or more spectral values at the bands most sensitive to the changes the vegetation index represents. While these indices are generalizable across many species, peatland species reflect light differently in NIR and SWIR bands and respond differently. The spectral bands at which changes in chlorophyll content, moisture, nitrogen, and other properties are most pronounced do not necessarily align with those of vascular vegetation. Therefore, vegetation indices may need to be adjusted for peatland observation. With this being said, it has been shown through several studies that the NDVI, along with a few other indices, may be promising as a general site-independent peatland GPP predictor, with the caveat that nonvascular peatland vegetation needs to have NDVI customized and calculated separately from vascular vegetation Lees et al. (2018).

In this project, a LAI VI is utilized for two purposes: (a) threshold detection of vegetation against backgrounds and (b) as a component in one of the spectral products. At such a high resolution as is given in this dataset, separating vegetation from the background signal is straightforward; each pixel usually covers a single material, allowing each pixel signal to be treated as an individual material signal. This is not completely true, as hidden layers of material will contribute to the single pixel's signal (i.e., the dirt or other materials underneath a leaf). An accurate LAI VI, however, should produce a stark difference between a pixel of a leaf and a pixel of dirt. Most VIs are designed for use in remote sensing where a single

pixel may represent a large swath of land or several meters square; LAI VIs aim to quantify the area of leaf cover represented in the pixel, typically using chlorophyll absorption bands, without reaching a saturation limit at high leaf cover densities. If high-resolution, leaf-scale hyperspectral images are available. One of these LAI predictor VIs could be used to separate non-vegetation pixels from vegetation pixels, where a vegetation pixel should exhibit very high LAI (100%) and a background pixel should ideally exhibit a low LAI. For this purpose, MTVI2, defined in (3.30), is utilized. Its background theory and success in predicting the LAI of various crops are detailed by its authors (Haboudane et al., 2004).

$$\text{MTVI2} = \frac{1.5[1.2(\rho_{800} - \rho_{550}) - 2.5(\rho_{670} - \rho_{550})]}{\sqrt{(2\rho_{800} + 1)^2 - (6\rho_{800} - 5\sqrt{\rho_{670}}) - 0.5}} \tag{3.30}$$

Other indices which aim to predict LAI are available; however, for this project, MTVI2 is used as it adequately separates the moss vegetation from background materials in test trials, is effective for high-resolution moss turf detection (Malenovský et al., 2015), and mainly takes into account signal contamination from soil. With moss canopies usually sitting close to the ground, accounting for soil signal effects will be advantageous in accurately selecting moss pixels and quantifying the LAI present.

The MTVI2 index was computed for all hyperspectral images using (3.30). In Figure 3.13, the four measurement session images of specimen *01B* are shown along with their MTVI2 results which have been thresholded into bins. The binning intervals are designed around the relationship between MTVI2 and LAI as empirically determined by the index's authors and shown in (3.31).

$$\text{MTVI2:LAI} = .2227exp(3.6566 \cdot \text{MTVI2}) \tag{3.31}$$

LAI is defined as the ratio of leaf area to ground area; its value can exceed 1 in areas where dense vegetation may contain multiple layers of overlapping leaves competing for light. It is a critical quantity for carbon cycle modeling, often used to estimate primary production. Bryophytes generally exhibit relatively high LAI values, commonly ranging from 6 to 25, though it is highly dependent on the ecosystem constraints (Proctor and Goffinet, 2008). During drought stress, moss bryophytes typically decrease leaf area as a desiccation-resisting tactic, curling and folding the leaves into a denser structure, reducing the effective photosynthetic area but protecting the capitulum from excessive water loss (Zotz and Kahler, 2007; Glime, 2013). While MTVI2 is designed to predict LAI, the prediction model should ideally be tailored to the use case. It should, nevertheless, be generally transferrable between species as it is based on the radiometric absorptions inherent

in photosynthesizing surfaces. As its name implies, MTVI2 is an enhanced version of former triangular indices, where the value corresponds to the geometrical area within a triangle formed by the reflectance values at 550nm, 670nm, and 800nm, with an adjustment made for soil effects. These wavelengths represent the spectral positions of the green, red, and infrared shoulders. The former triangular vegetation index (TVI) used 750nm as the infrared shoulder; however, due to effects from the chlorophyll absorption band widening as the concentration increased, the infrared shoulder was moved further out to be independent of chlorophyll effects.

Looking at Figure 3.13, the shaded triangles between these three points are shown for specimen samples from each session image of mesocosm *01B*, where the sample is spectra sample taken from a small circular area shown in red. *S. 01B* received irrigation from the experiment starting at measurement session 1 (T1) through session 2 (T2), like all other samples. It underwent extreme drought treatment until session 3 (T3) and a recovery irrigation period to session 4 (T4). The area of the MTVI2 triangles track the presumed health status of the plant throughout the four experimental stages, with the exception that the MTVI2 values decrease even more at S4 after having been watered following the drought, indicating that the mesocosm was either irreparably desiccated or was perhaps dormant and reluctant to revive.

3.5.2 Signal normalization

Consistent and reproducible hyperspectral image capture is challenging due to the nature of hyperspectral camera system designs. As explained in previous sections, the white reference standardizes the image as much as possible. Yet, the image signal can still contain systematic errors or offsets that make comparing samples challenging. In machine learning, sample normalization helps minimize these disparities' impacts. In signal preprocessing, however, if a target section of the spectral signal is selected for use, the signal can be internally normalized against a convex hull. The method known as continuum removal (Clark and Roush, 1984) has successfully been used in several remote sensing studies to process reflectance signals before using as a predictor for bioindicators such as nitrogen content, chlorophyll concentration, and vegetation quality (Huang et al., 2004; Kováč et al., 2012, 2013). It is performed by simply computing a convex hull over a region of signal and computing a modified ratio as shown in (3.32), where CR_i, r_i, and r_{ih} are the is the continuum removal value, reflectance, and convex hull value at band i respectively.

$$CR_i = |\frac{r_i}{r_i h} - 1| \tag{3.32}$$

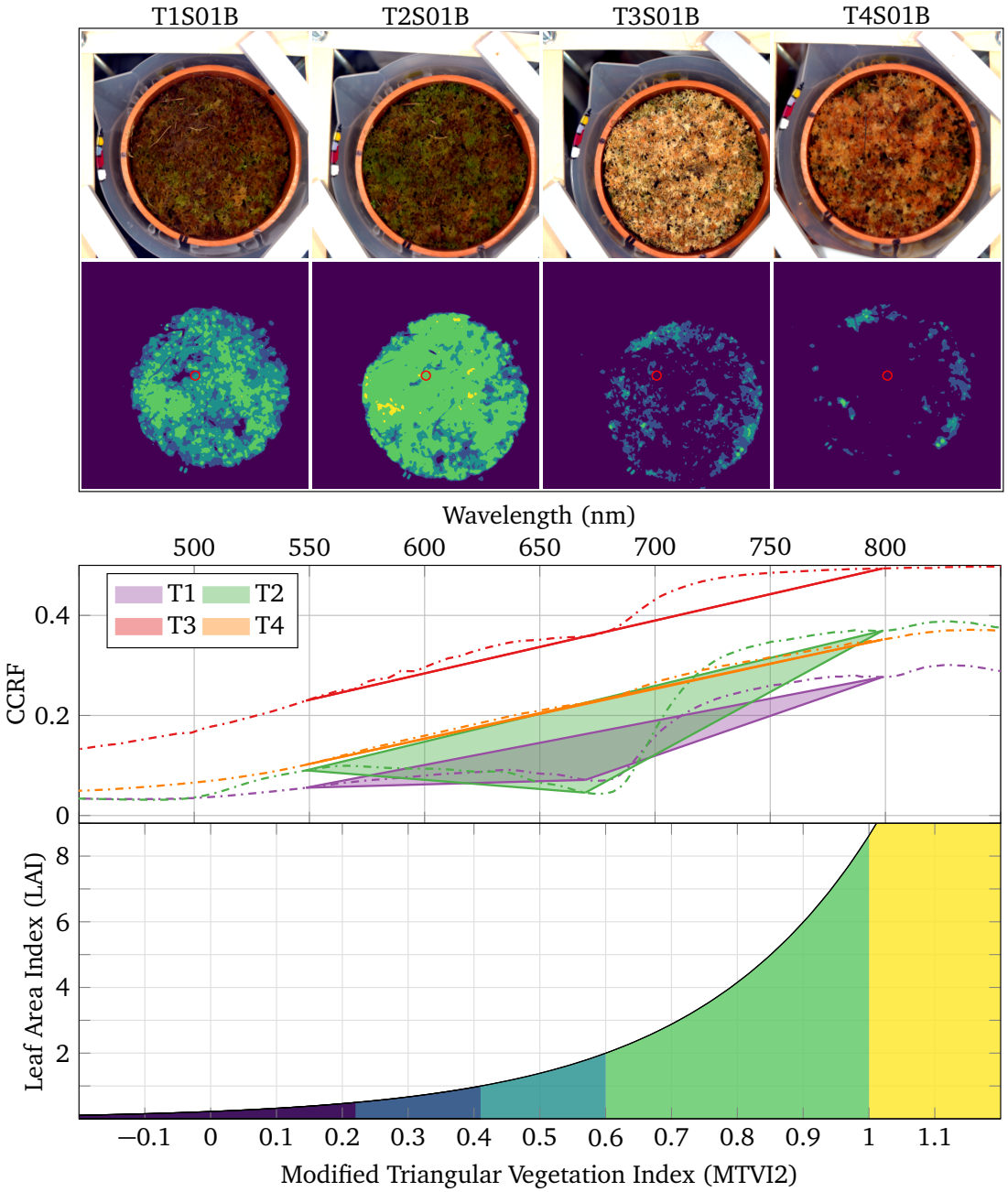


Figure 3.13: *From top to bottom: Mesocosm 01B image previews for each session. Binned MTVI2 results. Spectra from hyperspectral reflectance image patches (marked with red circles), with triangular areas used in MTVI2 computation* (3.30) *plotted. Relationship between MTVI2 and LAI as described by Haboudane et al. (2004), with shaded area colors corresponding to binning groups used in the MTVI2 sample renders.*

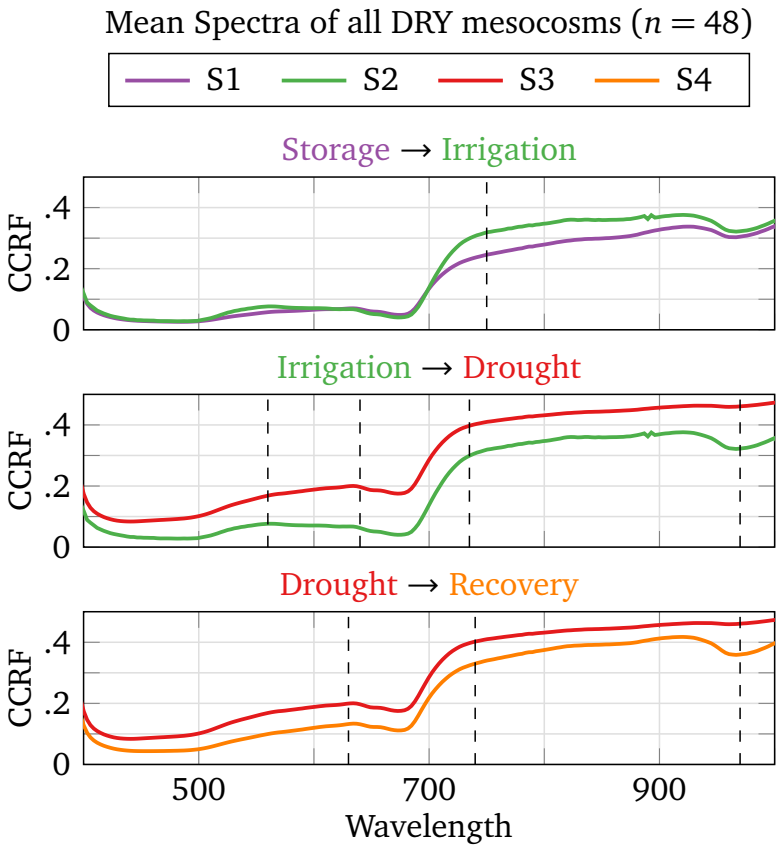


Figure 3.14: *Mean CCRF spectra plots for all DRY treatment mesocosm samples over each imaging session. Regions of significant variance*

Figure 3.15 shows an example of continuum removal being applied to mesocosm *10D*. The regions it is performed over are regions of interest where chlorophyll ($\boldsymbol{\lambda}_{Chl.}$), turf structure ($\boldsymbol{\lambda}_{LD}$), red/green chlorophyll pigmentation ($\boldsymbol{\lambda}_{RG}$), and water content ($\boldsymbol{\lambda}_{NIR}$) are suspected to exhibit effects resulting from changes in mesocosm health. Initial inspection on reflectance signals indicated significant changes in these spectral regions (Fig. 3.14).

It is evident from the results that signal normalization greatly aids in comparing regions of interest in the signals from different measurement times. Broadband absorptions strongly influence each spectral region in the source material. The narrow-band application of continuum removal helps to isolate a spectral feature from the dominant absorption effects in the signal.

Continuum removal was applied to all mesocosms' reflectance and absorbance images for the four spectral regions shown in Figure 3.15. This produced most spectral products used in training the prediction models (all products with CRx

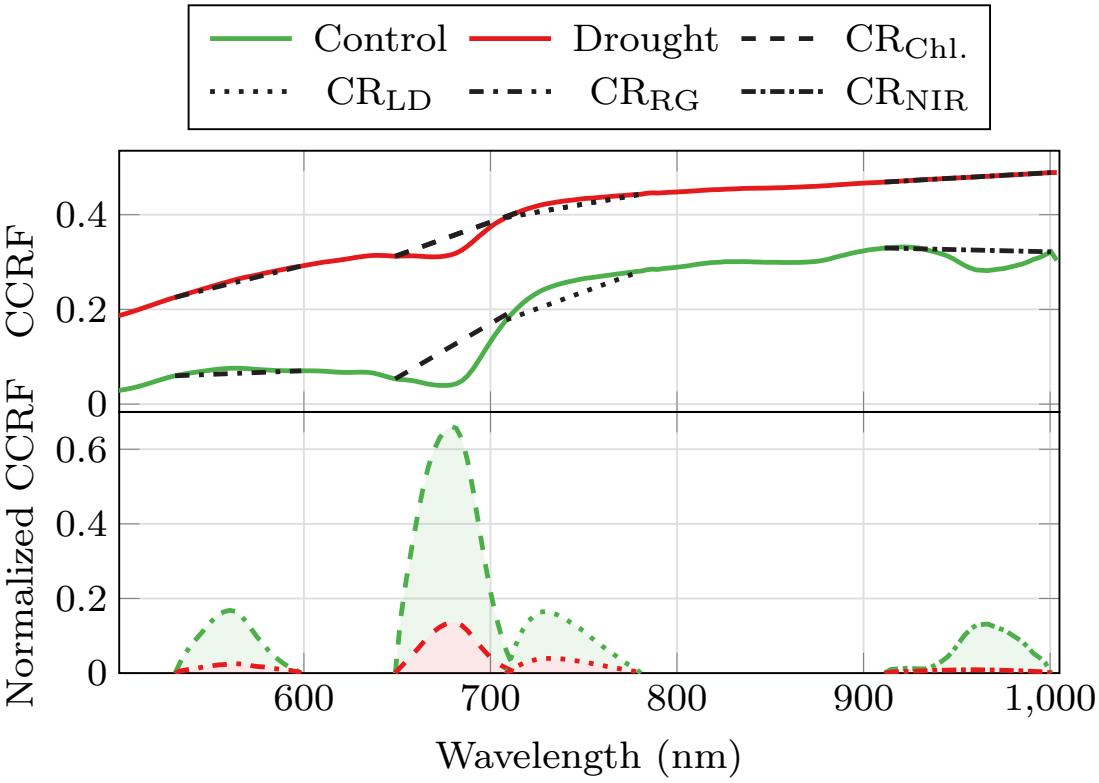


Figure 3.15: *Continuum-removed regions of interest in mesocosm 10D. The healthy sample from image capture session 2 while the drought sample from session 3, after the drought stress period.*

notation).

3.5.3 Machine learning predictive models

Given all the spectral products compiled up until this point, it makes sense to check if the relationships between them and the health indicator observations for the mesocosms are reliably correlated. Clear influences on the hyperspectral data are observed from session to session as mesocosms go from storage to sufficiently irrigated to drought stress and recovery. As an initial inspection, the Pearson correlation coefficients between all spectral products from Table 3.2 (except pure signals, MTVI2$_{sum}$, MTVI2A$_{sum}$, and the other continuum removal scalar products) and observed mesocosm traits from Table 3.1 were computed. The results are separated into reflectance products in Figure 3.16 and absorbance products in Figure 3.17 and grouped by drought and control treatment as well as peatland site history (WLD, dry, wet), site type (fen, bog), and species. While not entirely

indicative of all possible correlations, as they reveal linear correlation, the Pearson correlation coefficients indicate initial relationships between the data that may give way to promising predictive models.

The Pearson correlation values range from -1 to 1, with -1 representing a perfectly inverse linear relationship, 0 meaning no correlation, and 1 representing a perfect linear correlation.

The results reveal that mesocosms from the same site types (fen and bog) and the same species groups correlate more to the observed traits than those from the same peatland site history. They also indicate strong correlations between the photosynthetic capacity trait and the MTVI2, $CR_{Chl.}$, CR_{LD}, CR_{RG} from the reflectance signal, $\overline{r}_{SD}$ and even stronger correlation to some of the same products from the absorbance signal.

SVMs are a class of machine learning systems that have been used extensively for decades Boser et al. (1992) across a broad range of fields Cervantes et al. (2020) to solve high-dimensional classification and regression problems. Of interest, they have been used with success in vegetation bioindicator prediction Tuia et al. (2011); Pasolli et al. (2012), including assessment of bio-resilience Jafarzadeh et al. (2020). In this project, SVMs construct hyperplanes through the input data dimensions to linearly separate data into their known classes for classification problems or to group the data within an epsilon margin of error for regression problems. Regression SVMs, SVRs, were selected as this project's machine-learning design for several reasons. First, many spectral data products contain 20-30 dimensions, with some containing around 200 dimensions, making an optimized method attractive. SVRs only needs to operate on a subset of the data to construct hyperplanes, making them potentially more efficient than standard linear regression. Additionally, most SVR model implementations can use non-linear kernel functions on the data for cases where linear separation is impossible. Because of the potentially complex relationship between the spectral products and the measured bioindicator parameters as well as the derived ecosystem traits (resilience, resistance, recovery), the use of kernel functions may prove to be advantageous.

Detailed theory and application-specific implementations of SVMs can be found in related literature. For this project, a standard implementation with a configurable cost parameter (C), kernel parameters (gamma), and epsilon margins. In initial trials, it was found that linear SVRs performed just as well as kernel SVRs in some instances. For predicting the mesocosm traits, it was decided to compare the performance of Radial Basis Function (RBF) kernel SVRs against linear SVRs.

An experiment was set up, pairing all spectral products from Table 3.2 as predictors with each bioindicator parameter from Table 3.1, including the derived resistance, recovery, and resilience traits from photosynthetic capacity values. To build generalized, transferable predictors, the mesocosm specimens were not

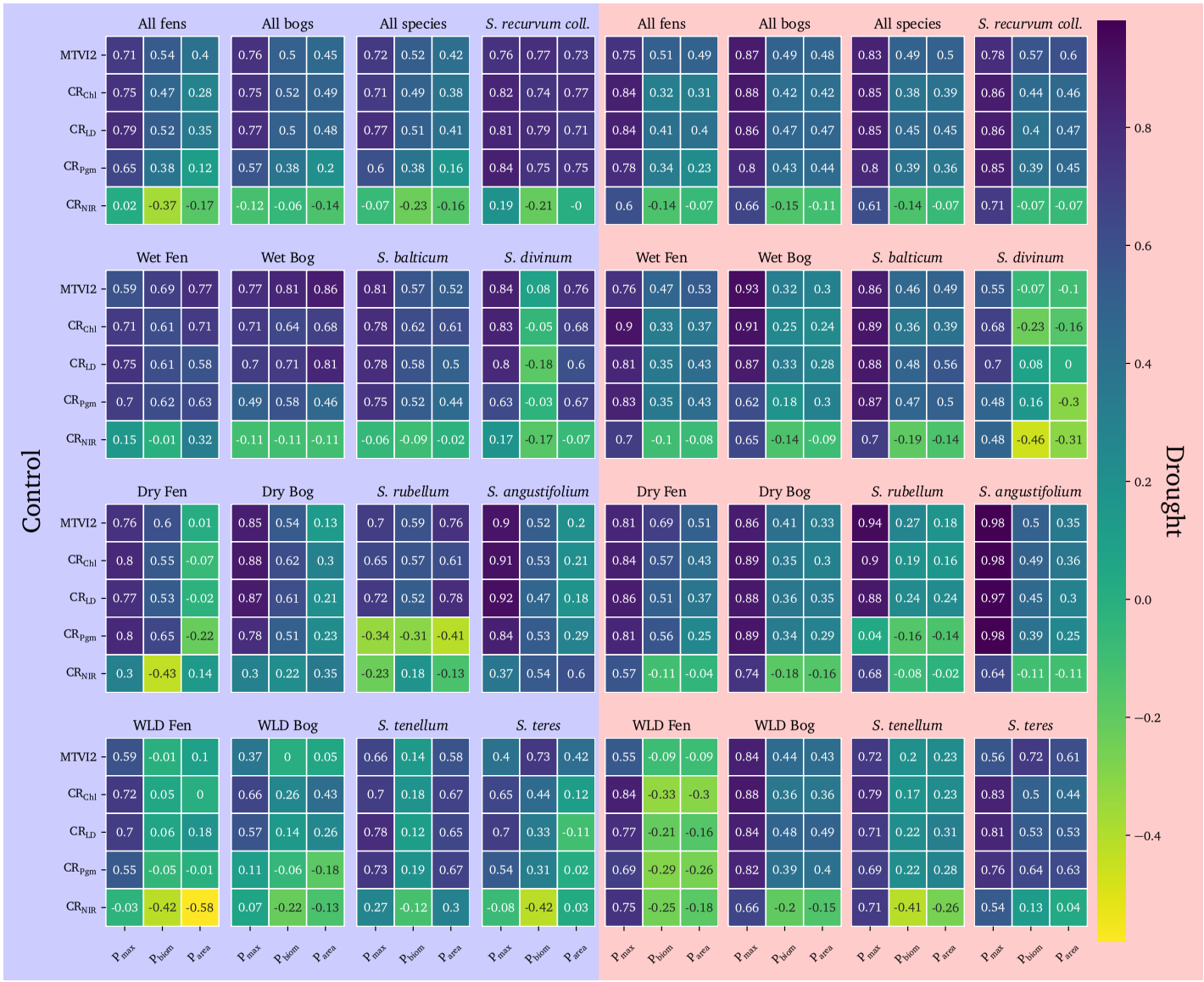


Figure 3.16: *Pearson correlation coefficients for all control specimens (sessions 2-4) between each of the scalar spectral products and the experiment mesocosm measurements and estimations (campaigns 1, 2, and 5). Pearson correlation assumes a linear relationship, which may not be true between the spectral data and biological characteristics.*

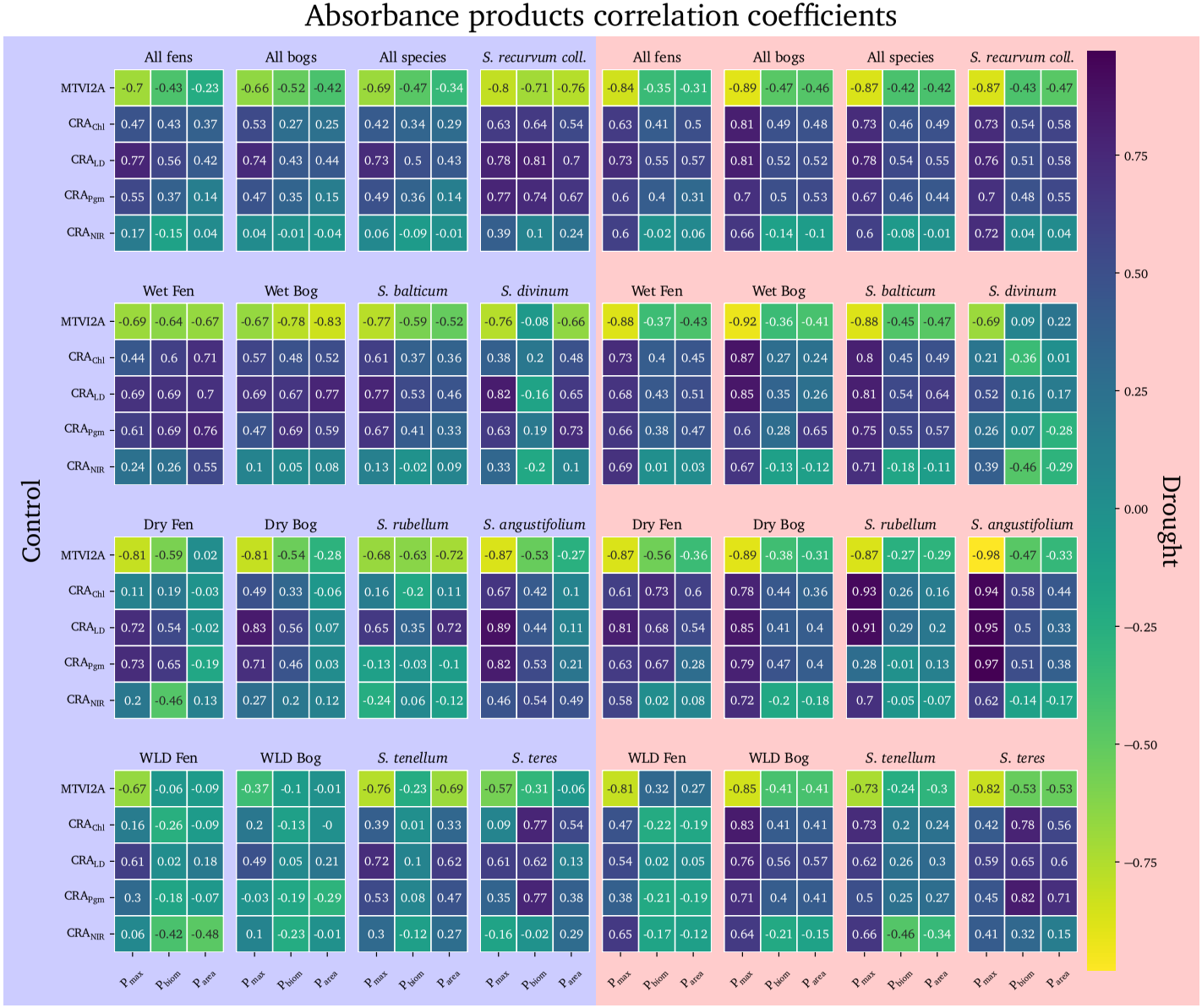


Figure 3.17: *Pearson correlation coefficients for all control specimens (sessions 2-4) between each of the scalar spectral products and the experiment mesocosm measurements and estimations (campaigns 1, 2, and 5). Pearson correlation assumes a linear relationship, which may not be true between the spectral data and biological characteristics.*

grouped by site history, site type, or species as they were in the Pearson correlation inspection. Grouping by input predictors, these classes would likely decrease data variance and contribute to more performant models, yet they might not be as practical for implementation in large-scale sensing. Nevertheless, these class labels available in the dataset could be used in a subsequent experiment where specimen origin and peatland type could be estimated from the spectral products. A potential photosynthetic capacity predictor model would complement peatland site predictors well. According to (Kokkonen et al., 2023), the findings from this dataset suggest that drying peatland sites will begin to shift turf composition towards drought-resilient species. Serving as a peatland vegetation health indicator, photosynthetic capacity estimations could indicate early what species may come to dominate the micro-topography of currently drought-stressed peatlands, but this shift would vary by peatland type. It would come with changes to the overall peatland carbon flux Belyea and Malmer (2004); Loisel and Bunsen (2020).

For the prediction models, an SVR was set up for every combination between the six bioindicator parameters and the 22 spectral products for a total of 132 observation-prediction combinations. Additionally, the same variable combinations were replicated for the linear SVR design for a total of 264 model combinations altogether. With many models to train, efficient methods would be used during the model optimization process to cut the time of model optimization down from days and weeks to several hours.

Each model combination needed to be optimized first. The cost and kernel parameters for the SVR will affect how efficiently and precisely it finds a solution in the data. A set of 33 cost parameter values, $[2^{-10}, 2^{-9}, ..., 2^{22}]$, and 23 kernel gamma values, $[2^{-18}, 2^{-17}, .., 2^{4}]$, were set up in a cross-fold validation grid-search optimizer. With a total of 759 parameter combinations and K-fold cross-validation, an efficient optimal model search method needed to be used. Successive halving was utilized for larger prediction sets and random search for smaller prediction sets. Successive halving trains all model configuration candidates on a small subset of the training data. After cross-validating, the best performers continue to the next round, where more samples are used to train. This goes on until the final two candidates are left with a large proportion of the training samples to use for selecting the best model.

The random search was instead used for the smaller sample-size spectral product sets. The resistance, recovery, and resilience observations were calculated per drought-stress-induced mesocosm over the whole course of the experiment, with one observation value calculated using mesocosm traits from sessions 2, 3, and 4. Due to this, session two samples were selected as the predictor of recovery, resistance, etc., as they represent the mesocosms before drought stress and could potentially indicate how well the specimen would respond to drought based on the current

state. The dataset size for these ecological stress response model combinations (eco-stress models) was 47 (the drought-variable half of the 96 mesocosms minus one culled sample). With a low sample size, successive halving is challenging to employ reliably without creative data augmentation. Random search optimization, however, proves to be highly efficient with a minor reduction in optimal selection probability (**?**).

Experimental data obtained from Kokkonen et al. (2023) contain observation values for their measurement campaigns 1, 3, and 5. These measurement campaigns correspond to hyperspectral imaging sessions 2, 3, and 4, respectively (see Fig. 1.3. This brings the dataset size for the remaining spectral products down to 3 sessions of 96 samples each for 287 predictor samples (288 minus one culled sample).

Initial optimization surveys augmented the datasets with a Gaussian noise modifier. This was used to increase the sample size by 2 and 3 fold. Model scoring from these augmented datasets did not appear to improve results reliably. A separate experiment could subsequently be run to determine advantageous methods for augmenting hyperspectral data for bioindicator estimation.

Before models were optimized, the individual datasets for each one needed to be divided into training and testing data. Except for the eco-stress models, all models have three imaging sessions worth of data. Each session replicates the others, except that the drought treatment mesocosms receive their treatment only in session 3; nevertheless, there are corresponding trait measurements for this. Within each session, four groups are overlapping four sets, each one a replica of all possible history x site x species type x treatment modifier combinations. Considering the time series of each session as another modifier, the test set could be built from a stratified 1/3 of the dataset (96 samples) without sacrificing any balancing of the test data. With the thorough experimental setup given to work with, it made sense to preserve a fully balanced replica for the test set, leaving 2/3 (192 samples minus one culled) for the training. The dataset was shuffled and then split into stratified groups, ensuring the same proportion of all modifier combinations, including imaging session number, were alloted. The training set was then used in the model optimization process.

During optimization and for the final prediction, the coefficient of determination (3.33), coefficient of determination, was used as the primary scoring method. The explained variance was also trialed for optimization searches, finding models with equally performant training scores as the coefficient of determination scored models, but proving not to be as reliable when validated on the test sets. The coefficient of determination metric can be used as a measure of how well a predictive model explains the observation data. It measures the variance of observed data that is explained by the predictions. It will typically range from 0 to 1, with one meaning that all the variance is accounted for and 0 indicating that the model is simply

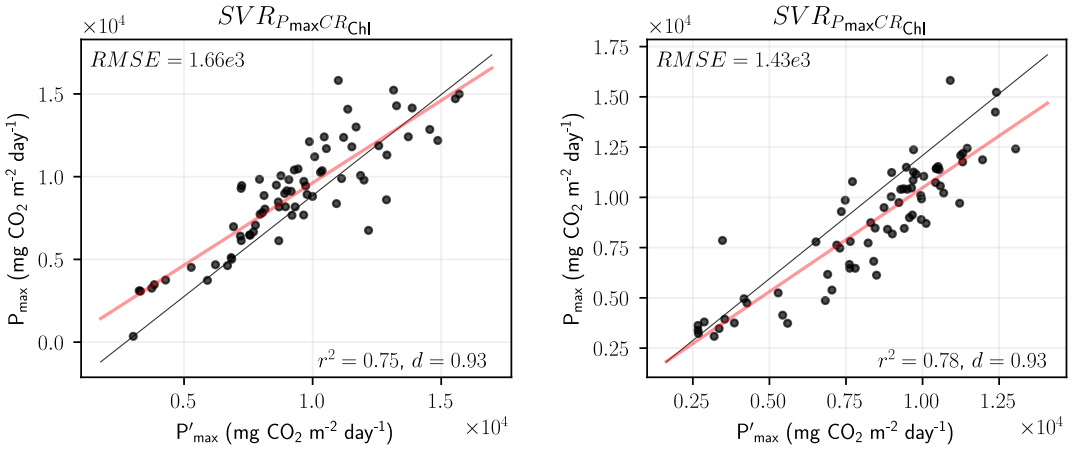


Figure 3.18: *Scores and RMSE error for the $SVR_{P_{max}CR_{Chl.}}$ model. The scatter point trendline is shown in red; the perfect linear trendline is shown in black.*

predicting the mean of the observations. As in this experiment, the value will extend to negative values for very poor-performing models.

$$R^2(y, y\prime) = 1 - \frac{\sum_{i=1}^{n}(y_i - y'_i)^2}{(y_i - \overline{y})^2} \tag{3.33}$$

$$d = 1 - \frac{\sum_{i}^{N} |y_i - y'_i|}{\sum_{i}^{N} |y'_i - \overline{y}| + |y_i - \overline{y}|} \tag{3.34}$$

where y_i is the observation value for prediction y'_i
and $\overline{y}$ is the mean of all y_n observation values

(3.35)

To complement the coefficient of determination measure of model performance, the Willmott index of agreement (3.34) was computed for each model's test set predictions. Assuming perfect predictions will be linearly related to the observations, index of agreement measures how well the magnitude and sign of deviation in observations agree with that of the predictions. If the observations and predictions are plotted against each other as in Figure 3.18, where a linear fit to the points in red can be seen aligned with the perfectly linear line, index of agreement is a measure of how closely the red linear trendline of the data fits the perfectly linear line.

All data was normalized to have a mean of zero and a variance of one. Upon optimal model selection from the optimization searches, the models were refitted onto the entire training dataset and subsequently used to predict the test set values.

Optimization of the models was the most time-consuming task. Still, it was cut down to around 3-4 hours of optimization time for all 264 model variants each time the optimization was run (several trials were run and subsequent attempts to verify and improve optimization convergence with search parameter settings). This, as well as all other tasks in this project, were performed on a personal computer with an Intel Core i7 processor, 16 GB of RAM, and an NVIDIA GeForce GTX 1060 Mobile - 6 GB VRAM GPU, though the GPU was not used for most tasks.

4 Results

Processing methods from Chapter 3 were applied to the dataset, and the resulting spectral products were used as input variables to SVR models as predictors of the experimental measurements and predictions of mesocosm bioindicator values obtained from researchers (Kokkonen et al., 2023). A comprehensive experiment was set up to answer the research questions posed earlier in Chapter 1, list 1.3.

4.1 Processed Dataset

The first result of this project work is a dataset of spectral products from the comprehensively designed peatland mesocosm experiment (Kokkonen et al., 2023), with the two primary products being the reflectance and absorbance hyperspectral cubes for each of the original mesocosm raw hyperspectral images.

The original experiment by Kokkonen et al. (2023) collected a thorough sampling of moss specimens from the three peatland sites of variable history (WET, DRY, and WLD). At each site, samples were collected from ombrotrophic areas (bogs) and minerotrophic areas (fens), with generalist and specialist species collected from each site. Hyperspectral images of each specimen over the four measurement sessions were used as the primary source of data in this project, used to derive the (reflectance factor) of the specimens. As a reflectance product, the dataset serves as a standard, transferrable derived product, capable of being used in further experiments to cross-train similar existing peatland or bryophyte prediction models. This dataset was initially used in this project to inspect variations in specimen response to the experimentally imposed climatic factors and aided in identifying spectral regions of interest where variations in response to climate were detected. In Figure 4.1, the mean reflectances of all species, irrespective of their site or history, are plotted for each measurement session. Apparent differences are observed between species, though it is important to remember that some species (*S. balticum*, *S. recurvum coll.*, *S. magellanicum*) are generalist species or were collected from multiple peatland sites with varying history or water source types. Therefore, the average of their spectra may not reveal peculiarities between responses concerning

the varying mesocosm environments.

To account for varying peatland sites, similar plots can be made for all samples grouped by history or site rather than species to study the effects of peatland site water table history and composition on the mesocosm climate responses. In this mesocosm experiment, (Kokkonen et al., 2023) found that naturally dry and WLD mesocosms did not display more resilience to drought than did the mesocosms from high water table peatlands. They concluded that species from the drier peatlands had not adapted a drought resistance, suggesting instead that peatlands experiencing growing periods of drought would alter turf cover towards more vascular species and hummock-adapted moss species. This would invariably have implications for the carbon cycle of these drought-stressed peatlands. Plotting the spectra of all grouping combinations from the dataset variables (history, site, treatment, species) would be copious. Instead, some of the predicted health metrics will be used to assess the performance of mesocosms across all of the experimental variables. First, the predictive models and their results will be discussed; then, spectral products with the strongest ability to predict some of the mesocosm photosynthesis bioindicators will be used to plot the relative performance of all mesocosm groupings throughout the experiment stages.

4.2 Predicting Bioindicators

SVR was selected as the machine learning model design for fitting the derived spectral products to the measured and estimated mesocosm experiment values from researchers' previous study from which this dataset was derived (Kokkonen et al., 2023). For every spectral product derived in this project (described in Table 3.2), an SVR model was created for each of the six experiment mesocosm values (described in Table 3.1). Hyperparameter optimization of the SVR kernel properties and the model cost function parameter for each model combination yielded an optimal configuration for predicting the experiment values given the particular spectral product. This totaled over 120 models, each iteratively optimizing several hundred parameter configuration candidates and fitting until convergence for each candidate. Using exhaustive optimization, the models would take days to optimize and fit on each run. Hyperparameter candidates could have been reduced to cut down optimization time. Still, an efficient optimization method was opted for instead, providing best-model selections, which were just as performant as exhaustively found best models.

Linear SVR models were fitted as well along with the originally selected RBF kernel SVR models. During the initial inspection, it was found that linear SVR models could predict the mesocosm photosynthetic metrics fairly accurately using some of the spectral products. The linear SVR models are less computationally

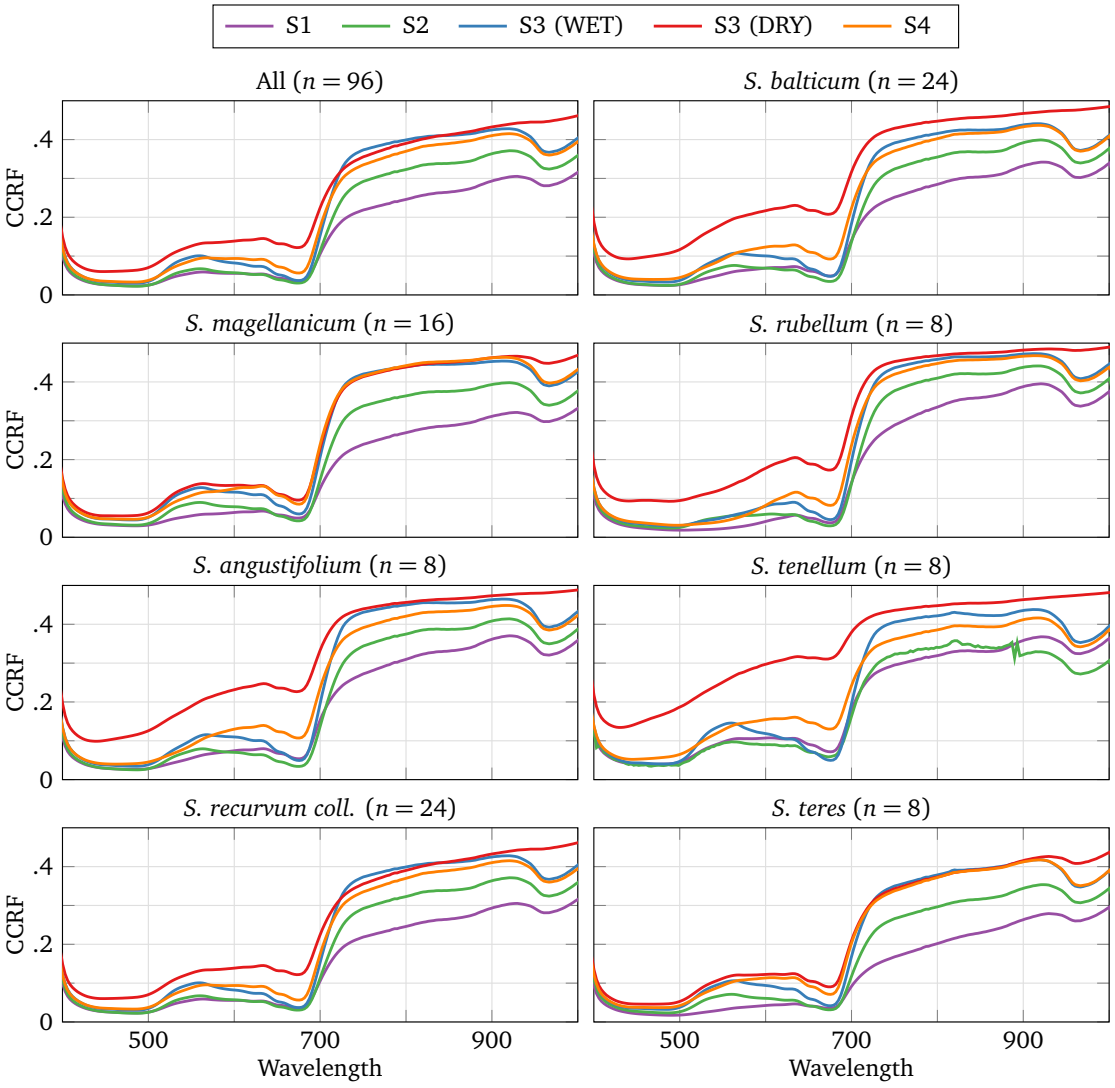


Figure 4.1: *Mean of each species over the four measurement sessions (S1-S4). Session 3 (S3) signals are divided into wet and dry sample signals, corresponding to the control half and variable half, respectively.*

expensive, generally optimizing and fitting more quickly than the RBF kernel SVR models.

Models were scored based on their coefficients of determination, r^2. This measures the proportion of unbiased variance in the dependent variable (e.g., P_{max}, P_{bio}, etc.) that is predictable by the independent variable (e.g., Rchl, RLD, etc.).

Below, in Figures 4.2 and 4.3, the r^2 and r^2 are shown for all of the optimized models' prediction results on the test set inputs. As the mesocosm experiment measurement values were only available for three measurement campaigns corresponding to hyperspectral imaging sessions 2, 3, and 4, there were 287 (96 mesocosm images from each session minus one culled image).

The top performing model, $SVR_{CRA_{Chl}P_{max}}$ (ranked first by r^2 and second by r^2), was subsequently used to predict P_{max} values for all mesocosms. P_{max} is plotted for all mesocosm hyperspectral images in Figure 4.4, grouped by their species and imaging session. They are shown alongside the observed P_{max} values obtained from Kokkonen et al. (2023).

4.3 Spatial Health Mapping

The best predictor model was then used to spatially predict the P_{max} across some hyperspectral images. The resulting P_{max} images are binned and shown alongside the MTVI2 images, measured mesocosm P_{max} and predicted P_{max} of the same samples for comparison. In Figure 4.6, control sample *01A* is rendered; P_{max} visibly has a solid correlation to MTVI2.

Looking at the similarly rendered results for the drought variable mesocosm sample *04C* in Figure 4.7, the correlation to MTVI2 is repeated, with some exceptions in areas with very low P_{max} values. The S2 MTVI2 rendering exhibits MTVI2 values of $.4 - -.8$ in some areas where the P_{max} is only $7.5e3 - -9e3$. Remember from Figure 3.13 that MTVI2 values of .4-.8 correspond to LAI values of around 1 to 4. LAI is crucial to ecosystem carbon flux modeling, but, if the leaf area is not actively photosynthesizing, respiration will be reduced. Perhaps that is revealed here with MTVI2 registering leaf area up to four layers thick in areas where the P_{max} is relatively low. In the lower range of MTVI2, it may become insensitive and unable to differentiate some inactive vegetation.

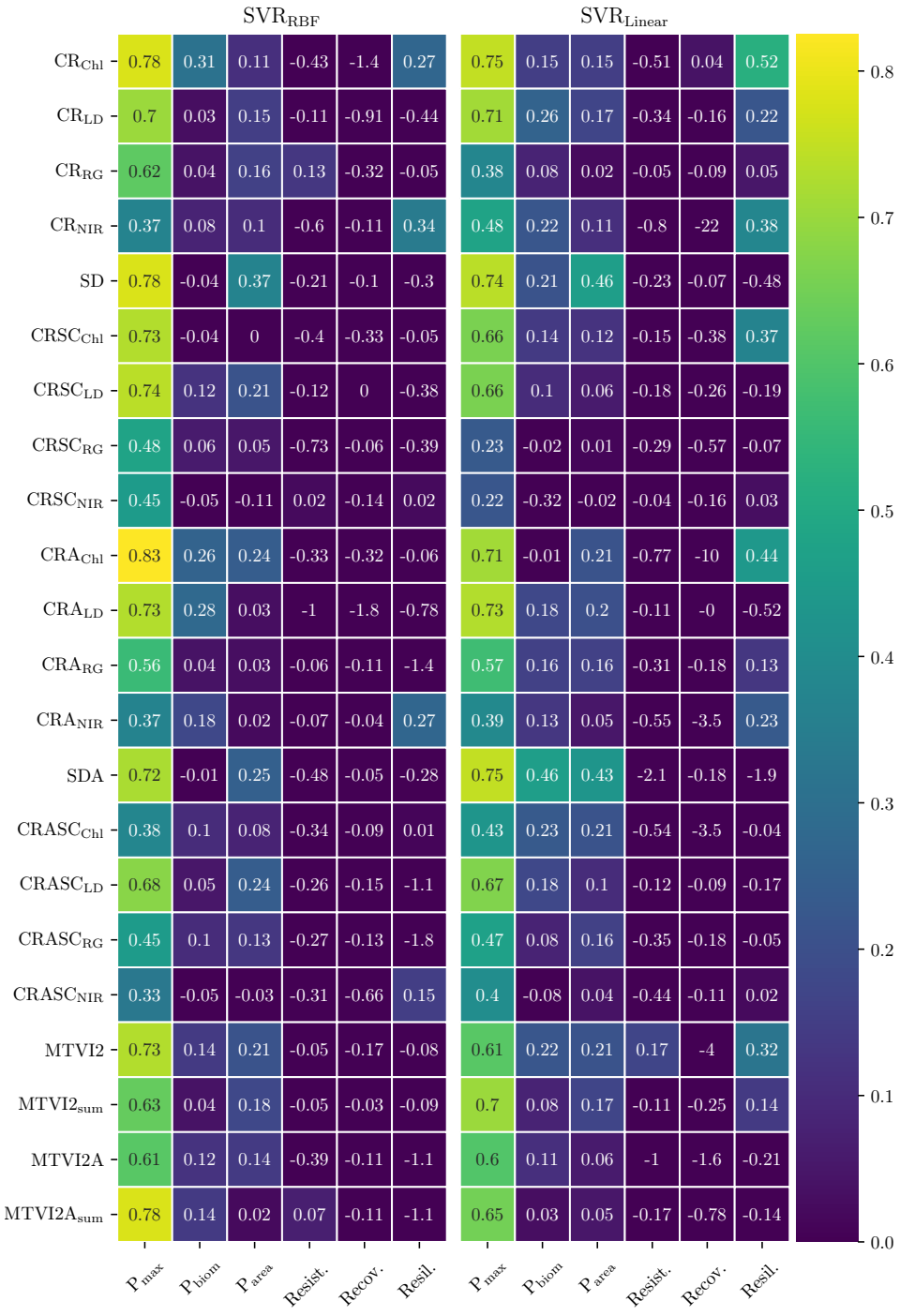


Figure 4.2: *Coefficient of determination, r^2 for all SVR model combinations*

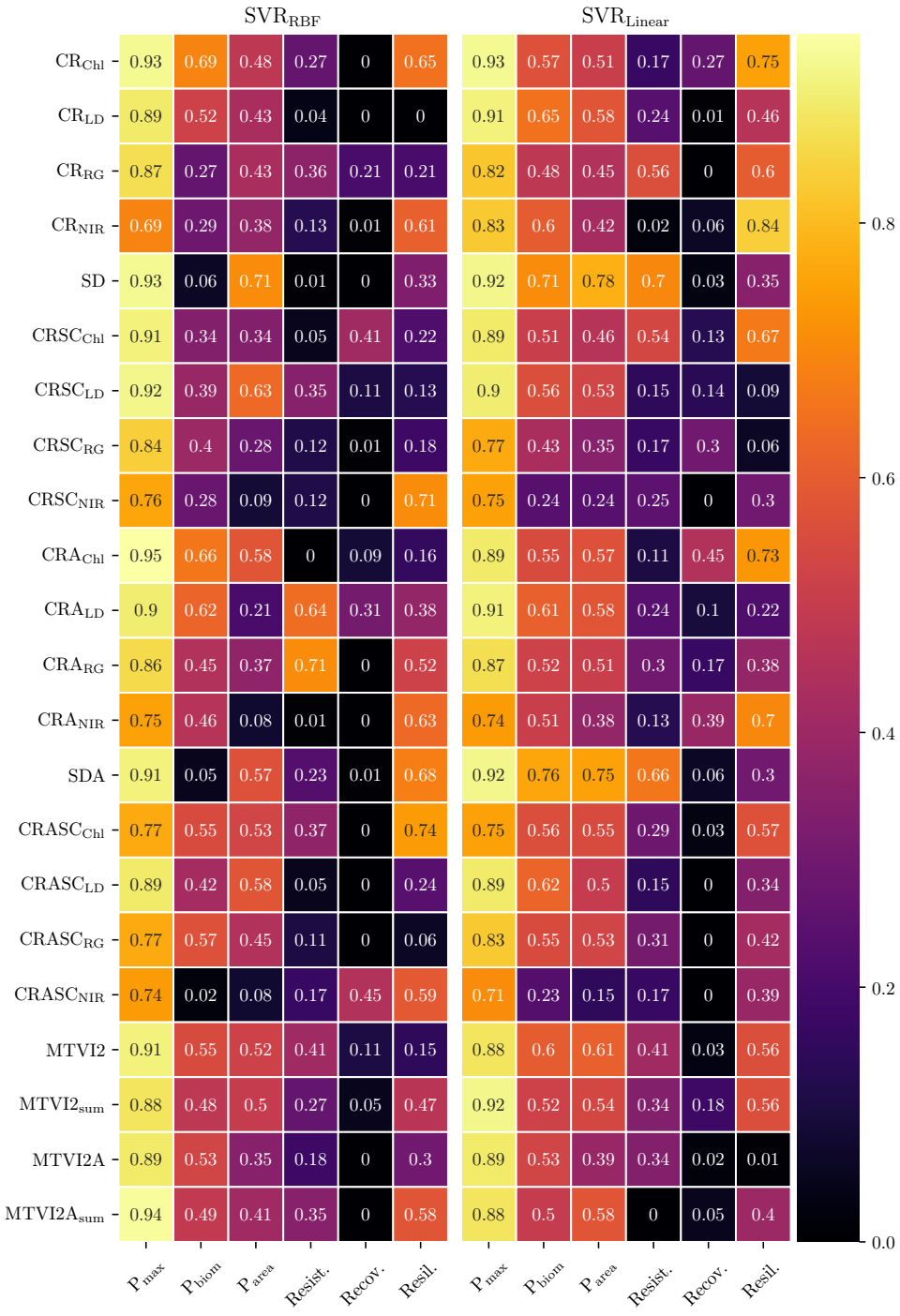


Figure 4.3: *Index of agreement, d, for all SVR model combinations*

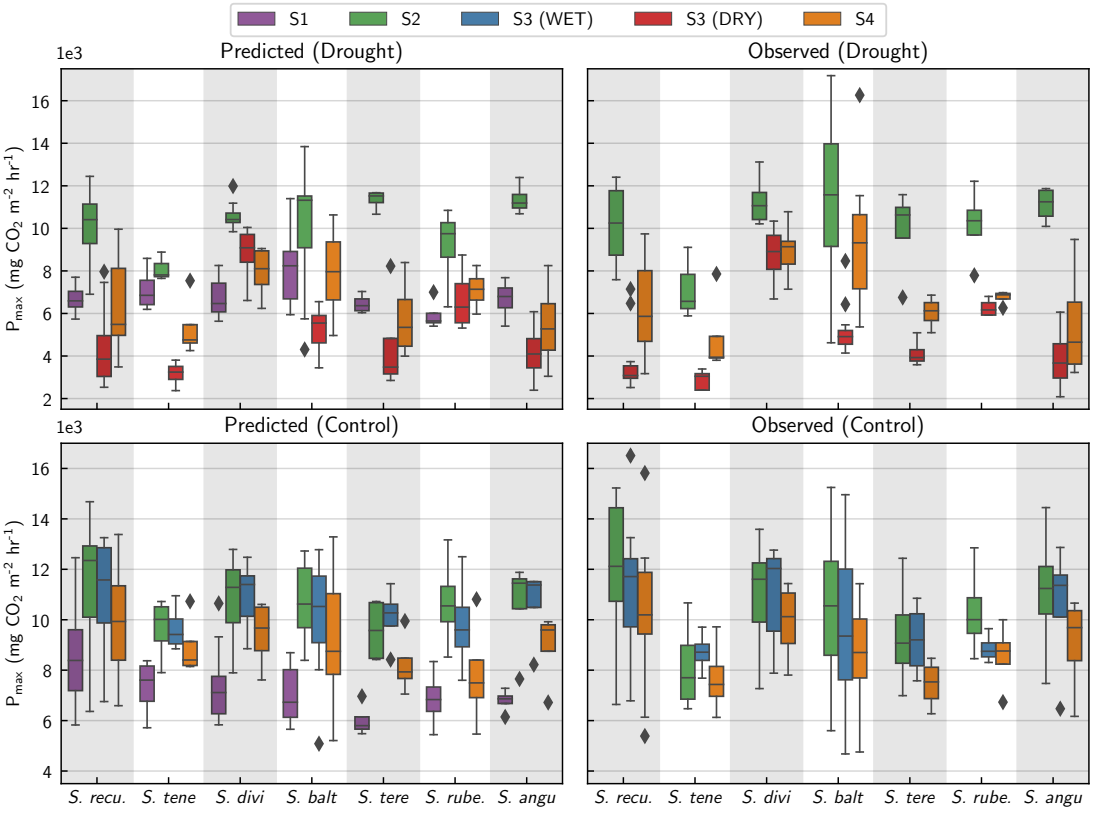


Figure 4.4: *Predicted and observed values of P_{max} for all specimens, grouped by species and imaging session.*

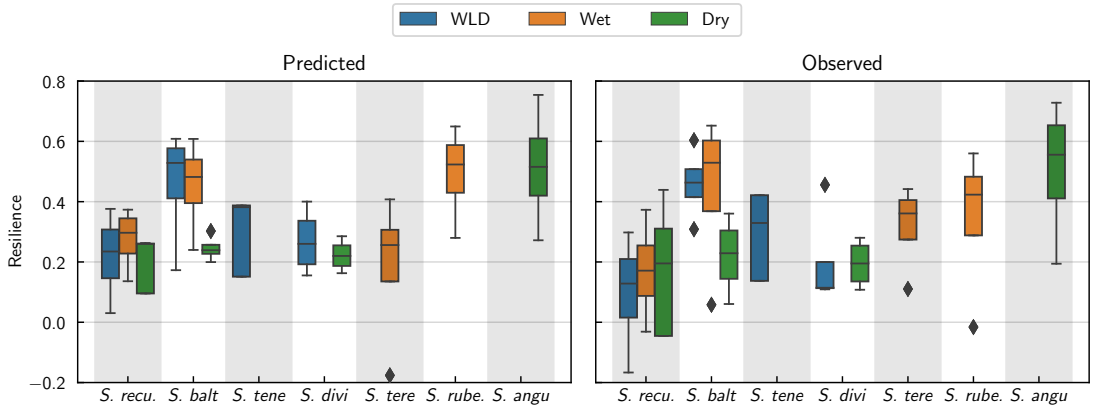


Figure 4.5: *Derived resilience from predicted and observed P_{max} of all variable specimens (drought stress), grouped by peatland site history.*

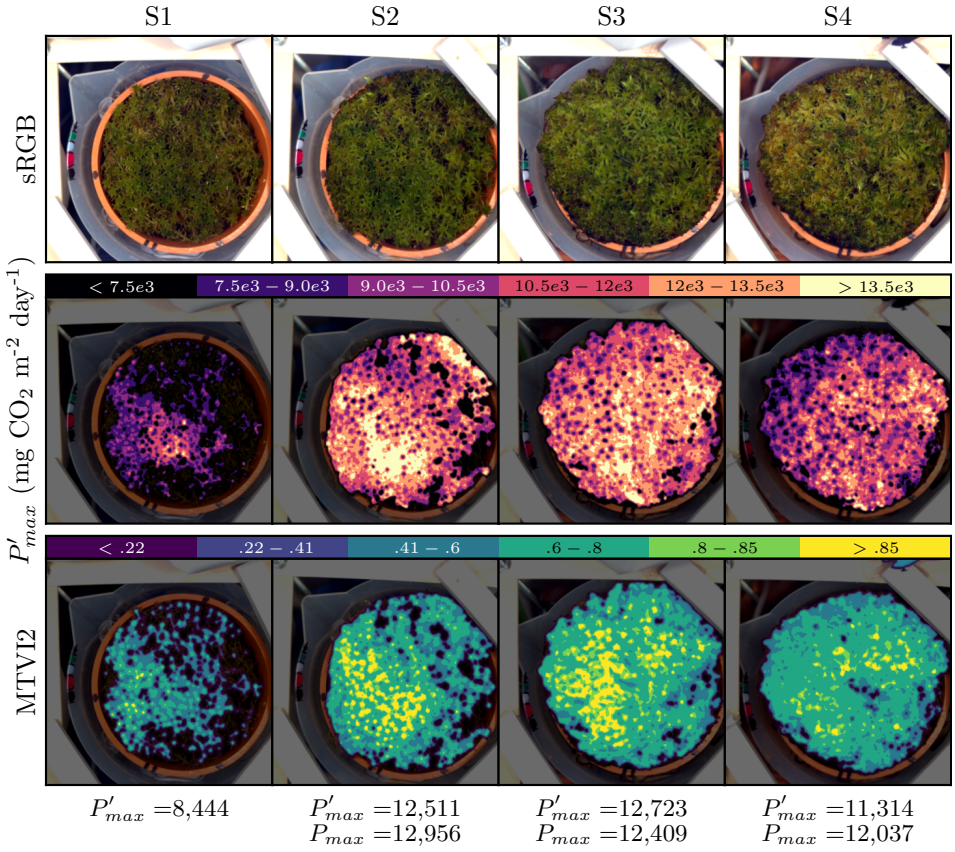


Figure 4.6: *Spatial prediction of P_{max} with model $SVR'_{P_{max}CR_{Chl.}}$*

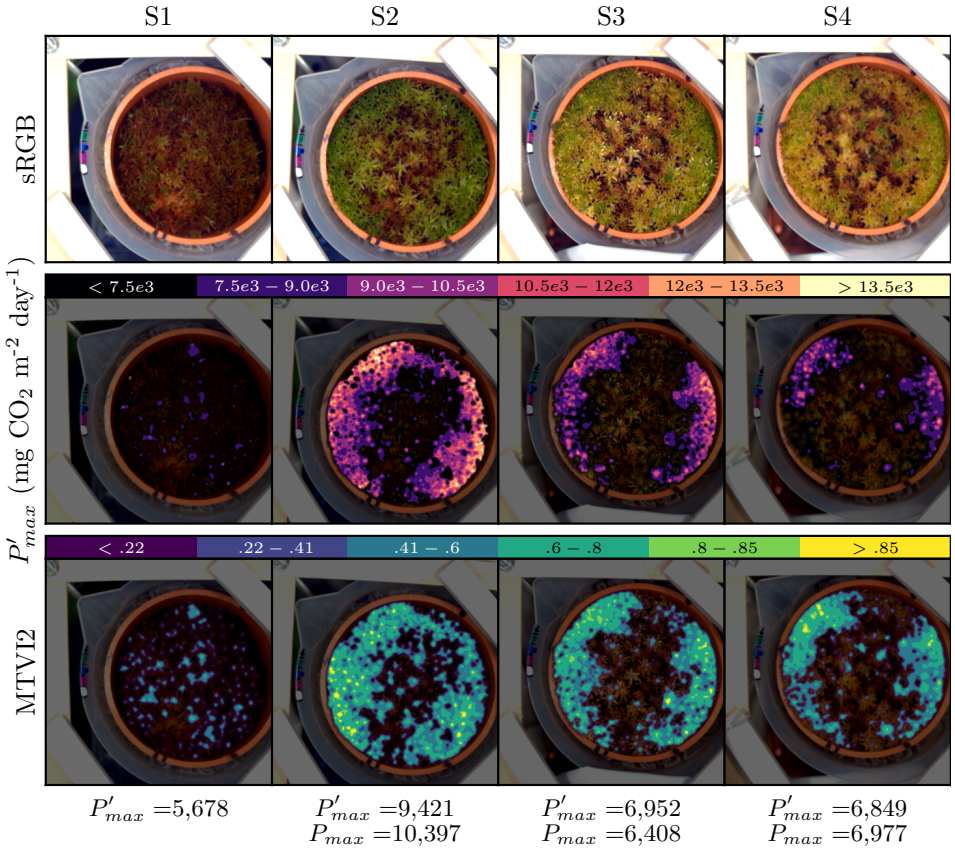


Figure 4.7: *Spatial prediction of sample 04C P_{max} with model $SVR'_{P_{max}CR_{Chl.}}$*

5 Discussion

The results obtained in this project answer some of the posed questions and open new ones. In response to the research questions raised throughout this report:

1. How do peatland bryophyte species respond to climate stresses?
2. Do hyperspectral data and spectral products derived thereof to serve as accurate covariate predictors of health during climate stress?
    - This has been shown
3. Are absorbance signals advantageous to reflectance when modeling biochemical parameters from spectral data?
    - Results indicate that training regression models on absorbance spectral products can produce better predictions for some parameters.
4. How do vegetation indices such as MTVI2 compare to spectral data in predictive ability?
    - In the results, MTVI2 products performed very well, as well as most of the normalized and pure spectral products
5. Can moss bio-resilience be predicted from hyperspectral data?
    - Some of the SVR models produced acceptable results for predicting the bio-resilience of the moss species

It was particularly interesting to find that the absorbance signal could produce better prediction results in some cases than the reflectance signal. The SVR models with nonlinear kernels are capable of reshaping input data to separable distributions, implying that an SVRs could produce the same results from reflectance and absorbance signals as the absorbance is simply a log of the reflectance images. None of the reflectance product-trained models, however, ever achieved scores as high as did $SVR_{P_{max}CRA_{Chl.}}$ with coefficient of determination=.83 and index of agreement=.95.

5.0.1 Enhancing results

The results obtained from the optimized and trained models could be improved with additional data processing, such as augmentation. Basic limited augmentation was trialed but did not seem to enhance model optimization.

The best predictive spectral products could be combined into one training set. Additionally, as spectral data is highly correlated, the most significant bands for training could be computed, reducing the size of training inputs and opening up more possibilities for input capture systems (multispectral or even three-channel RGB systems).

5.0.2 Further work

With the comprehensively designed mesocosm drought experiment and its associated dataset, there are several ways to extend this study.

1. Use specimen site and history labels to train a peatland classifier model.
2. Some countries will begin to require landowners to assess their forest biodiversity. This dataset can be used to build a moss species classifier. Moss species are known to be strong indicators of ecosystem succession (Vitt and House, 2021) and biodiversity (Pardow et al., 2012; Zechmeister et al., 2003) and a moss classifier could potentially quantify biodiversity in field-scale forestry.
3. Absorbance products were better predictors than reflectance products for some signals and traits. An experiment could be performed to assess the performance of absorbance VIs in comparison to standard reflectance VIs

Bibliography

Anderson, M., Motta, R., Chandrasekar, S., and Stokes, M. (1996). Proposal for a standard default color space for the internet—srgb. In *Color and imaging conference*, volume 1996, pages 238–245. Society for Imaging Science and Technology. Issue: 1. (cited on page 23)

Andriesse, J. (1988). *Nature and management of tropical peat soils*. Food & Agriculture Org. (cited on page 3)

Bartczak, P., Fält, P., Penttinen, N., Ylitepsa, P., Laaksonen, L., Lensu, L., Hauta-Kasari, M., and Uusitalo, H. (2017). Spectrally optimal illuminations for diabetic retinopathy detection in retinal imaging. *Optical Review*, 24:105–116. Publisher: Springer. (cited on page 25)

Belyea, L. R. and Malmer, N. (2004). Carbon sequestration in peatland: patterns and mechanisms of response to climate change. *Global Change Biology*, 10(7):1043–1052. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1529-8817.2003.00783.x. (cited on page 57)

Boser, B. E., Guyon, I. M., and Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152. (cited on page 54)

Bubier, J. L., Rock, B. N., and Crill, P. M. (1997). Spectral reflectance measurements of boreal wetland and forest mosses. *Journal of Geophysical Research: Atmospheres*, 102(D24):29483–29494. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/97JD02316. (cited on page 11)

Burdun, I., Bechtold, M., Sagris, V., Lohila, A., Humphreys, E., Desai, A. R., Nilsson, M. B., De Lannoy, G., and Mander, Ü. (2020). Satellite Determination of Peatland Water Table Temporal Dynamics by Localizing Representative Pixels of A SWIR-Based Moisture Index. *Remote Sensing*, 12(18):2936. Number: 18 Publisher: Multidisciplinary Digital Publishing Institute. (cited on pages 12 and 19)

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., and Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, 408:189–215. (cited on page 54)

Clark, R. N. and Roush, T. L. (1984). Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. *Journal of Geophysical Research: Solid Earth*, 89(B7):6329–6340. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/JB089iB07p06329. (cited on page 50)

Cocchi, M., Seeber, R., and Ulrici, A. (2003). Multivariate calibration of analytical signals by WILMA (wavelet interface to linear modelling analysis). *Journal of Chemometrics*, 17(8-9):512–527. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cem.819. (cited on page 17)

Dargie, G. C., Lewis, S. L., Lawson, I. T., Mitchard, E. T. A., Page, S. E., Bocko, Y. E., and Ifo, S. A. (2017). Age, extent and carbon storage of the central Congo Basin peatland complex. *Nature*, 542(7639):86–90. Number: 7639 Publisher: Nature Publishing Group. (cited on page 13)

English, A. R. (2023). Automated Design of Task-Dedicated Illumination with Particle Swarm Optimization. In *2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS)*, pages 416–421. IEEE. (cited on page 25)

Fenton, J. H. C. (1980). The Rate of Peat Accumulation in Antarctic Moss Banks. *Journal of Ecology*, 68(1):211–228. Publisher: [Wiley, British Ecological Society]. (cited on page 14)

Glime, J. M. (2013). Light: Effects of High Intensity. In *Bryophyte Ecology*, volume 1. Michigan Technological University. Section: 9-3. (cited on page 49)

González, E., Henstra, S. W., Rochefort, L., Bradfield, G. E., and Poulin, M. (2014). Is rewetting enough to recover Sphagnum and associated peat-accumulating species in traditionally exploited bogs? *Wetlands Ecol Manage*, 22(1):49–62. (cited on page 1)

Haboudane, D., Miller, J. R., Pattey, E., Zarco-Tejada, P. J., and Strachan, I. B. (2004). Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sensing of Environment*, 90(3):337–352. (cited on pages 49, 51, and 80)

Harris, L. I., Richardson, K., Bona, K. A., Davidson, S. J., Finkelstein, S. A., Garneau, M., McLaughlin, J., Nwaishi, F., Olefeldt, D., Packalen, M., Roulet, N. T., Southee, F. M., Strack, M., Webster, K. L., Wilkinson, S. L., and Ray, J. C. (2022). The essential carbon service provided by northern peatlands. *Frontiers in Ecology and the Environment*, 20(4):222–230. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/fee.2437. (cited on page 13)

Hicks Pries, C. E., Schuur, E. A., and Crummer, K. G. (2013). Thawing permafrost increases old soil and autotrophic respiration in tundra: Partitioning ecosystem respiration using δ13C and? 14C. *Global Change Biology*, 19(2):649–661. Publisher: Wiley Online Library. (cited on page 1)

Hicks Pries, C. E., Van Logtestijn, R. S., Schuur, E. A., Natali, S. M., Cornelissen, J. H., Aerts, R., and Dorrepaal, E. (2015). Decadal warming causes a consistent and persistent shift from heterotrophic to autotrophic respiration in contrasting permafrost ecosystems. *Global change biology*, 21(12):4508–4519. Publisher: Wiley Online Library. (cited on page 1)

HOLLAS, J. (1984). FUNDAMENTALS OF MOLECULAR-SPECTROSCOPY-BANWELL, CN. (cited on page 44)

Huang, Z., Turner, B. J., Dury, S. J., Wallis, I. R., and Foley, W. J. (2004). Estimating foliage nitrogen concentration from HYMAP data using continuum removal analysis. *Remote Sensing of Environment*, 93(1-2):18–29. Publisher: Elsevier. (cited on page 50)

Jafarzadeh, N., Rajaee, T., Daneh Kar, A., Robati, M., and others (2020). Evaluation of river bio-resilience with artificial intelligence models (case study: Aliabad river). *Iranian Water Researches Journal*, 14(3):61–70. Publisher: Shahrekord University. (cited on page 54)

Kalacska, M., Lalonde, M., and Moore, T. R. (2015). Estimation of foliar chlorophyll and nitrogen content in an ombrotrophic bog from hyperspectral data: Scaling from leaf to image. *Remote Sensing of Environment*, 169:270–279. (cited on pages 16, 17, and 83)

Kalacska, M., Sanchez-Azofeifa, G. A., Rivard, B., Caelli, T., White, H. P., and Calvo-Alvarado, J. C. (2007). Ecological fingerprinting of ecosystem succession: Estimating secondary tropical dry forest structure and diversity using imaging spectroscopy. *Remote Sensing of Environment*, 108(1):82–96. (cited on page 17)

Keenan, T. F., Migliavacca, M., Papale, D., Baldocchi, D., Reichstein, M., Torn, M., and Wutzler, T. (2019). Widespread inhibition of daytime ecosystem respiration.

Nat Ecol Evol, 3(3):407–415. Number: 3 Publisher: Nature Publishing Group. (cited on page 12)

Kokkonen, N., Laine, A. M., Korrensalo, A., Nijp, J., Limpens, J., Metselaar, K., Mehtätalo, L., Männistö, E., and Tuittila, E.-S. (2023). The photosynthetic response of lawn Sphagna to experimental drought and subsequent recovery as impacted by moss traits and peatland ecohydrology. manuscript provided by Kokkonen. (cited on pages 5, 9, 46, 47, 57, 58, 61, 62, 64, and 83)

Kokkonen, N. A. K., Laine, A. M., Laine, J., Vasander, H., Kurki, K., Gong, J., and Tuittila, E.-S. (2019). Responses of peatland vegetation to 15-year water level drawdown as mediated by fertility level. *Journal of Vegetation Science*, 30(6):1206–1216. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jvs.12794. (cited on page 5)

Kováč, D., Malenovskỳ, Z., Urban, O., Špunda, V., Kalina, J., Ač, A., Kaplan, V., and Hanuš, J. (2013). Response of green reflectance continuum removal index to the xanthophyll de-epoxidation cycle in Norway spruce needles. *Journal of experimental botany*, 64(7):1817–1827. Publisher: Oxford University Press UK. (cited on page 50)

Kováč, D., Navratil, M., Malenovskỳ, Z., Štroch, M., Špunda, V., and Urban, O. (2012). Reflectance continuum removal spectral index tracking the xanthophyll cycle photoprotective reactions in Norway spruce needles. *Functional Plant Biology*, 39(12):987–998. Publisher: CSIRO Publishing. (cited on page 50)

Lees, K. J., Artz, R. R. E., Khomik, M., Clark, J. M., Ritson, J., Hancock, M. H., Cowie, N. R., and Quaife, T. (2020). Using Spectral Indices to Estimate Water Content and GPP in Sphagnum Moss and Other Peatland Vegetation. *IEEE Transactions on Geoscience and Remote Sensing*, 58(7):4547–4557. Conference Name: IEEE Transactions on Geoscience and Remote Sensing. (cited on page 20)

Lees, K. J., Quaife, T., Artz, R. R. E., Khomik, M., and Clark, J. M. (2018). Potential for using remote sensing to estimate carbon fluxes across northern peatlands – A review. *Science of The Total Environment*, 615:857–874. (cited on pages 2, 11, and 48)

Leifeld, J. and Menichetti, L. (2018). The underappreciated potential of peatlands in global climate change mitigation strategies. *Nat Commun*, 9(1):1071. Number: 1 Publisher: Nature Publishing Group. (cited on page 13)

Loisel, J. and Bunsen, M. (2020). Abrupt fen-bog transition across southern Patagonia: Timing, causes, and impacts on carbon sequestration. *Frontiers in Ecology and Evolution*, 8:273. Publisher: Frontiers Media SA. (cited on page 57)

Loisel, J., Yu, Z., Beilman, D. W., Kaiser, K., and Parnikoza, I. (2017). Peatland Ecosystem Processes in the Maritime Antarctic During Warm Climates. *Sci Rep*, 7(1):12344. Number: 1 Publisher: Nature Publishing Group. (cited on page 14)

Malenovský, Z., Turnbull, J. D., Lucieer, A., and Robinson, S. A. (2015). Antarctic moss stress assessment based on chlorophyll content and leaf density retrieved from imaging spectroscopy data. *New Phytologist*, 208(2):608–624. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/nph.13524. (cited on pages 6, 14, 15, 16, and 49)

Minasny, B., Berglund, Ö., Connolly, J., Hedley, C., de Vries, F., Gimona, A., Kempen, B., Kidd, D., Lilja, H., Malone, B., McBratney, A., Roudier, P., O'Rourke, S., Rudiyanto, Padarian, J., Poggio, L., ten Caten, A., Thompson, D., Tuve, C., and Widyatmanti, W. (2019). Digital mapping of peatlands – A critical review. *Earth-Science Reviews*, 196:102870. (cited on pages 2, 3, and 11)

Pardow, A., Gehrig-Downie, C., Gradstein, R., and Lakatos, M. (2012). Functional diversity of epiphytes in two tropical lowland rainforests, French Guiana: using bryophyte life-forms to detect areas of high biodiversity. *Biodiversity and Conservation*, 21:3637–3655. Publisher: Springer. (cited on pages 4 and 72)

Pasolli, E., Melgani, F., Alajlan, N., and Bazi, Y. (2012). Active Learning Methods for Biophysical Parameter Estimation. *IEEE Transactions on Geoscience and Remote Sensing*, 50(10):4071–4084. Conference Name: IEEE Transactions on Geoscience and Remote Sensing. (cited on page 54)

Proctor, M. C. F. and Goffinet, B. (2008). Physiological ecology. In Shaw, A. J., editor, *Bryophyte Biology*, pages 237–268. Cambridge University Press, 2 edition. (cited on page 49)

Robinson, S. A., Turnbull, J. D., and Lovelock, C. E. (2005). Impact of changes in natural ultraviolet radiation on pigment composition, physiological and morphological characteristics of the Antarctic moss, Grimmia antarctici. *Global Change Biology*, 11(3):476–489. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2486.2005.00911.x. (cited on page 16)

Rydin, H. and Jeglum, J. (2013). *The biology of peatlands, second edition.* (cited on page 13)

Salko, S.-S., Juola, J., Burdun, I., Vasander, H., and Rautiainen, M. (2023). Intra- and interspecific variation in spectral properties of dominant Sphagnum moss species in boreal peatlands. *Ecology and Evolution*, 13(6):e10197. _eprint:

https://onlinelibrary.wiley.com/doi/pdf/10.1002/ece3.10197. (cited on pages 19 and 83)

Semenov, D. (2019). Hyperspectral Dataset Setup Code for Lakkasuo Mesocosm Experiment. Dataset exploratory code provided by Dmitry Semenov. (cited on page 5)

Tuia, D., Verrelst, J., Alonso, L., Pérez-Cruz, F., and Camps-Valls, G. (2011). Multioutput support vector regression for remote sensing biophysical parameter estimation. *IEEE Geoscience and Remote Sensing Letters*, 8(4):804–808. Publisher: IEEE. (cited on page 54)

van Breemen, N. (1995). How Sphagnum bogs down other plants. *Trends in Ecology & Evolution*, 10(7):270–275. (cited on page 1)

Viitakoski, M. (2023). Specim IQ Technical Specifications. (cited on page 7)

Vitt, D. and House, M. (2021). Bryophytes as key indicators of ecosystem function and structure of northern peatlands. *Bryophyte Diversity and Evolution*, 43. (cited on pages 4, 13, and 72)

Xu, J., Morris, P., Liu, J., and Holden, J. (2018). PEATMAP: Refining estimates of global peatland distribution based on a meta-analysis. *Catena*, 160:134–140. (cited on page 1)

Zechmeister, H., Tribsch, A., Moser, D., Peterseil, J., and Wrbka, T. (2003). Biodiversity 'hot spots' for bryophytes in landscapes dominated by agriculture in Austria. *Agriculture, ecosystems & environment*, 94(2):159–167. Publisher: Elsevier. (cited on pages 4 and 72)

Zotz, G. and Kahler, H. (2007). A moss "canopy" – Small-scale differences in microclimate and physiological traits in Tortula ruralis. *Flora - Morphology, Distribution, Functional Ecology of Plants*, 202(8):661–666. (cited on page 49)

List of Figures

LIST OF FIGURES

List of Tables