Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Faculty of Landscape Architecture, Horticulture and Crop Production Science

PlantSpec

– An R package for interactive analysis of spectral data for high-throughput plant phenotyping

Alexander Koc

Degree Project • 30 hec
Alnarp 2017

PlantSpec – An R package for interactive analysis of spectral data for high-throughput plant phenotyping

PlantSpec – Ett tilläggspaket för R för interaktiv analys av spektraldata för höghastighetsfenotypning

Alexander Koc

Supervisor: Aakash Chawade, SLU, Department of Plant Breeding
Examiner: Erik Alexandersson, SLU, Department of Plant Protection Biology

Credits: 30 hec
Project level: A2E
Course Title: Master's Thesis Project in Biology
Course Code: EX0800
Subject: Biology

Place of Publication: Alnarp
Year of Publication: 2017
Cover Art: N/A
Online Publication: http://stud.epsilon.slu.se

Keywords: Hyperspectral data, high throughput phenotyping, R programming, Bioinformatics

SLU, Swedish University of Agricultural Sciences
Faculty of Landscape Architecture, Horticulture and Crop Production Science
Department of Plant Breeding

Summary

Hyperspectral phenotyping is a promising high-throughput plant phenotyping method that addresses the phenotyping bottleneck, in fields of plant science like crop improvement or study of plant diseases. However, high-throughput hyperspectral data poses a data management and analysis problem due to the sheer volume of data generated in the data collection process. The R Package presented in this paper seeks to address this problem by implementing a tool for easy and quick handling of large spectral datasets, and to provide functions for processing and analyzing of spectral data. This project also aims to implement functions in the package for calculating vegetation indices, and data visualization functions. Furthermore, this R package is extended into an online tool, that provides and interactive graphical user interface to the underlying R code.

Acknowledgements

I would like to thank my supervisor Aakash Chawade for his support, patience, and guidance. Furthermore, I would like to offer my gratitude to Tina Henriksson at Lantmännen Lantbruk, for permission to use the case study data. I would also like to extend my thanks to everyone at the Department of Plant Breeding for a fun and supportive work environment, and to Tanja, for hunting down all my stray commas again.

Table of contents

Abbreviations

- NDVI = Normalized Difference Vegetation Index
- NIR = Near Infrared
- PCA = Principal Component Analysis
- PRI = Photochemical Reflectance Index
- REIP = Red Edge Inflection Point
- WI = Water Index

Introduction

The high-throughput plant phenomics (HTPP) techniques aim to address the plant-phenotyping bottleneck (Furbank & Tester, 2011), by designing phenotyping methods and platforms that are cost-effective, specific, rapid and adaptable to robotics and automation (Fahlgren, Gehan, & Baxter, 2015). For example, the study of plant disease in fields like crop improvement relies on manual visual scoring, while molecular techniques, such as Polymerase Chain Reaction and Enzyme-Linked Immunosorbent Assays, used for studying and identifying plant pathogens, require lab-work and pathogen-specific reagents. While in forestry, the assessment of tree health is primarily done through scouting (Sankaran, Mishra, Ehsani, & Davis, 2010).

Precision agriculture

Rapid assessment of plant traits is also of interest in precision farming. In conventional phenotyping practices, the variation of crop conditions and resources that exists in a field is often not considered, and fields are often treated as homogenous units (Diacono, Rubino, & Montemurro, 2013). For example, fertilizer is often over-applied by farmers, who want to ensure that their crops get enough nitrogen. This can result in the contamination of water bodies as a result of nitrogen runoff and economic loss due to inefficient use of nitrogen (Diacono et al., 2013; Khosla, Fleming, Delgado, Shaver, & Westfall, 2002). According to Diacono et al (2013), in the case of wheat, the over-application of nitrogen may also result in "weed problems and could result in an increased risk of lodging, delayed maturity and greater wheat susceptibility to diseases".

Precision agriculture seeks to address the abovementioned problems by providing a sustainable and site-specific approach to farming, that considers the variation in the field when applying fertilizer or pesticide treatments (Pierce & Nowak, 1999). Among the key technologies to achieve this are better sensor-based phenotyping technologies for accurately quantifying site-specific crop properties like disease outbreaks, or nitrogen status (Diacono et al., 2013). Thus, better HTPP methods can enable precision farming and contribute to more sustainable and economic farming practices. Furthermore, sensor-based technologies could be extended to be used on unmanned or autonomous vehicles, such as drones,, enabling automated assessment of crop status.

Spectral phenotyping

Spectral HTPP methods focus on collecting spectral and imaging data from plant tissue using remote sensing data from satellites or ground-based sensors (Sankaran et al., 2010). The idea is

to collect data on light reflected from the plant tissue, or data from fluorescence experiments, and use that data to estimate a trait or a plant condition like biotic or abiotic stresses. Reflectance occurs when light is reflected from a surface (as opposed to absorbance). For example, as is commonly seen in case of plant tissue, wavelengths of light that correspond to the colors red and blue are absorbed by pigments, while green light is reflected (Furbank & Tester, 2011). This is what makes plant tissue like leaves appear green to the human eye. Hyperspectral phenotyping goes beyond the visible spectrum (400 – 700 nm) by collecting reflectance data from a large amount of narrow spectral-bands in the visible spectrum and beyond (Fahlgren et al., 2015), for example in the near-infrared (NIR) region (700 – 1000 nm), where plant tissue typically shows high reflectance (Furbank & Tester, 2011).

The idea is to identify changes in reflectance in specific plants caused by specific traits or stresses, then characterizing those changes, for example by identifying wavelengths that are most affected, and using that data to build predictive models that can estimate a trait or a stress condition (Fahlgren et al., 2015). A typical plant leaf will show low reflectance in the visible region (400 – 700 nm), high reflectance in the NIR region and a sharp increase in reflectance around 700 nm, known as the Red Edge, as shown in Figure 1. As previously mentioned, the low reflectance in the visible region is due to light absorbance by leaf pigments. Some examples of important pigments are chlorophylls and carotenoids, which are crucial for light harvesting in photosynthesis. Other examples of pigments include the flavonoids and anthocyanins, which can protect plant tissue against both UV-radiation and excessive radiation in the visible spectrum (Merzlyak, et al., 2003). The high reflectance in the NIR region is due to the internal cell structure of the leaf, which reflects about 40-60% of the light in this range, while the rest is transmitted (Knipling, 1970; Josep Peñuelas & Filella, 1998).

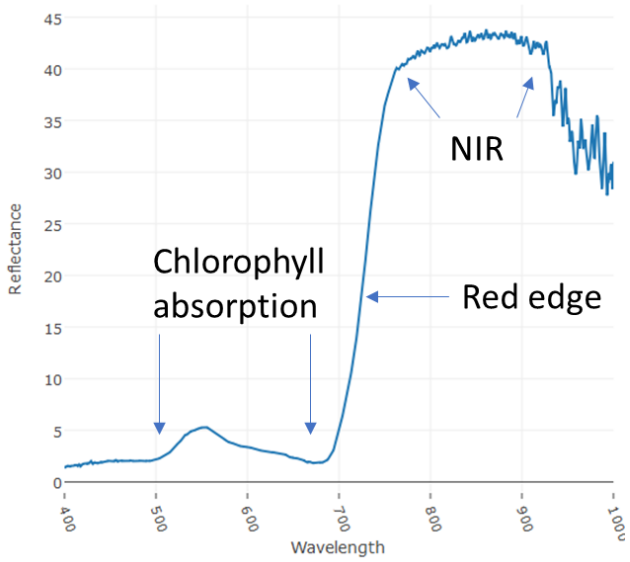


Figure. 1 An example of a reflectance spectrum of plant tissue.

Reflectance thus depends on structural and biochemical properties of the leaf, and it is thought that studying reflectance would allow measuring those underlying plant properties in a non-

destructive way. For example, the red edge region can be used as an indicator of chlorophyll content in a leaf, which is associated with plant traits such as plant nitrogen status, but also affected by other forms of plant stress, while individual wavelengths in the visible spectrum are associated with specific pigments (Merzlyak et al., 2003; Josep Peñuelas & Filella, 1998).

Spectral data

Spectral data can be collected in an imaging and non-imaging format (Furbank & Tester, 2011). Imaging data contains the spatial distribution of measured intensities of a range of wavelengths, much like color values per pixel in a photograph from a consumer-grade digital camera. Non-imaging spectral data forgoes the spatial information and instead collects the average intensity data across the whole measured area. Typically, a spectral readout will contain an intensity value for each measured band.

Vegetation indices

As mentioned previously, the reflectance at individual spectral bands in vegetation can be tied to specific biochemical or structural properties, such as for example the chlorophyll concentration. The reflectance at these specific bands can be summarized in vegetation indices, which are unitless ratios or differences between measured wavelengths. The indices are used to enhance or highlight specific plant properties (Agapiou, Hadjimitsis, & Alexakis, 2012; Sankaran et al., 2010). There are over 100 vegetation indices with different uses, and the most commonly used index is the normalized difference vegetation index (NDVI, Rouse, Haas, Schell, & Deering, 1973), which is expressed as the "quotient of the difference and sum of the reflectance in NIR and red regions" (Wojtowicz, Wojtowics, & Piekarczyk, 2015). NDVI has been shown to be related to other plant properties such as biomass, nitrogen and chlorophyll (Hansen & Schjoerring, 2003; Resea, Molero, Group, Agricultur, & Nogu, 2014). Other examples of specific vegetation indices include the water index (WI), that measures plant water status (J. Peñuelas, Filella, Biel, Serrano, & Savé, 1993), or the Photochemical Reflectance Index (PRI), which is affected by carbon dioxide uptake or photosynthetic efficiency (Gamon, Serrano, & Surfus, 1997).

In conclusion, spectral phenotyping shows high potential as a HTPP method. Both fluorescence and reflectance can be measured without destroying plant tissue. The measurements themselves are rapid and avoid the time-consuming laboratory work associated with molecular techniques, and thus lend themselves well to initial screening for a trait of interest. The cost-effectiveness depends on the imaging device used for the measurements. Typically, the case is that the

broader the spectrum of the measured wavelengths, the more expensive the equipment. Wavelengths in the visible spectrum can be measured using a consumer-grade camera, while hyperspectral measurements require more advanced and expensive equipment (Fahlgren et al., 2015; Sankaran et al., 2010).

Example application of spectral phenotyping

An example of successful application of spectral phenotyping is a paper published in 2014 by (Ashourloo, Mobasheri, & Huete, 2014), which demonstrates the use of vegetation indices combined with hyperspectral imaging for detecting leaf rust in wheat. Light reflectance was measured within the range of 450 – 1000 nm for different stages of leaf rust infection. The authors designed two indices based on the observed change in reflectance at the three wavelengths: 605, 695 and 455 nm, which could be used to estimate different stages of the disease with an R^2 value as high as 0.94 (Ashourloo et al., 2014).

Spectral analysis software

A multitude of open-source software already exists for analysis of hyperspectral spectroscopic or imaging data. For example, the R package hsdar (Lehnert, Meyer, & Bendix, 2016) offers a library for managing, analyzing, and simulating hyperspectral data, with focus on plant remote sensing data. Hsdar offers functions for handling spectral data: users can load their spectral data, subset it and aggregate it by attribute data. Hsdar also offers basic data analysis functionality, calculation of vegetation indices, machine learning functionality, and more, but requires that the user is familiar with coding in R to use it.

The Python equivalent HyperSpy (de la Peña et al., 2015) likewise provides well-documented tools to analyze both non-imaging and imaging hyperspectral datasets, complete with plotting functions and machine learning functionality. However, just like hsdar requires knowledge of R programming, HyperSpy requires that the user is familiar with programming in Python to use it.

Aim of the project

The aim of this project is to implement an R Package called PlantSpec, which streamlines the handling of large hyperspectral datasets by combining spectral measurements for the user, and which provides a variety of data processing and analysis functions geared towards spectral data. The software should be able to load large spectral data sets with ease and handle or be extended to handle different spectral file formats from the various types of devices and measurement software available to users.

Besides the file reading functionality, PlantSpec should provide data processing functions like the calculation of vegetation indices from loaded spectral data. PlantSpec should also provide data visualization functions in the form of meaningful plots that highlight the difference in spectra between different groups of spectral samples and individual spectral samples in the data set.

This project should make the abovementioned functionality accessible to users with minimal or no prior programming knowledge, by extending the R package into a web application that provided a graphical user interface to the underlying R code.

The functionality of the package and web application will be illustrated by a case study with an analysis of spectral data collected in a field trial of winter wheat subjected to different levels of nitrogen treatment, with the aim of showing how the visual analysis functions can show differences in the data between the treatment groups.

Methods

R package

PlantSpec is implemented in several distinct layers, as shown in Figure 2. At the core of PlantSpec lies an R package that provides a file-reading, processing and visualizing functionality. The PlantSpec R package is implemented in R (3.3.2), which is a programming language developed for statistical analysis. It evolved out of the statistical programming language S around 1993 and has become one if not the most widely used tool in the world of science for statistical analysis. R comes with a wide variety of libraries implementing statistical and graphical tools that are made available through the Comprehensive R Archive Network (CRAN), or projects like Bioconductor (R Core Team, 2016). Implementing the software in R allows for making use of the many open-source libraries that are made available on CRAN. Furthermore, it also allows for distributing the software as an R package in the future, so that users that are familiar with R can download it and run in their statistical analyses.

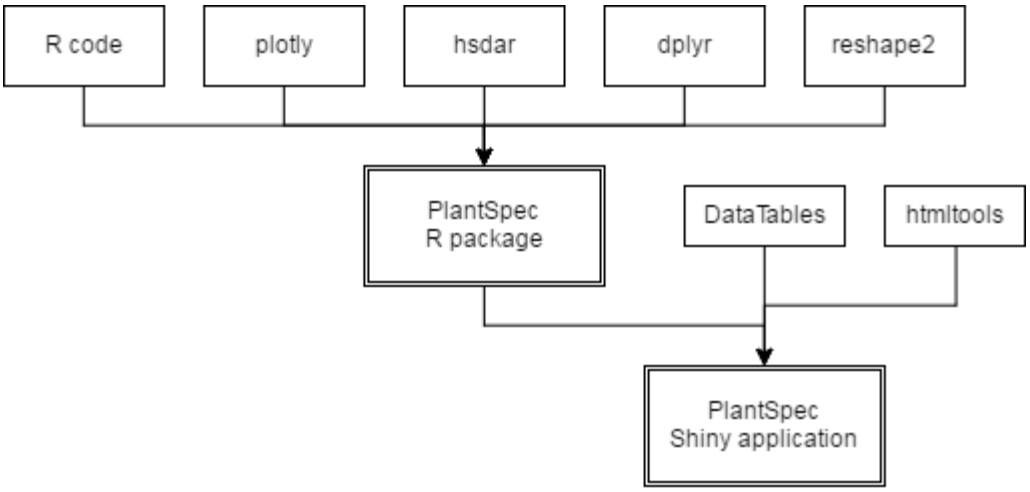


Figure 2. A map of tools and packages used to implement the PlantSpec R package, Shiny application and complete web application.

The R package relies on the R library hsdar (0.5.1) for the Speclib class used to store and manipulate spectral data. PlantSpec primarily makes use of the helper functions for Speclib objects provided by hsdar for manipulating spectral data, but also implements its own thin wrapper function around hsdar's vegindex function, that expands the list of available vegetation indices. PlantSpec uses the 115 indices included with the hsdar package[1] (Lehnert et al., 2016) and adds 21 indices that are described in Table I in the Appendix.

Interactive plots are implemented using the plotly (4.5.6) package, which provides an R API to plotly.js (Sievert et al., 2016). Plotly is an open-source data visualization tool that enables creating interactive java-script based charts and diagrams from R. PlantSpec uses plotly to provide interactive plots instead of static image plots. This means users can hover over data points in charts for more information, zoom in and out of plots, while also being able to save the plots to disk. Furthermore, the PlantSpec R package makes use of the dplyr R package (Wickham & Francois, 2016) for the internal functions, mainly related to subsetting data for the purpose of building plots. PlantSpec also makes extensive use of the melt function from the reshape2 R library (Wickham, 2007), which is used for converting data from wide format, where all spectral bands are organized into their own columns, to long format, where all the bands are collapsed into one column. This is mainly used in some of the implemented PlantSpec plotting functions, where having spectral bands in one column is necessary for plotting reflectance values as a function of wavelength.

[1] For a reference list of indices included in hsdar, please refer to this link: https://cran.r-project.org/web/packages/hsdar/vignettes/References.pdf

Shiny application

The R package described above is extended using the open source R package Shiny, which provides a framework for implementing web applications using R code (Chang, Cheng, Allaire, Xie, & McPherson, 2017). Shiny makes it possible to implement a web application that allows the user to interface with the R code that underlies the handling and processing of the spectral data. The implemented Shiny application makes use of the DT package, which serves as an interface to the javascript DataTables library (Xie, 2016). DataTables is used to display tabular data in web documents, while also providing advanced interactive features such as selecting, filtering or searching for data in the tables. Finally, the Shiny application relies on the htmltools R package, which is used to generate HTML output in the user interface (RStudio & Inc., 2016).

Web server

The data upload is handled by a CGI-script coded in the scripting language Python 3.5, that unpacks the user-uploaded archive, generates a job ID and associated folder structure, and passes the user data to an R script that processes the data before loading in the Shiny app. The Shiny application is served using the free Shiny server software and the website is hosted and coordinated using Apache2.

Case study data

25 varieties of winter wheat were sown in a field trial in Svalöv, Skåne in Southern Sweden in 2015/2016 by Tina Henriksson, Lantmännen Lantbruk. The varieties were randomized into an alpha lattice design, with three nitrogen treatments (140, 180 and 220 kg Ha^{-1}) and two replicates per treatment. The plant material was a combination of breeding lines and elite varieties.

Spectral reflectance measurement

The hyperspectral reflectance of the canopies was measured using the handheld Apogee PS-100 visible to near infrared range spectroradiometer (SpectraWiz PS-100, Apogee, Roseville, CA). The spectroradiometer calibrated against the default white reference, with additional calibration performed every second reading. Measurements were made in the range of 339.0 – 1177.5 nm. Due to a low signal-to-noise ratio in the areas around the edges of the measured spectral interval, the collected spectral data was restricted to the interval of 400 – 1000 nm. The measurements were made in 11 June 2016, during midday, under a clear sky. The spectroradiometer was held approximately one meter above the canopies while collecting the reflectance data.

Data analysis

The spectral data from the field trial was loaded into PlantSpec and analyzed using the available data visualization and processing functions. The mean spectra were compared between the three treatment groups. The spectral data was processed into four vegetation indices, as seen in Table 1, which were compared between the three treatment groups.

Table 1. Vegetation indices used in a case study analysis of wheat reflectance spectra.

Index name	Formula
Normalized Difference Vegetation Index (NDVI)	$\frac{R680 - R800}{R680 + R800}$
Red edge inflection point approximating index (REIP)	$\frac{\frac{(R670 + R780)}{2} - R700}{R740 - R700}$
Water Index (WI)	$\frac{R900}{R700}$
Custom index	$\frac{R400}{R500}$

The nitrogen status analysis was performed using the Normalized Difference Vegetation Index (NDVI, Rouse, Haas, Schell, & Deering, 1973b) and the Red Edge Inflection Point approximating index (REIP). NDVI is an indicator of green biomass in remote sensing data, and sensitive to chlorophyll content in vegetation. REIP is a standard index developed by (Guyot, et al, 1988 as cited by Heege, Reusch, & Thiessen, 2008) and shown to be better correlated with nitrogen status than other standard indices, including NDVI (Heege et al., 2008). The Water Index (WI) is a simple ratio proposed to be sensitive to plant water status (J. Peñuelas et al., 1993). WI is used to compare a trait other than nitrogen status between the three groups. Finally, the final custom index is not known to be indicative of any plant property, and is used to illustrate an index that does not indicate any pattern in the data.

Results

The result is an R package and an implemented web application following the specifications laid out in the methods. The R package contains high-level functions that allow users to analyze their hyperspectral data following the workflow shown in figure 3. The web application provides a high-level interface to the functions specified in the R package.

Data processing

The PlantSpec R package implements a function for reading spectral data, where users can specify the file format and file or directory path. The function will read and combine individual files into a Speclib data object (from the hsdar package) or a matrix with spectral data. Currently, PlantSpec only supports spectral data obtained from SpectraWiz software bundled with Apogee spectroradiometers, and standard long and wide-format data. The package implements a function for calculating vegetation indices from Speclib objects. The function acts as a thin wrapper around hsdar's vegetation index function: it expands hsdar's built-in vegetation index calculation function with an expanded list of built-in indices, of which PlantSpec offers 140.. PlantSpec implements a function that preprocesses and performs a Principal Component Analysis (PCA) on the spectral data. The preprocessing involves removing all spectral bands that show no variance from the spectral data and passing the remaining data to R's built-in prcomp function.

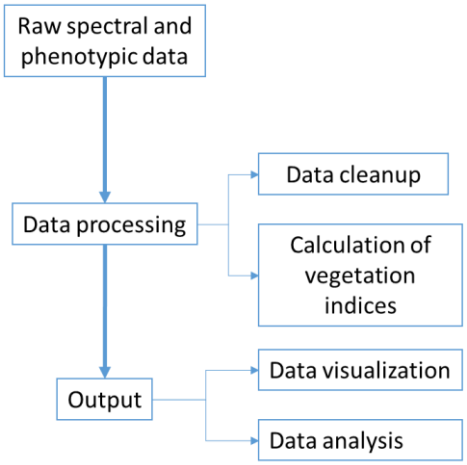


Figure 3. The PlantSpec data processing and visualization pipeline.

Data visualization functions

The package offers several high-level visualization functions for producing interactive plotly plots from the spectral data, with some examples shown in figure 4. Users can visualize the reflectance as a function of wavelength. The reflectance can be plotted by individual samples, or aggregated as a mean, median or variance. Figure 4a shows an example plot of mean reflectance. The aggregated plots can be split by attributes. The package also implements a function for visualizing PCA plots created from the spectral data, as seen in figure 4c. Users can use it to plot the two first principal components against each other, and map the marker aesthetics to attribute data.

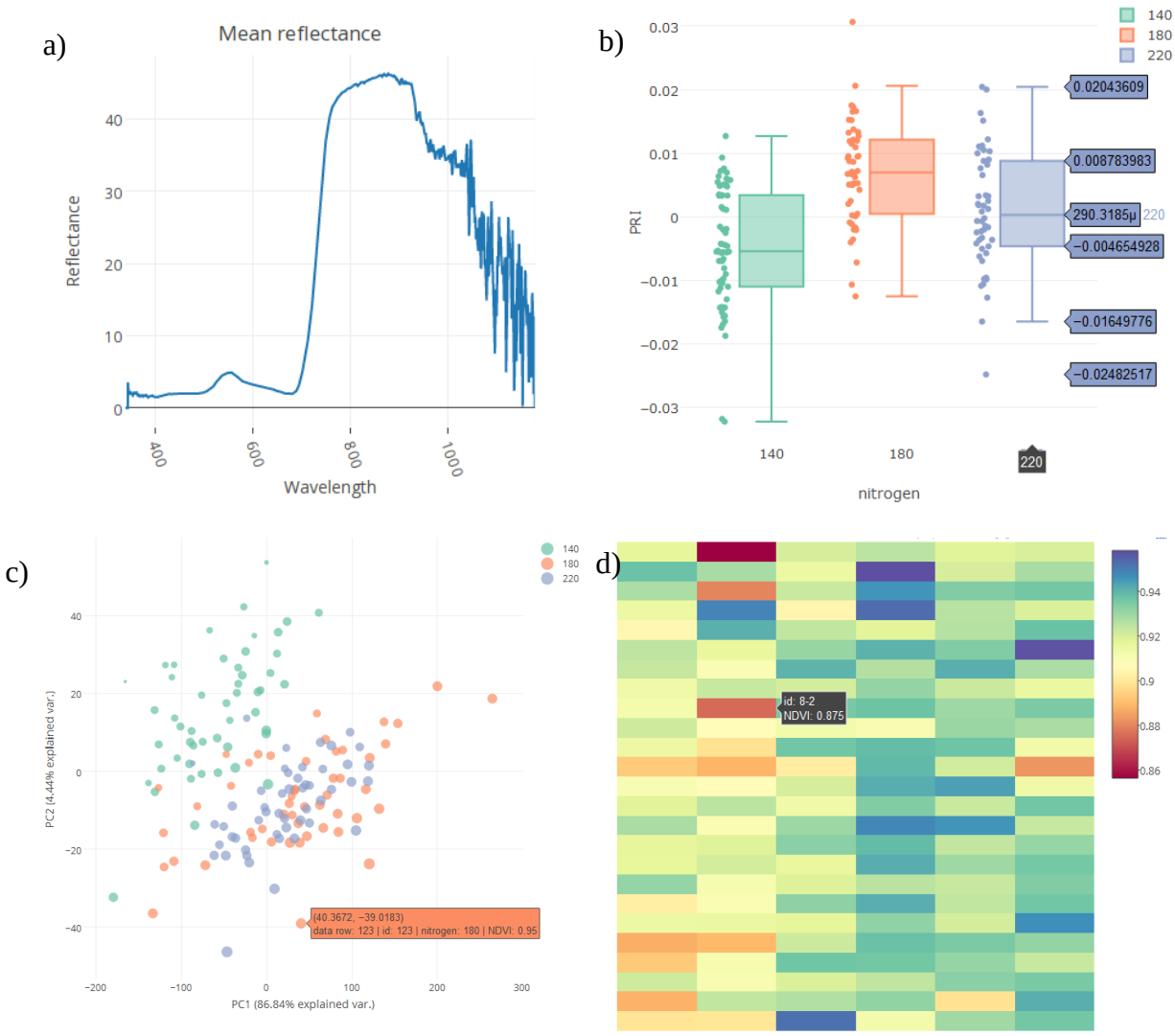


Figure 4. Example plots of spectral data made in PlantSpec, from left to right, top to bottom: a) Mean reflectance spectrum of a winter wheat canopy. b) PRI values between three nitrogen treatment groups, c) PCA plot of spectral data visualized by nitrogen treatment group colorwise and size mapped to NDVI values. d) Distribution of NDVI values in the field where the spectral data was collected.

The PlantSpec package offers several vegetation index visualizations. The first one are basic vegetation index plots, where users can visualize a selected vegetation index against attribute data, or for individual samples. Vegetation indices can be visualized in the form of scatter plots or boxplots, with an example boxplot shown in figure 4b. A basic dot-plot is used when visualizing all calculated index values in a data set. Furthermore, as seen in figure 4d, users can supply a matrix representing the spatial distribution of their samples to a function that produces a spatial "field map" of index values, where each square corresponds to for example a wheat variety or plot.

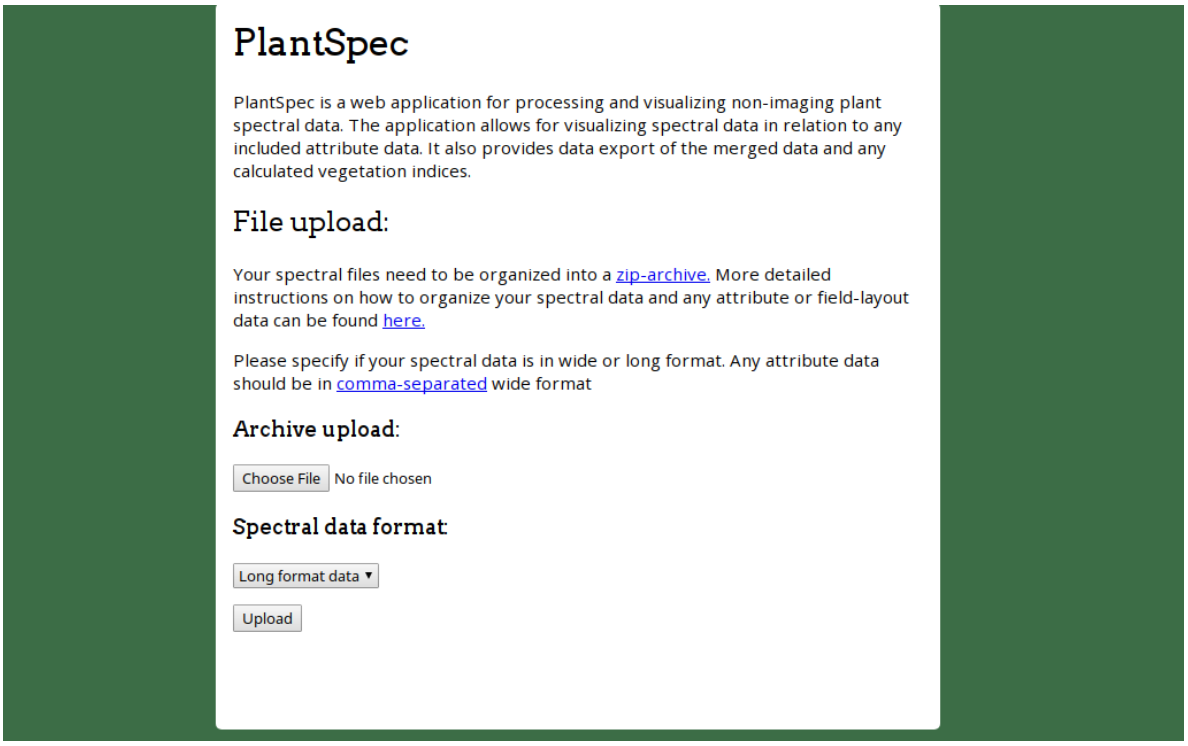


Figure 5. The homepage of PlantSpec. Users select a data-format and upload a zip-archive with the data.

Web application

The web application serves as an interactive high-level interface to the functions outlined above. The Shiny web application that implements the interface to the R code is hosted on a Shiny Server. The users upload their data in the form of a zip archive on the homepage of PlantSpec, seen in figure 5, and select the format of their data. The data is processed by a CGI script coded in Python and R, where the user data is assigned a unique job ID, processed into a serialized Speclib object, and loaded in the Shiny application.

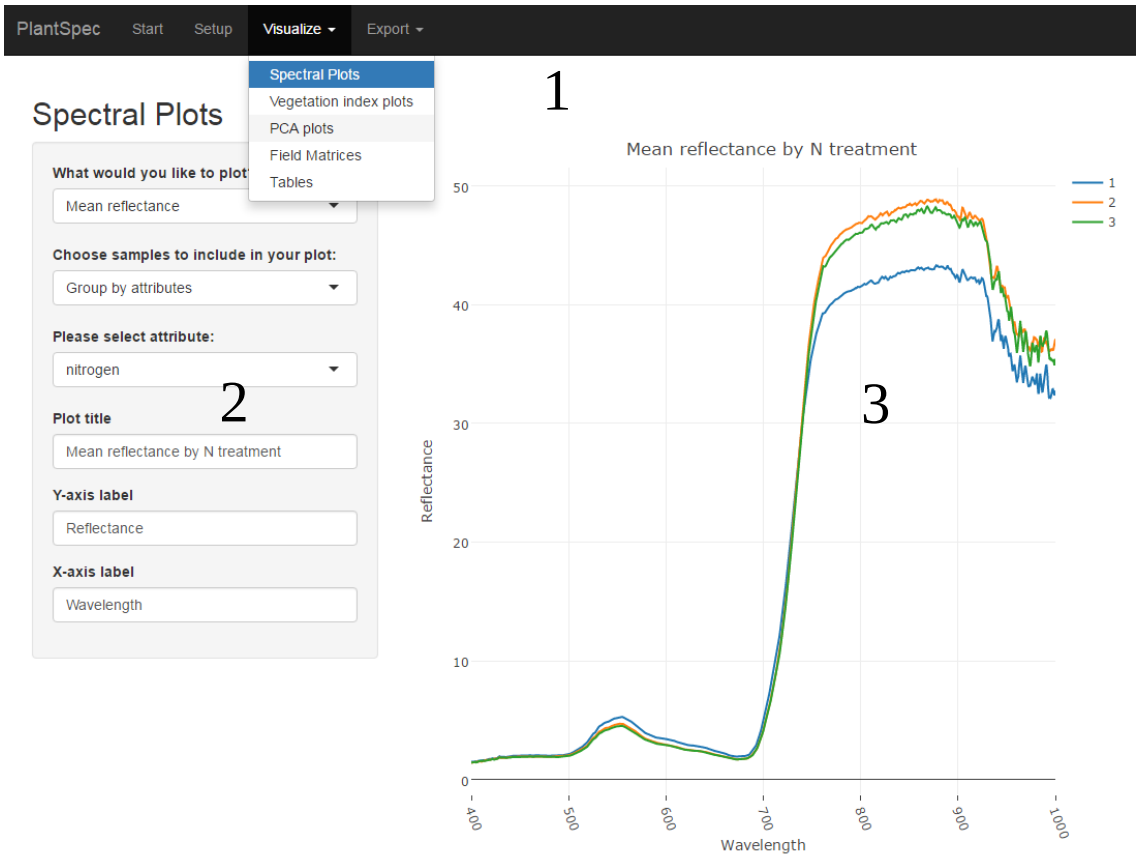


Figure 6. The spectral plot visualization menu in the PlantSpec web application. Users access other sections of the site using the (1) top navigation bar. Current plots and tables can be adjusted using options in the sidebar (2) and plots are visualized in the main view (3)

Figure 6 shows the implementation of PlantSpec as a Shiny application. Users can access the different functions of the application using the navigation bar at the top of the view (1). Most pages implement a sidebar (2) for adjusting data and plot options and a main view (3) where tables and plots appear. The Start menu is where users arrive after successfully uploading their data, and shows a summary of the spectral data and any attributes. The Setup page lets users modify their data into a working data set. The Visualize menu contains links to various visualization options that reflect the implemented plot types in the R package, and the Export menu lets the user export their processed spectral data.

Figure 7. PlantSpec setup page, for modifying uploaded data into a working data set.

The setup page, seen in figure 7, lets users view their raw data, which includes reflectance data and any attribute data. Here, users can modify their data into a working data set that will be used in the subsequent visualizations and calculations of vegetation indices, by removing unwanted samples, or specifying which attributes they are interested in in subsequent analysis. Users can also mask undesired wavelength intervals in their spectral data.

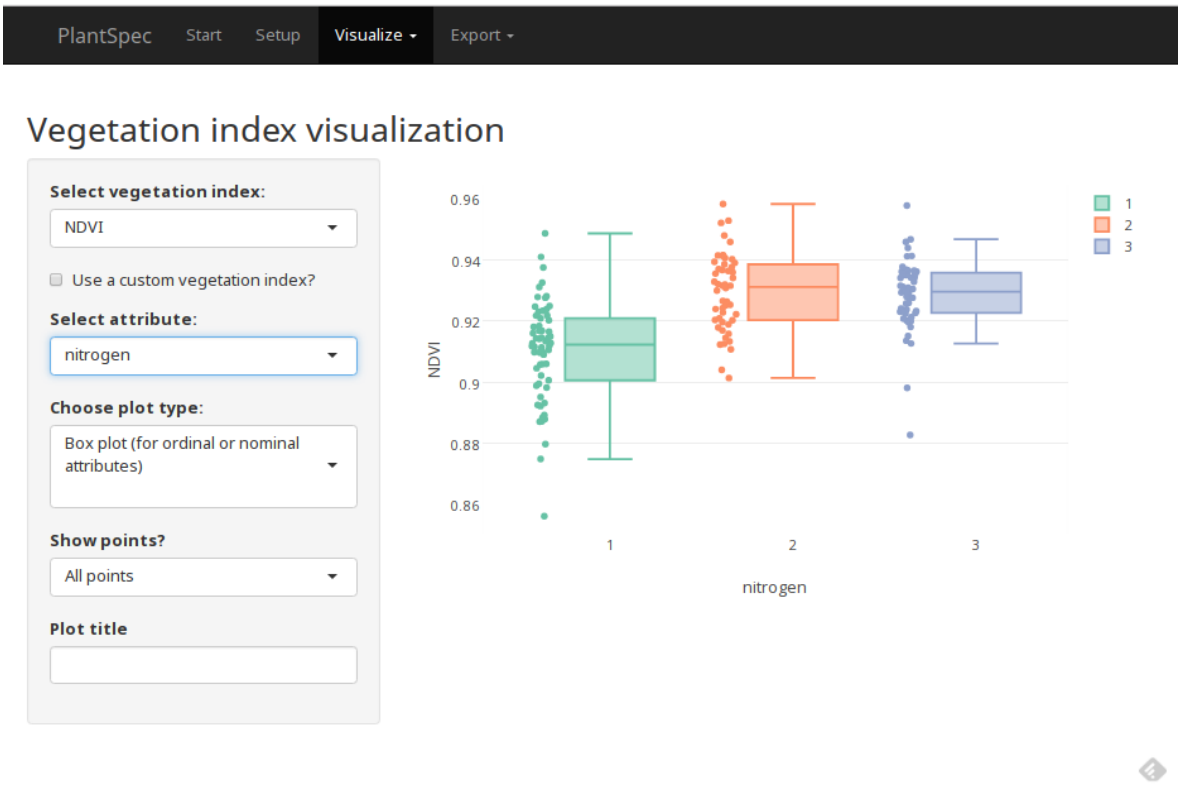


Figure 8. PlantSpec vegetation index visualization menu.

The vegetation index visualization section, seen in figure 8, allows the users to calculate and visualize any of the built-in or custom vegetation indices from their data, and plot them against categorical attributes in a boxplot, or continuous attributes in a scatter plot. The view in figure 8 shows the menu for visualizing box plots. Users can choose a predefined or custom vegetation index they would like to visualize, and the attribute by which the index values should be grouped. Then the user can adjust plot-specific options, like mapping attributes to plot-aesthetics.

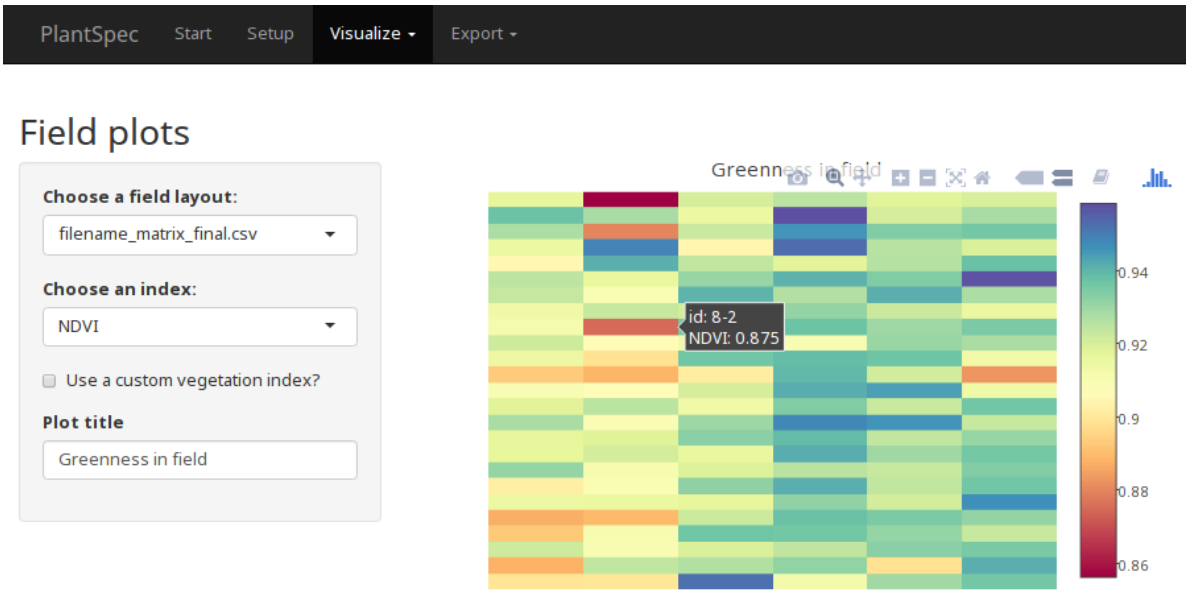
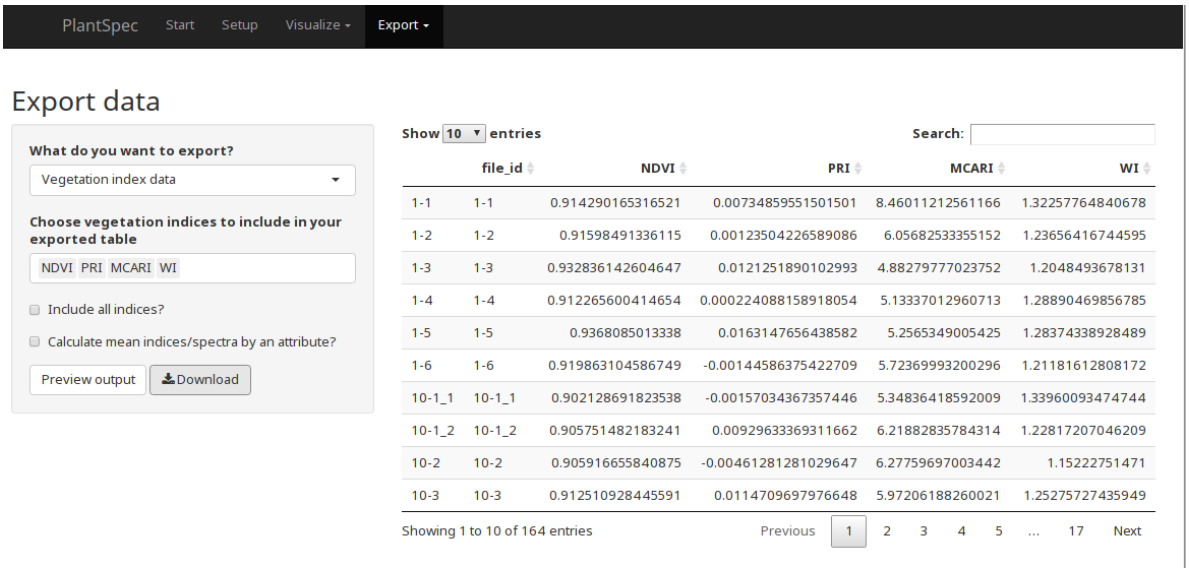


Figure 9. Field vegetation index plot menu in PlantSpec.

The web application implements the option to let users visualize field maps of vegetation index values, as seen in figure 9. Users can map a predefined or custom index to a matrix that represents the distribution of samples in a theoretical field. Users can upload several field matrices, if their uploaded data set consists of data from more than one field, or timepoint.

Figure 10. PlantSpec PCA plot menu

Finally, for web application users interested in clustering the spectral data, the web application implements an interface to creating previously mentioned PCA plots based on the spectral data, as seen in figure 10. Users also have the option to map the size and color of markers in the PCA plot to attributes, allowing the visualization of the clustered spectral data in relation to attribute data.



	file_id	NDVI	PRI	MCARI	WI
1-1	1-1	0.914290165316521	0.00734859551501501	8.46011212561166	1.32257764840678
1-2	1-2	0.91598491336115	0.00123504226589086	6.05682533355152	1.23656416744595
1-3	1-3	0.932836142604647	0.0121251890102993	4.88279777023752	1.2048493678131
1-4	1-4	0.912265600414654	0.000224088158918054	5.13337012960713	1.28890469856785
1-5	1-5	0.9368085013338	0.0163147656438582	5.2565349005425	1.28374338928489
1-6	1-6	0.919863104586749	-0.00144586375422709	5.72369993200296	1.21181612808172
10-1_1	10-1_1	0.902128691823538	-0.00157034367357446	5.34836418592009	1.33960093474744
10-1_2	10-1_2	0.905751482183241	0.00929633369311662	6.21882835784314	1.22817207046209
10-2	10-2	0.905916655840875	-0.00461281281029647	6.27759697003442	1.15222751471
10-3	10-3	0.912510928445591	0.0114709697976648	5.97206188260021	1.25275727435949

Figure 11. PlantSpec export menu.

As seen in figure 11, after the Visualization menu comes the Export menu, where users can re-export their spectral data as a comma separated value file, and build a custom table of selected vegetation indices for each sample, or calculate the mean vegetation index value by a categorical attribute of choice.

Case study

The case study data outlined in the Methods section was loaded into and analyzed in PlantSpec. In figure 12, the mean reflectance was split between the three nitrogen treatment groups. The figure shows that on average the reflectance varies the most between the first treatment group and the other two. The first treatment group has a higher reflectance in the photosynthetic region of 400.0 – 700.0 nm, and a lower reflectance in the NIR region of 700.0 nm and upwards. The two remaining groups seem to vary noticeably in the NIR region, with treatment group two showing a higher reflectance than group three.

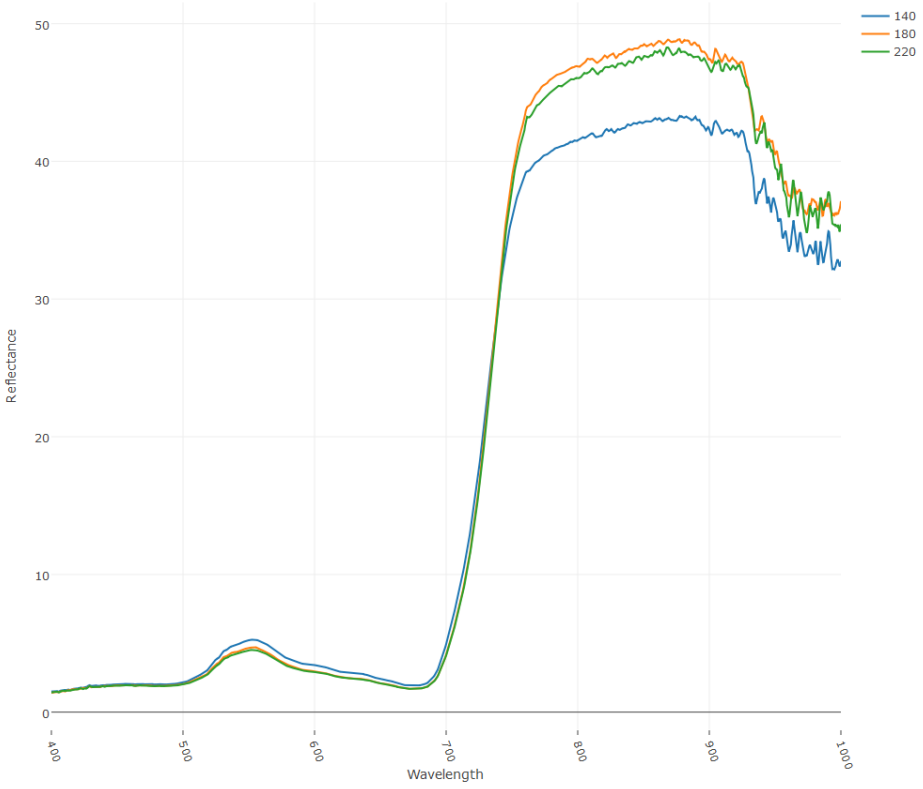


Figure 12. Mean reflectance spectra of winter wheat subject to three different nitrogen treatments (140, 180 and 220 kg Ha^{-1})

The reflectance data was analyzed in a PCA as seen in figure 13. Each point in the figure represents a sample and is colored by the nitrogen treatment group. In the figure it is seen, again, that the samples collected from the first treatment group cluster away from the second and third treatment groups, while treatment groups two and three cluster together.

Figure 13. PCA plot based on spectral reflectance data from winter wheat subject to three different nitrogen treatments (140, 180 and 220 kg Ha^{-1})

PlantSpec was used to compare the three treatment groups relative to the calculated vegetation index values using the indices outlined in table 1. Figure 14 shows the four indices compared using boxplots. The NDVI (figure 14a) and REIP (figure 14b) results both show that the 140 kg Ha^{-1} separates from the remaining two treatment groups with higher applied nitrogen. The REIP index shows higher separation between the 1st treatment group and the other two. The WI (fig 14c) plot shows more variation in the index values of the low nitrogen treatment group compared to the two other groups. The custom index shows roughly the same distribution of values for each group.

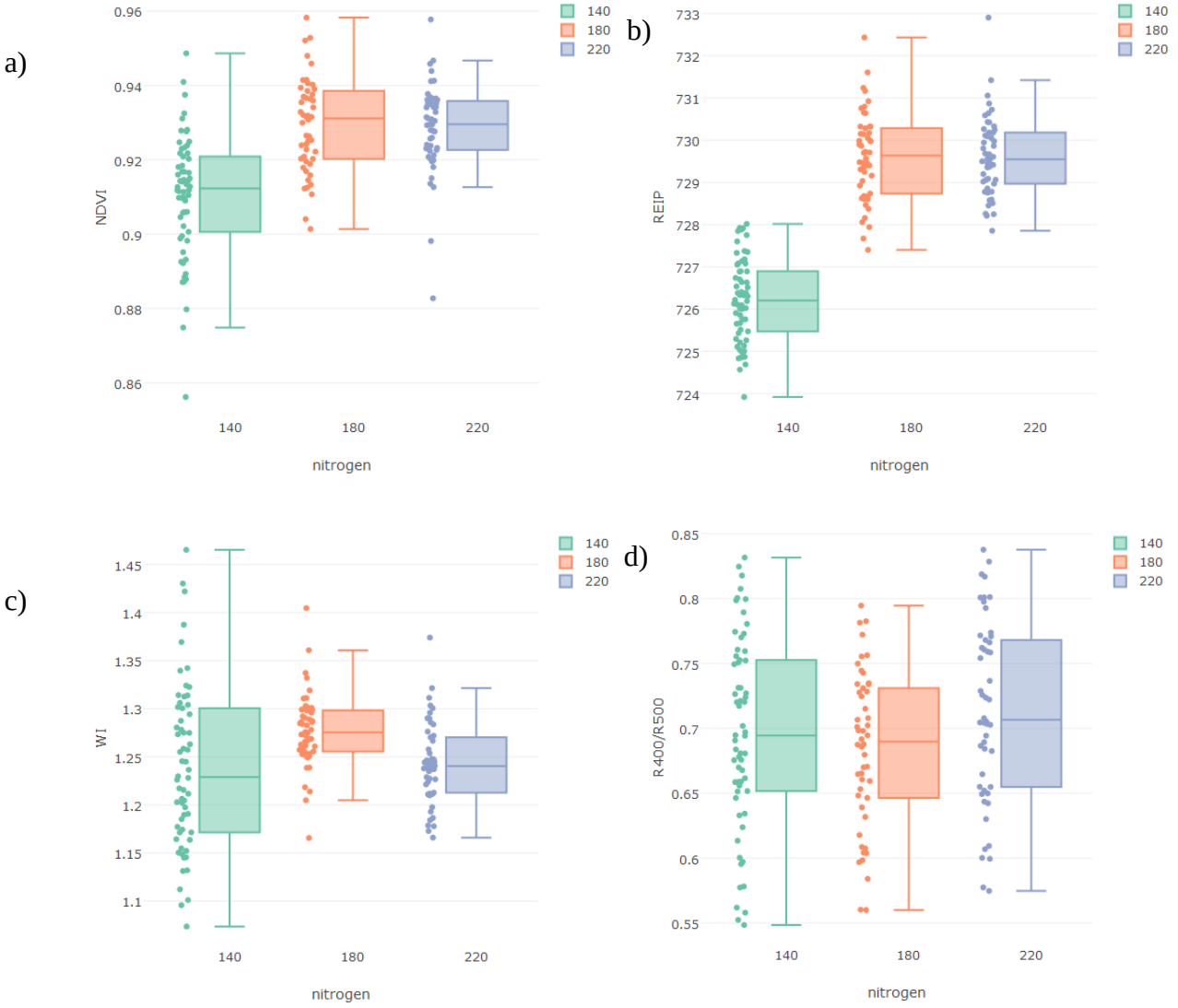


Figure 14. Vegetation index data from winter wheat reflectance spectra subject to three different nitrogen treatments (140, 180 and 220 kg Ha^{-1}): a) NDVI b) REIP c) WI d) Custom index (R400/R500)

In figure 15 the same indices are plotted in a field map, where the position of each value in the matrix corresponds to the position of the wheat plot from which it was collected in the field. The low, medium, and high nitrogen treatment groups follow each other from left to right. Meaning that rows one and two belong to the low nitrogen treatment group, three and four are the medium treatment, and so on. Again, the distribution of NDVI values reveals that the first treatment has lower NDVI values compared to the rows with medium and high nitrogen treatment, but this time it is revealed how individual wheat plots respond to the different treatment groups. The REIP matrix mirrors this pattern but shows a more pronounced difference between the low nitrogen group and the other two. The WI field map suggests lower WI values in the 2nd and 6th rows, while the custom index does not reveal any pattern.

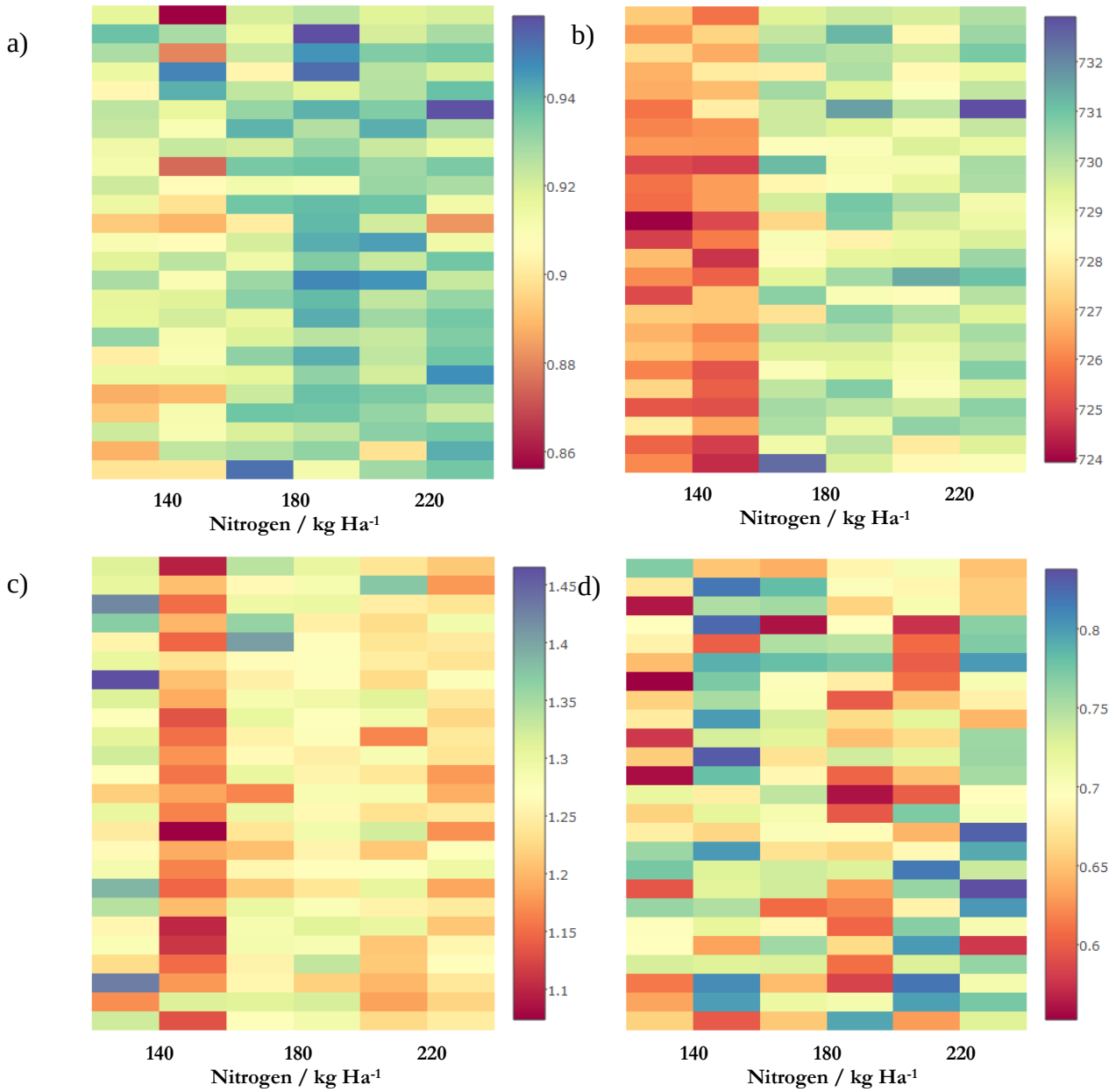


Figure 15. Vegetation index data from winter wheat reflectance spectra subject to three different nitrogen treatments (140, 180 and 220 kg Ha^{-1}). Each square represents a different wheat line. a) NDVI b) REIP c) WI d) Custom index (R400/R500)

Discussion

Spectral HTPP methods generate substantial amounts of data, creating a challenge in merging the data together with any phenotypic data into a workable format. This problem can be resolved by user-friendly software that simplifies the task of handling these large data sets and making the information contained within available for processing and analysis. High interactivity and meaningful analysis functionality are necessary in such software. PlantSpec accomplishes the aim of providing an R package extended into a web application that meets the criteria above. High-level functions allow users to visualize patterns in their spectral data by either looking at

the reflectance profiles or calculated vegetation indices. Users can combine spectral data from files or directories, and extract vegetation indices from their data — both using the R package and the web application. The graphical interface provided by the web application allows users with no prior knowledge of programming in R to gain access to the data visualization and processing functions.

The usefulness of the data visualization and processing functions are shown in the figures produced in the case study. Figure 12 shows how the mean reflectance can be compared between spectral samples partitioned into different treatment groups. The option to plot PCA results and map the plot aesthetics to recorded attributes likewise allows for visualizing trends in the spectral data, but on a finer level compared to the aggregated reflectance. PCA plots also come with the added benefit of visualizing potential outliers. Returning to the case study results, the PCA plot in figure 13 uses data obtained from a PCA of the raw spectral data to illustrate if and how the data clusters in relation to the three fertilizer levels. Users can do the same analysis in relation to any categorical or continuous attributes, with the added benefit of visualizing outliers. In the case of the case study, the two figures reveal that the reflectance of the canopies that received the 140 kg Ha^{-1} treatment differ from the two other groups (180 and 220 kg Ha^{-1}).

PlantSpec allows users to analyze their data using over 100 vegetation indices. The case study results show how spectra can be processed into vegetation index values and compared in relation to attribute data. In the case study, the fertilizer groups were compared using two indices known to correlate with plant nitrogen status (Hansen & Schjoerring, 2003; Li et al., 2010). Vegetation indices that have been identified as correlated with, for example, nitrogen status, as is the case with NDVI and REIP, or traits like yield or water status, can be used to try to predict those attributes in all samples or between treatment groups. Figures like the figures 14 and 15 show examples of visualization of vegetation indices in PlantSpec, and can be useful for identifying vegetation indices that are correlated to differences in plant trait data. This is useful for variable selection, in the sense of identifying indices to explore when building statistical phenotyping models. Figure 14 helps discern the distribution of index values between treatment groups, allowing the user to identify indices that are correlated to selected attribute data. Figure 15 shows the spatial distribution of the same values in a representation of the field that the data was collected in. In the context of the case study, figure 15 is useful for identifying how the selected vegetation index value differs between varieties in the three different groups.

The implemented Shiny web application fulfils the goal of enabling users to access PlantSpec without having to program. The code is made interactive via simple navigation and sidebar menus which helps in keeping things simple for the end-user. As seen in figure 11, the web-application also allows users to easily extract their combined or processed data as a csv-file, for use outside of PlantSpec.

As stated previously, hyperspectral data provides a promising way of detecting diseases and other forms of plant stress in vegetation (Sankaran et al., 2010). Hyperspectral data can be processed into hyperspectral indices and related to diseases, but also crop traits like yield or leaf cover (Thenkabail, Smith, & De Pauw, 2000). PlantSpec offers an open-source tool for exploratory data analysis and visualization of high-throughput hyperspectral data. It offers data visualization functions that allow the end-user to visualize how the spectral data behaves in relation to any traits of interest and to identify spectral bands or vegetation indices that can be investigated in statistical models of the same traits. Most importantly, PlantSpec lowers the entry barrier to hyperspectral data analysis, by packaging the software in the form of a fully functional online tool, that abstracts all required programming into an easy-to-use graphical user interface.

Future work

Although PlantSpec in its current state is useful for basic exploratory analysis of hyperspectral data, the application has its limitations. PlantSpec currently only supports three filetypes (SpectraWiz files and tab separated long and wide-form data). However, it can be easily extended to support more filetypes, and work is underway to add support for ASD FieldSpec binary data files, by incorporating the asdreader R package (Roudier, 2016) into PlantSpec.

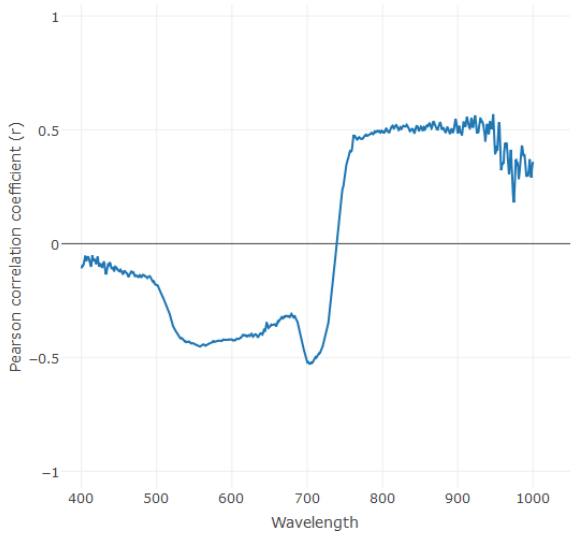


Figure 16. Correlation between measured spectral reflectance at individual bands and three levels (140, 180 and 220 kg Ha^{-1}) of nitrogen treatment in a field trial of winter wheat.

PlantSpec currently provides no high-level functions for visualizing correlation between numeric attributes and individual spectral bands. This sort of relationship is usually visualized in a correlogram, as seen in figure 16, and is seen in use in a wide array of papers trying to identify spectral bands for characterizing crop traits, for example in (Thenkabail et al., 2000). This is also currently in the works and will be included in a future version.

The web-application does not offer the possibility of aggregating boxplots and spectra by more than one variable. This makes it impossible to study something like the mean spectra of three different treatments over a second attribute, like for example a temporal variable. Adding this functionality would make it easier to study spectral data by attributes over time.

Another crucial feature missing from PlantSpec is functionality for handling replicates. In cases where users want to merge biological or technical replicates into their respective mean values, they cannot yet merge replicates in the web application, but must do it beforehand. This goes against the goal of minimal data pre-processing required to use the application.

Aside from the concerns outlines above, the R package and web application code requires more testing and more robust error handling. Plots like the PCA plot, although useful, could be updated to support mapping more aesthetics to the attributes, like marker type, and easily mapping any calculated index values to plot aesthetics, to uncover more patterns in the data.

Conclusion

In conclusion, PlantSpec offers a toolset in the form of an R package or online tool to easily load, process and analyze plant spectral data in relation to any traits. More work needs to be done, and a few crucial functions need to be added to the package and online tool, but until then the PlantSpec R package and web application are both capable of basic processing and exploration of hyperspectral data in relation to non-spectral attribute data. Users can visualize difference in reflectance in relation to non-spectral attribute data and spatial data, and process their spectral data using over 100 vegetation indices. Most crucially, the web application makes those functions accessible to users with minimal programming skills, making exploration of hyperspectral phenotyping more streamlined and accessible to the end user.

References

Agapiou, A., Hadjimitsis, D. G., & Alexakis, D. D. (2012). Evaluation of broadband and narrowband vegetation indices for the identification of archaeological crop marks. *Remote Sensing*, *4*(12), 3892–3919. https://doi.org/10.3390/rs4123892

Ashourloo, D., Mobasheri, M. R., & Huete, A. (2014). Developing two spectral disease indices for detection of wheat leaf rust (Pucciniatriticina). *Remote Sensing*, *6*(6), 4723–4740. https://doi.org/10.3390/rs6064723

Chang, W., Cheng, J., Allaire, J. J., Xie, Y., & McPherson, J. (2017). shiny: Web Application Framework for R. Retrieved from https://cran.r-project.org/package=shiny

de la Peña, F., Burdet, P., Sarahan, M., Nord, M., Ostasevicius, T., Taillon, J., … Walls, M. (2015). Hyperspy 0.8. *Doi.org*. https://doi.org/10.5281/zenodo.16850

Diacono, M., Rubino, P., & Montemurro, F. (2013). Precision nitrogen management of wheat. A review. *Agronomy for Sustainable Development*, *33*(1), 219–241. https://doi.org/10.1007/s13593-012-0111-z

Fahlgren, N., Gehan, M. A., & Baxter, I. (2015). Lights, camera, action: High-throughput plant phenotyping is ready for a close-up. *Current Opinion in Plant Biology*, *24*, 93–99. https://doi.org/10.1016/j.pbi.2015.02.006

Furbank, R. T., & Tester, M. (2011). Phenomics - technologies to relieve the phenotyping bottleneck. *Trends in Plant Science*, *16*(12), 635–644. https://doi.org/10.1016/j.tplants.2011.09.005

Gamon, J. A., Serrano, L., & Surfus, J. S. (1997). The photochemical reflectance index: An optical indicator of photosynthetic radiation use efficiency across species, functional types, and nutrient levels. *Oecologia*, *112*(4), 492–501. https://doi.org/10.1007/s004420050337

Guyot, G., Baret, F., & Major, D. J. (1988). High spectral resolution: determination of spectral shifts between the red and the near infrared. *International Archives of Photogrammetry and Remote Sensing*, *11*(750–760).

Hansen, P. M. M., & Schjoerring, J. K. K. (2003). Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. *Remote Sensing of Environment*, *86*(4), 542–553. https://doi.org/10.1016/S0034-4257(03)00131-7

Heege, H. J., Reusch, S., & Thiessen, E. (2008). Prospects and results for optical systems for site-specific on-the-go control of nitrogen-top-dressing in Germany. *Precision Agriculture*, *9*(3), 115–131. https://doi.org/10.1007/s11119-008-9055-3

Khosla, R., Fleming, K., Delgado, J. a, Shaver, T. M., & Westfall, D. G. (2002). Use of site-specific management zones to improve nitrogen management for precision agriculture. *Journal of Soil and Water Conservation*, *57*(6), 513–518.

Knipling, E. B. (1970). Physical and Physiological Basis for the Reflectance of Visible and Near Infrared Radiation from Vegetation. *Remote Sensing of Environment*, *1*, 155–159. Retrieved from https://pubag.nal.usda.gov/pubag/downloadPDF.xhtml?id=37948&content=PDF

Lehnert, L. W., Meyer, H., & Bendix, J. (2016). hsdar: Manage, Analyse and Simulate Hyperspectral Data. Retrieved from https://cran.r-project.org/web/packages/hsdar/index.html

Li, F., Miao, Y., Hennig, S. D., Gnyp, M. L., Chen, X., Jia, L., & Bareth, G. (2010). Evaluating hyperspectral vegetation indices for estimating nitrogen concentration of winter wheat at different growth stages. *Precision Agriculture*, *11*(4), 335–357. https://doi.org/10.1007/s11119-010-9165-6

Merzlyak, M. N., Gitelson, A. A., Chivkunova, O. B., Solovchenko, A. E., & Pogosyan, S. I. (2003). Application of Reflectance Spectroscopy for Analysis of Higher Plant Pigments. *Text*, *50*(5), 704–710. https://doi.org/10.1023/a:1025608728405

Peñuelas, J., Filella, I., Biel, C., Serrano, L., & Savé, R. (1993). The reflectance at the 950–970 nm region as an indicator of plant water status. *International Journal of Remote Sensing*. https://doi.org/10.1080/01431169308954010

Peñuelas, J., & Filella, L. (1998). Visible and near-infrared reflectance techniques for diagnosing plant physiological status. *Trends in Plant Science*, *3*(4), 151–156. https://doi.org/10.1016/S1360-1385(98)01213-8

Pierce, F. J., & Nowak, P. (1999). Aspects of Precision Agriculture. *Advances in Agronomy*, *67*(C), 1–85. https://doi.org/10.1016/S0065-2113(08)60513-1

R Core Team. (2016). R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*, *version 3*, 3503. https://doi.org/10.1007/978-3-540-74686-7

Resea, A., Molero, G., Group, C., Agricultur, I., & Nogu, S. (2014). NDVI as a Potential Tool for Predicting Biomass , Plant Nitrogen Content and Growth in Wheat Genotypes Subjected to Different ..., (March 2011). https://doi.org/10.1556/CRC.39.2011.1.15

Roudier, P. (2016). asdreader: Reading ASD Binary Files in R. Retrieved from https://cran.r-project.org/package=asdreader

Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1973). Monitoring Vegetation

Systems in the Great Plains with ERTS. *Third Earth Resources Technology Satellite-1 Symposium*, *1*, 325–333.

RStudio, & Inc. (2016). htmltools: Tools for HTML. Retrieved from https://cran.r-project.org/package=htmltools

Sankaran, S., Mishra, A., Ehsani, R., & Davis, C. (2010). A review of advanced techniques for detecting plant diseases. *Computers and Electronics in Agriculture*, *72*(1), 1–13. https://doi.org/10.1016/j.compag.2010.02.007

Sievert, C., Parmer, C., Hocking, T., Chamberlain, S., Ram, K., Corvellec, M., & Despouy, P. (2016). plotly: Create Interactive Web Graphics via "plotly.js." Retrieved from https://cran.r-project.org/package=plotly

Thenkabail, P. S., Smith, R. B., & De Pauw, E. (2000). Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. *Remote Sensing of Environment*, *71*(2), 158–182. https://doi.org/10.1016/S0034-4257(99)00067-X

Wickham, H. (2007). Reshaping Data with the reshape package. *Journal of Statistical Software*, *21*(12), 1–20. https://doi.org/10.1016/S0142-1123(99)00007-9

Wickham, H., & Francois, R. (2016). dplyr: A Grammar of Data Manipulation. Retrieved from https://cran.r-project.org/package=dplyr

Wojtowicz, M., Wojtowics, A., & Piekarczyk, J. (2015). Application of remote sensing methods in Agriculture. *Communication in Biometry and Crop Science*, *11*(1), 31–50.

Xie, Y. (2016). DT: A Wrapper of the JavaScript Library "DataTables." Retrieved from https://cran.r-project.org/package=DT

Appendix

Vegetation indices added to PlantSpec

Table I. List of additional vegetation indices included in PlantSpec

Name/Code	Equation	Reference
ARI	(1/R550)/(1/R700)	(Anatoly A. Gitelson, Mark N. Merzlyak, & Chivkunova, 2001)
ARI2	R800*(1/R550)-(1/R700)	(Anatoly A. Gitelson et al., 2001)
BGI	R450/R550	(Zarco-Tejada et al., 2005)
BRI	R450/R690	(Zarco-Tejada et al., 2005)
CRI	(1/R510)-(1/R550)	(Gitelson, Zur, Chivkunova, & Merzlyak, 2002)
LRDSI1	4.2*(R695/R455)-0.38	(Ashourloo, Mobasheri, & Huete, 2014)
LRDSI2	(R750-R705)/(R750+705-2*R445)	(Ashourloo et al., 2014)
MCARI1	1.2*(2.5*(R790- R670) - 1.3 * (R790- R550))	(Haboudane, Miller, Tremblay, Zarco-Tejada, & Dextraze, 2002)
mNDVI2	(R800/(R670 - 1))/(R800/(R670 + 1))^0.5	(Sims & Gamon, 2002)
mSR3	R723/R700	(Sims & Gamon, 2002)
NPQI	(R415-R435)/(R415+R435)	(Barnes, Balaguer, Manrique, Elvira, & Davison, 1992)

PhRI	(R550-R531)/(R550+R531)	(Gamon, Peñuelas, & Field, 1992)
PRI2	(R570-R539)/(R570+R539)	(Filella, Amaro, Araus, & Peñuelas, 1996)
RGI	R690/R550	(Zarco-Tejada et al., 2005)
SIPI2	(R800 - R440)/(R800 - R680)	(Pen¯Uelas, Filella, Lloret, Mun¯Oz, & Vilajeliu, 1995)
SIPI3	(R800 - R445)/(R800 - R650)	(Penuelas, Baret, & Filella, 1995)
SR9	R690/R655	(Zarco-Tejada, Pushnik, Dobrowski, & Ustin, 2003)
SR10	R685/R655	(Zarco-Tejada et al., 2003)
SR705	R750/R705	(Castro-Esau, Sánchez-Azofeifa, & Rivard, 2006)
VS	R725/R702	(White, Williams, & Barr, 2008)
WI	R900/R970	(Peñuelas, Filella, Biel, Serrano, & Savé, 1993)

Index reference list

Anatoly A. Gitelson, Mark N. Merzlyak, & Chivkunova, O. B. (2001). Optical Properties and Nondestructive Estimation of Anthocyanin Content in Plant Leaves. *Photochemistry and Photobiology*, *74*(1), 38–45. https://doi.org/10.1562/0031-8655(2001)

Ashourloo, D., Mobasheri, M. R., & Huete, A. (2014). Developing two spectral disease indices for detection of wheat leaf rust (Pucciniatriticina). *Remote Sensing*, *6*(6), 4723–4740. https://doi.org/10.3390/rs6064723

Barnes, J. D., Balaguer, L., Manrique, E., Elvira, S., & Davison, A. W. (1992). A reappraisal of the use of DMSO for the extraction and determination of chlorophylls a and b in lichens and higher plants. *Environmental and Experimental Botany*, *32*(2), 85–100. https://doi.org/10.1016/0098-8472(92)90034-Y

Castro-Esau, K. L., Sánchez-Azofeifa, G. A., & Rivard, B. (2006). Comparison of spectral indices obtained using multiple spectroradiometers. *Remote Sensing of Environment*,

103(3), 276–288. https://doi.org/10.1016/j.rse.2005.01.019

Filella, I., Amaro, T., Araus, J. L., & Peñuelas, J. (1996). Relationship between photosynthetic radiation-use efficiency of barley canopies and the photochemical reflectance index (PRI). *Physiologia Plantarum*, *96*(2), 211–216. https://doi.org/10.1111/j.1399-3054.1996.tb00204.x

Gamon, J. A., Peñuelas, J., & Field, C. B. (1992). A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. *Remote Sensing of Environment*. https://doi.org/10.1016/0034-4257(92)90059-S

Gitelson, A. a, Zur, Y., Chivkunova, O. B., & Merzlyak, M. N. (2002). Assessing carotenoid content in plant leaves with reflectance spectroscopy. *Photochemistry and Photobiology*, *75*(3), 272–281. https://doi.org/10.1562/0031-8655(2002)0750272ACCIPL2.0.CO2

Haboudane, D., Miller, J. R., Tremblay, N., Zarco-Tejada, P. J., & Dextraze, L. (2002). Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sensing of Environment*, *81*(2–3), 416–426. https://doi.org/10.1016/S0034-4257(02)00018-4

Pen¯Uelas, J., Filella, I., Lloret, P., Mun¯Oz, F., & Vilajeliu, M. (1995). Reflectance assessment of mite effects on apple trees. *International Journal of Remote Sensing*, *16*(14), 2727–2733. https://doi.org/10.1080/01431169508954588

Penuelas, J., Baret, F., & Filella, I. (1995). Semi-empirical indices to assess carotenoids/chlorophyll a ratio from leaf spectral reflectance. *Photosynthetica*.

Peñuelas, J., Filella, I., Biel, C., Serrano, L., & Savé, R. (1993). The reflectance at the 950–970 nm region as an indicator of plant water status. *International Journal of Remote Sensing*. https://doi.org/10.1080/01431169308954010

Sims, D. A., & Gamon, J. A. (2002). Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sensing of Environment*, *81*(2–3), 337–354. https://doi.org/10.1016/S0034-4257(02)00010-X

White, D., Williams, M., & Barr, S. (2008). Detecting sub-surface soil disturbance using hyperspectral first derivative band rations of associated vegetation stress. ... . *Photogramm. Remote Sens. Spat. Inf. Sci*, *1980*(Figure 1), 243–248.

Zarco-Tejada, P. J., Berjón, A., López-Lozano, R., Miller, J. R., Martín, P., Cachorro, V., … De Frutos, A. (2005). Assessing vineyard condition with hyperspectral indices: Leaf and canopy reflectance simulation in a row-structured discontinuous canopy. *Remote Sensing of Environment*, *99*(3), 271–287. https://doi.org/10.1016/j.rse.2005.09.002

Zarco-Tejada, P. J., Pushnik, J. C., Dobrowski, S., & Ustin, S. L. (2003). Steady-state chlorophyll a fluorescence detection from canopy derivative reflectance and double-peak red-edge effects. *Remote Sensing of Environment*, *84*(2), 283–294. https://doi.org/10.1016/S0034-4257(02)00113-X