



MSU Graduate Theses

Fall 2018

Modeling the Response of Black Walnut -dominant Mixed Hardwoods to Seasonal Climate Effects with UAV-based Hyperspectral Sensor and Aerial Photogrammetry

Tyler G. Bradford

Missouri State University, tyler.bradford95@gmail.com

As with any intellectual project, the content and views expressed in this thesis may be considered objectionable by some readers. However, this student-scholar's work has been judged to have academic value by the student's thesis committee members trained in the discipline. The content and views expressed in this thesis are those of the student-scholar and are not endorsed by Missouri State University, its Graduate College, or its employees.

Follow this and additional works at: <https://bearworks.missouristate.edu/theses>



Part of the [Agricultural Science Commons](#), [Agriculture Commons](#), [Environmental Monitoring Commons](#), [Forest Management Commons](#), and the [Natural Resources and Conservation Commons](#)

Recommended Citation

Bradford, Tyler G., "Modeling the Response of Black Walnut -dominant Mixed Hardwoods to Seasonal Climate Effects with UAV-based Hyperspectral Sensor and Aerial Photogrammetry" (2018). *MSU Graduate Theses*. 3321.

<https://bearworks.missouristate.edu/theses/3321>

This article or document was made available through BearWorks, the institutional repository of Missouri State University. The work contained in it may be protected by copyright and require permission of the copyright holder for reuse or redistribution.

For more information, please contact BearWorks@library.missouristate.edu.

**MODELING THE RESPONSE OF BLACK WALNUT -DOMINANT MIXED
HARDWOODS TO SEASONAL CLIMATE EFFECTS WITH UAV-BASED
HYPERSPPECTRAL SENSOR AND AERIAL PHOTOGRAMMETRY**

A Master's Thesis

Presented to

The Graduate College of

Missouri State University

In Partial Fulfillment

Of the Requirements for the Degree

Master of Natural and Applied Science

By

Tyler G. Bradford

December 2018

© 2018 by Tyler Gene Bradford

MODELING THE RESPONSE OF BLACK WALNUT -DOMINANT MIXED HARDWOODS TO SEASONAL CLIMATE EFFECTS WITH UAV-BASED HYPERSPECTRAL SENSOR AND AERIAL PHOTOGRAMMETRY

Environmental Plant Science and Natural Resources; Geography, Geology, and Planning

Missouri State University, December 2018

Master of Natural and Applied Science

Tyler G. Bradford

ABSTRACT

The development of compact sensors in recent years has inspired the use of UAS-based hyperspectral and aerial imaging techniques for small-scale remote sensing applications. With increasing concerns about climate change, spectrally-derived vegetation indices (VIs) have proven useful for quantifying stress-induced vegetation response. The goal of this study was to develop predictive models and assess methodology for modeling the biological response of a black walnut -dominant mixed hardwood stand to seasonal climate events using UAV-based hyperspectral remote-sensing. The derived VIs were evaluated against the means of four seasonal measures of climate calculated for a two-week period prior to the flight date. A best subsets regression was used to create best fitting linear regression models according to Bayesian Information Criterion (BIC). The highest-ranked model for total precipitation had an AdjR² of 0.0839 and RMSE of 0.0827 inches. The highest-ranked model for maximum air temperature had an AdjR² of 0.9922 and RMSE of 0.5485 °F. The highest-ranked model for average air temperature had an AdjR² of 0.9987 and RMSE of 0.2256 °F. The highest-ranked model for total solar radiation had an AdjR² of 0.9961 and RMSE of 0.06405 MJ/M². The results indicate that select VIs measured at the canopy level may be useful in estimating the response to at least some measures seasonal climate. The proposed regression models could help local researchers and landowners in making short-term management decisions, as well as further our understanding of climate-induced tree stress for maintaining sustainable forests in Missouri.

KEYWORDS: tree health, forestry, UAV, UAS, black walnut, land management, seasonal effects, plant stress, hyperspectral, photogrammetry, vegetation indices, remote sensing

**MODELING THE RESPONSE OF BLACK WALNUT -DOMINANT MIXED
HARDWOODS TO SEASONAL CLIMATE EFFECTS WITH UAV-BASED
HYPERSPSPECTRAL SENSOR AND AERIAL PHOTOGRAMMETRY**

By

Tyler G. Bradford

A Master's Thesis
Submitted to the Graduate College
Of Missouri State University
In Partial Fulfillment of the Requirements
For the Degree of Master of Natural and Applied Science, Agriculture and Geospatial Sciences

December 2018

Approved:

Michael Goerndt, Ph.D., Thesis Committee Chair

William McClain, Ph.D., Committee Member

Xin Miao, Ph.D., Committee Member

Julie Masterson, Ph.D., Dean of the Graduate College

In the interest of academic freedom and the principle of free speech, approval of this thesis indicates the format is acceptable and meets the academic criteria for the discipline as determined by the faculty that constitute the thesis committee. The content and views expressed in this thesis are those of the student-scholar and are not endorsed by Missouri State University, its Graduate College, or its employees.

ACKNOWLEDGEMENTS

Firstly, I would like to acknowledge my academic advisor and mentor, Dr. Michael Goerndt. Your commitment to ensuring my professional success has not gone unnoticed. The intellectual support and personal encouragement you have provided during my studies is unmatched. You have not only been a remarkable teacher and role model, but honored friend as well. Continue your noble quest of “intelligent tinkering,” my friend. I wish you all the best in life, and not just because you deserve it.

Secondly, I would like to acknowledge all of the students, friends, and family who contributed in some fashion to the consummation of this project. Whether it be helping with field measurements in the sweltering summer heat, offering constructive advice in the writing of my thesis, or providing vital encouragement, your involvements are greatly appreciated.

Lastly, I would like to recognize the Missouri Transect of the Experimental Program to Stimulate Competitive Research (EPSCoR)—funded by the National Science Foundation—for financing this research project and my graduate studies. Without whom, this small step of scientific exploration would not have been possible.

I dedicate this thesis to my two best and most loyal friends, Mr. Robert Frost and Mr. Michael Clutter. Bob, thank you for inspiring me with your unwavering character and for proving that a man’s accomplishments are only limited by his will to succeed. Mike, thank you for teaching me to always be kind despite what life throws at you and for always sporting a listening ear. Thank you both for stoking the fire of my ambitions and for continuing to offer loyal friendship. Our bond of brotherhood is invaluable to me. A-team for life!

TABLE OF CONTENTS

Introduction	Page 1
Literature Review	Page 1
Research Objectives	Page 3
Study Site	Page 3
Equipment	Page 5
Methods	Page 7
UAS Data Collection	Page 7
Development of an Orthomosaic	Page 7
Image Rectification and Sampling Design	Page 8
Multiple Linear Regression for Climate Metrics	Page 11
Results	Page 19
Criterion-based Model Selection	Page 19
Model Statistics	Page 22
Ranking Predictors for Select Models	Page 26
Discussion	Page 30
Limitations	Page 30
Conclusions	Page 33
Future Recommendations	Page 36
Beneficiaries	Page 37
References	Page 38
Appendix	Page 43

LIST OF TABLES

Table 1. Stand-level means and stand density index (SDI) for the study site and adjacent stand.	Page 5
Table 2. Summary statistics for select climate metrics used in linear modeling.	Page 13
Table 3. Derived vegetation indices of the multi-band hyperspectral scene.	Page 15
Table 4. Relative importance metrics for bootstrap analysis.	Page 18
Table 5. Criterion values for highest ranked regression models.	Page 19
Table 6. Regression model for Green Normalized Difference Vegetation Index (GNDVI) and total precipitation.	Page 23
Table 7. Summary statistics for regression of Green Normalized Difference Vegetation Index (GNDVI) and total precipitation	Page 23
Table 8. Regression model for $GDVI \times MTVI2 \times RENDVI \times SAVI \times TDVI$ and maximum air temperature.	Page 24
Table 9. Summary statistics for regression of $GDVI \times MTVI2 \times RENDVI \times SAVI \times TDVI$ and maximum air temperature.	Page 24
Table 10. Regression model for $GDVI \times MTVI2 \times RENDVI \times SAVI \times TDVI$ and average temperature.	Page 24
Table 11. Summary statistics for regression of $GDVI \times MTVI2 \times RENDVI \times SAVI \times TDVI$ and average temperature.	Page 25
Table 12. Regression model for $GDVI \times MTVI2 \times RENDVI \times SAVI \times TDVI$ and total solar radiation.	Page 25
Table 13. Summary statistics for regression of $GDVI \times MTVI2 \times RENDVI \times SAVI \times TDVI$ and total solar radiation.	Page 25

LIST OF FIGURES

Figure 1. Study site at Journagan Ranch.	Page 4
Figure 2. Sampling Layout for the Study Site.	Page 10
Figure 3. ModelBuilder workflow for extracting descriptive statistics for bands of the sample units.	Page 11
Figure 4. BIC ranking of regression models.	Page 20
Figure 5. Mallow's C_p ranking of regression models.	Page 21
Figure 6. Residual plots for highest ranked models.	Page 26
Figure 7. Contribution of predictors to the highest ranked model for maximum air temperature using relative importance metrics.	Page 27
Figure 8. Contribution of predictors to the highest ranked model for average temperature using relative importance metrics.	Page 28
Figure 9. Contribution of predictors to the highest ranked model for total solar radiation using relative importance metrics.	Page 29

INTRODUCTION

Literature Review

Within the last couple of decades, remote sensing has been substantiated as a relatively accurate and time-efficient approach to forest inventories and the investigation of forest variables (Bohlin et al., 2012; Gobakken et al., 2015; Magnussen, 2016; Puliti et al., 2015). Both spaceborne and airborne remote sensing have been important elements to vegetation spatial modeling in several innovative studies (Breidenbach & Astrup, 2012; Goerndt et al., 2010, 2011; Latifi et al., 2012), and airborne photogrammetry has proven to be a viable alternative to airborne laser scanning (ALS) for modeling forest biophysical properties (Bohlin et al., 2012; Gobakken et al., 2015). The use of unmanned aerial systems (UAS) for forest inventories with a multi-temporal aspect has also been explored (Wallace et al., 2014).

Line scanners using area array sensors have been flown on traditional fixed-wing aircraft for vegetation studies (Berni et al., 2009; Hruska et al., 2012), but it is their use with multi-rotor UAV (Lucieer et al., 2014) that has generated interest in exploring the practicality of this type of research. Hyperspectral imaging has gained attention in recent years due to increased spectral resolution (Latifi et al., 2012), and the development of compact sensors for unmanned aerial vehicles (UAV) has made gathering spectral information in the field more practical (Burkart et al. 2014; Nackaerts et al. 2010).

As pointed out in the literature, the accuracy of data produced by aerial remote sensing techniques is affected greatly by both flight altitude and the degree of image overlap between flight lines (Dandois et al., 2015). It has been suggested that the acquisition of aerial data at specific times during the vegetative season may increase the predictive power of forest inventory

models. Based on the recommendations of such work, flight dates for this study were chosen strategically. With regard to developing forest inventory models using biophysical and spectral variables, it was found that the accuracy of models including a spectral component was slightly higher than the accuracy of those that did not. However, several reasonable drawbacks of UAS-based spectral remote sensing have been identified, including: (1) the cost of equipment, (2) the limitations to coverage area resulting from battery capacity, and (3) the need for technical expertise to effectively process the data. (Puliti, et al., 2015)

With increasing concern for climate change, stress-related VIs are of growing interest in plant science. Concerns about drought, disease, and other biotic and abiotic factors have greatly contributed to their use. Past studies have reported a variety of visual indicators of plant stress, and it has been established that plant stress alters the reflectance of light on foliage (Nilsson, 1995; Weber & Jorg, 1991). It is this relationship that makes it possible for researchers to monitor and model vegetation response to stress. Several studies have explored the use of spaceborne multispectral remote sensing in the calculation of structural and stress-related VIs (Ishimura et al., 2011; Marx and Kleinschmit, 2017; Mohd Razali et al., 2016; Moon and Choi, 2015; Peng et al., 2016; Zhou et al., 2014), and more still have explored the use of UAS-based spectral imagery for precision forestry applications (Dash et al., 2018; DeWitt et al., 2017; Gini et al., 2012; Tao et al., 2011). Yet few have investigated the use of UAS-based spectral remote sensing for deriving stress or nutrient -related VIs (Goodwin et al., 2018; Sripada, 2005; Zarco-Tejada et al., 2012).

Research Objectives

The objective of this study was to develop useful models and assess methodology for modeling the response of a black walnut -dominant mixed hardwood stand to seasonal fluctuations in climate by the successful completion of four critical phases: (1) Conduct UAS flights to gather raw high-resolution hyperspectral and RGB aerial data at regular intervals for critical phenological shifts during the growing season; (2) develop a high-resolution orthomosaic from the aerial images; (3) perform corrections on the raw hyperspectral imagery; and (4) develop a spatial-temporal model to examine the relationship between derived vegetation stress metrics and select seasonal climate variables.

This study offers insight related to current methodologies for monitoring the health of hardwood forests using spectral remote sensing. The resulting model will be used to help local landowners predict the response of black walnut -dominant mixed hardwoods to seasonal climate stress and make informed management decisions. The model is not only useful to local landowners and landscape managers, but to local and regional researchers as well, for monitoring vegetation stress and response to the seasonal climate events and for ultimately maintaining sustainable forests in southern Missouri.

Study Site

The study area is located at Journagan Ranch, a Missouri State University-managed property approximately 10 miles south Mountain Grove, MO (Figure 1). The site was chosen for its high abundance of black walnut (*Juglans nigra*). Although black walnut is rarely found in the natural forest landscape, its sensitivity to drought stemming from its evolutionary adaptation to humid conditions and semi-moist soils makes it a good biological indicator of climate-induced

stress. Furthermore, its status as an economic staple species of agroforestry operations in the Midwest makes it an especially suitable candidate for study.

The study site consists primarily of mature, dominant black walnut; yet contains a few codominant, and several intermediate or suppressed, hardwood species. The black walnut were deliberately planted along the fence line of an open field approximately 60 years prior, and due to lack of management, several dominant hardwood species found in the stand canopy have become established in the sub-canopy of the study site.

Ground data was collected for a single fixed-area rectangular plot encompassing the study site. Since the site is relatively confined, ground data was collected for tenth-acre round plots evenly distributed across the adjacent stand as well. Diameter at breast height (DBH) was measured for each tree within the field plot(s) and stand-level statistics were calculated from the data for both stands (Table 1).

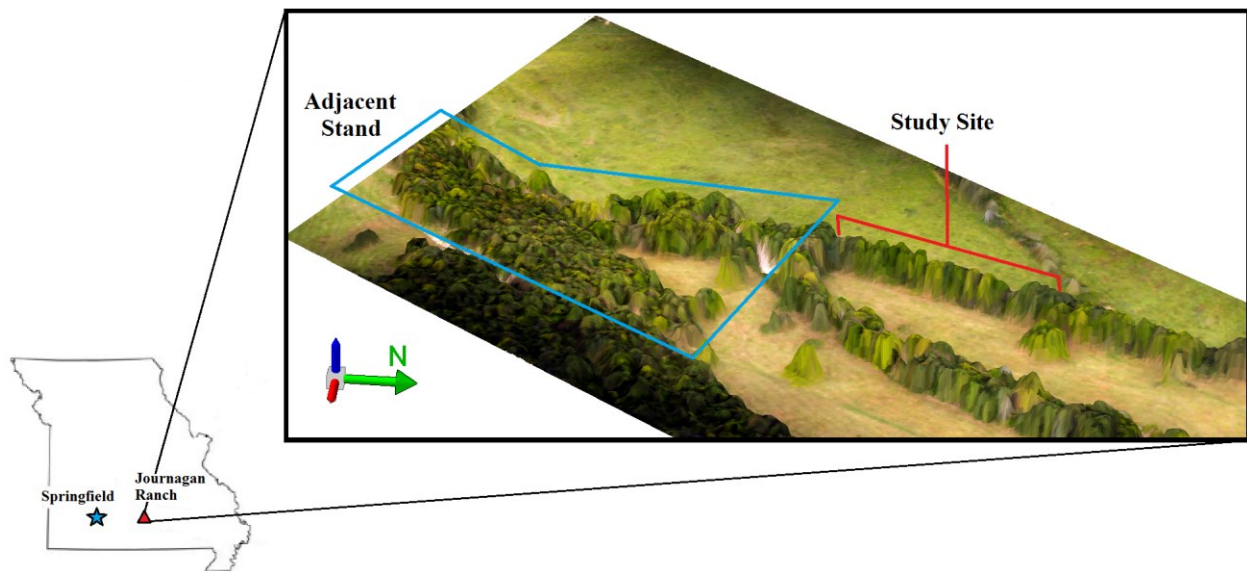


Figure 1. Study site at Journagan Ranch.

Table 1. Stand-level means and stand density index (SDI) for the study site and adjacent stand.

	n	DBH_μ	QMD	SDI
Study Site	35	11.03	12.90	93
Adjacent Stand	157	9.48	10.80	181

*n, number of observations; DBH_μ, mean diameter for trees; QMD, quadratic mean diameter for trees; SDI, stand density index.

Equipment

Assembled with a compact diffraction grating and concave mirror, this study utilizes a Headwall Nano-Hyperspec push-broom sensor that offers 270 spectral bands from 400-1000nm and 640 spatial bands across the track. The sensor was mounted to a DJI Matrice 600 Pro airframe. Flight plans were constructed in DJI Flight Planner, and DJI Ultimate Flight version 3 was used to execute the plans. The position of the UAV was measured by the XSens MTi-G-710 onboard IMU. With high precision, the IMU recorded the GPS location, roll, pitch, and yaw of the aircraft for the duration of the flight, and exported the information to a file for use in the geo- and ortho-rectification process.

A second UAS was mounted with a Sony a7R II camera and followed a similar flight path to collect aerial images of the forest canopy. The sensor was mounted to a DJI Spreading Wings S1000+ airframe using a Pixhawk autopilot board. The latest release of Mission Planner by Michael Osborne was used for loading flight plans to the Pixhawk autopilot board. Designed for various UAV, Mission Planner is a Windows-compatible feature-rich ground station software developed for the ArduPilot open source autopilot project (Mission Planner, 2017). Flight plans were developed with 80 percent overlap between neighboring photos for the ground control points (GCPs) to be witnessed in multiple images, improving the image matching process.

Both UAS were remotely controlled using DJI Ground Station Pro, a feature-rich ground station application for DJI aircraft, which defined a flying altitude of 120m and speed of 10m/s for both UAS. Both UAS are used in this study are considered portable, safe, and stable: They possess retractable landing gear, vibration dampers, angled arms, and a gimbal mount that offers a 360-degree view from camera or sensor. By assuring these conditions for high-resolution imagery, select vegetation metrics were successfully derived from the data.

The position of GCPs for each flight were collected using a handheld Trimble Nomad unit, running TerraSync v5.81, in tandem with a Trimble Pro 6H receiver, model 98850. Using four local ground stations of known position, a differential correction was performed in GPS Pathfinder Office 5.81 to produce submeter accuracy for GCPs. The G12BUS geoid model, which contains spatial information for WGS 1984 (UTM Zone 15N), was used in this process.

METHODS

UAS Data Collection

Both UAS were flown over the study site for three different flight dates in 2017: June 6th, July 31st, and September 12th. Dates were strategically chosen based on the beginning, middle, and end of the growing season, respectively. Each flight was scheduled approximately a month apart and were rescheduled accordingly in the event of inclement weather. Flights were conducted between approximately 10:00AM and 2:00PM, when the solar elevation angle was greater than 45 degrees, to reduce the effects of shadow. The hyperspectral sensor was calibrated using a standard white reference plate prior to each flight.

Development of an Orthomosaic

A digital surface model (DSM) was developed with the aerial imagery for each flight using AgiSoft PhotoScan Professional 1.3.4 (64-bit). The photogrammetry software uses feature matching techniques across the images, solves for intrinsic and extrinsic camera orientation parameters, performs a dense surface reconstruction using a combination of user-imported GCPs and software-generated tie points, and maps the texture to the resulting 3D surface (AgiSoft, 2017).

Next, an orthomosaic of the study site was created using the DSM and the image was exported in TIFF format. The image was then imported into the working project directory of ESRI's ArcGIS Desktop ArcMap 10.5.1. The processing parameters used for the development of an orthomosaic are outlined in the following list:

Align Photos

Accuracy: Medium
Generic Preselection: Yes
Key point limit: 40,000
Tie point limit: 2,000

Build Dense Cloud

Quality: Medium
Depth filtering: Aggressive

Build Mesh

Surface type: Height field
Source data: Dense cloud
Face count: Medium
Interpolation: Enabled

Build Orthomosaic

Surface: Mesh
Blending mode: Mosaic
Enable color correction: No
Enable hole filling: Yes

The default setting/value was used for all parameters not mentioned above.

Image Rectification and Sampling Design

The hyperspectral data was initially processed using Headwall SpectralView 5.5.1 (64-bit), a proprietary software designed for the sensor used. Due to the limitations of the software, it was used only to initially convert the raw image digital number values to at-sensor radiance values, as well as perform a crude geo- and ortho-rectification of the images before mosaicking them together using the *multi-ortho* tool. The hyperspectral mosaic for each of the three flights was then imported into ENVI 5.4.1 (64-bit), and a white reference calibration was performed. (Headwall Photonics, 2018)

Due to insufficiencies in the geo- and ortho-rectification of the mosaicked scene, ENVI was used to prepare the hyperspectral data for further image registration. This was done by first

visually identifying several features notable in both the uncorrected hyperspectral image, as well as the RGB orthomosaic created by PhotoScan. Secondly, a single pixel region of interest (ROI) was created for each of the identifying features, or tie points, in the hyperspectral image. Twenty features around the study site were identified and marked for each set of images. Next, a raster mask was built using the ROIs and a combination of 30 broadband and narrowband VIs were calculated from the spectral imagery using the inverse mask. This was done so that the resulting 30 VI band layers displayed a visual marker pixel at each notable feature; in effect, highlighting the location of 20 tie points per flight for rectification of the image. The resulting multiband raster image was exported in IMG format. This process was repeated for each flight. (EVIS, 2018)

In ESRI's ArcGIS Desktop ArcMap 10.5.1., the multiband raster image was imported, and a point feature class was created for both the multiband raster and RGB orthomosaic. A point was created for each blank pixel representing a tie point in the multiband raster. Similarly, a point was also created at the location each of the corresponding tie features in the RGB orthomosaic. Next, the *Warp* tool was used to better geo- and ortho-rectify the multiband raster image using the coordinates of the tie points from the hyperspectral image as *source* points and coordinates of the same tie features in the RGB orthomosaic as *target* points. The best results using the 20 selected tie points in the *Warp* tool were achieved with *transformation type* set to POLYORDER2 and *sampling type* set to BILINEAR. This process was repeated for each flight. (ESRI, 2018)

Once a sufficiently geo- and ortho- rectified multiband raster product was achieved for each flight, VI means were derived for several uniformly distributed sample units defined by a systematic centric sampling approach. In ArcMap, the Create Fishnet tool was used to construct

a north-to-south oriented grid over the study area. Using the Buffer tool, circular sample units were generated at the centers of grid squares so that all sample units were 2 meters in diameter. Sample units existing just outside the study area were removed. During the June flight, an anomaly in data collection resulted in a small sliver of missing data for the study area. Therefore, sample units having greater than fifty percent overlap with the small tract of missing data were removed from the sampling population for all flights. This resulted in a total of 105 sample units for each sample group—or flight (Figure 2).

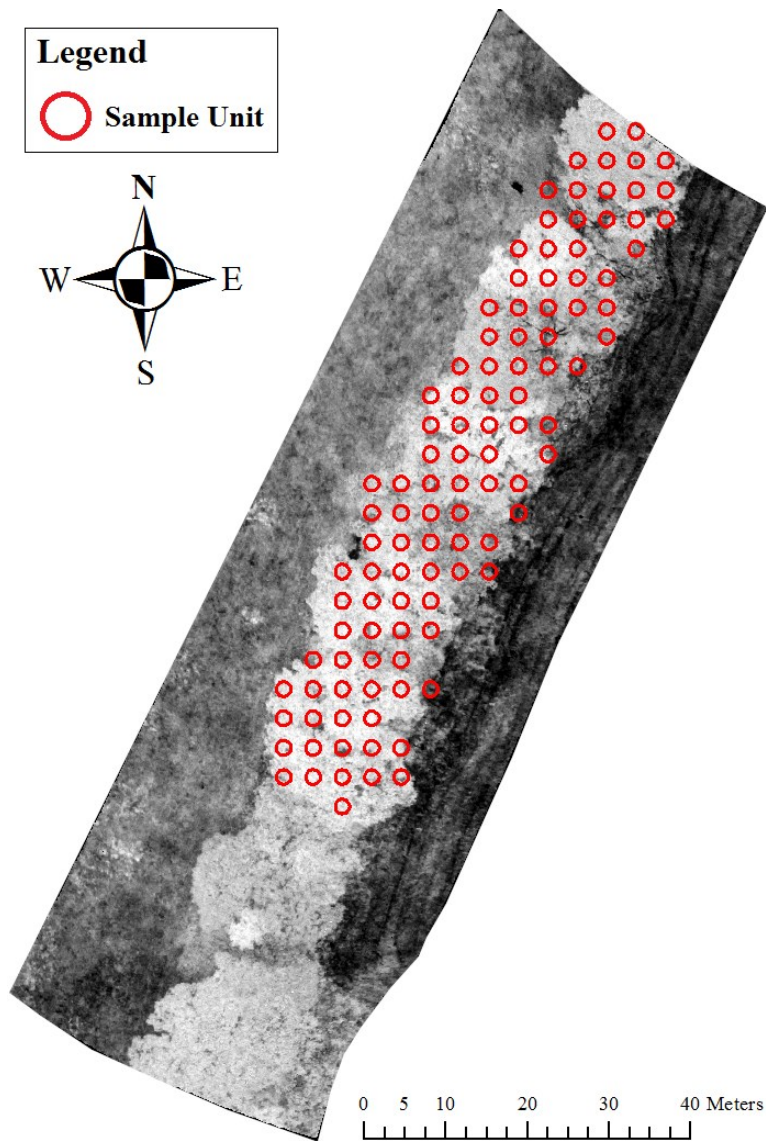


Figure 2. Sampling Layout for the Study Site.

Next, ModelBuilder for ArcMap was used to create a processing workflow that computes descriptive statistics for the bands of each sample unit within the multi-band raster image and outputs them to a text file (Figure 3).

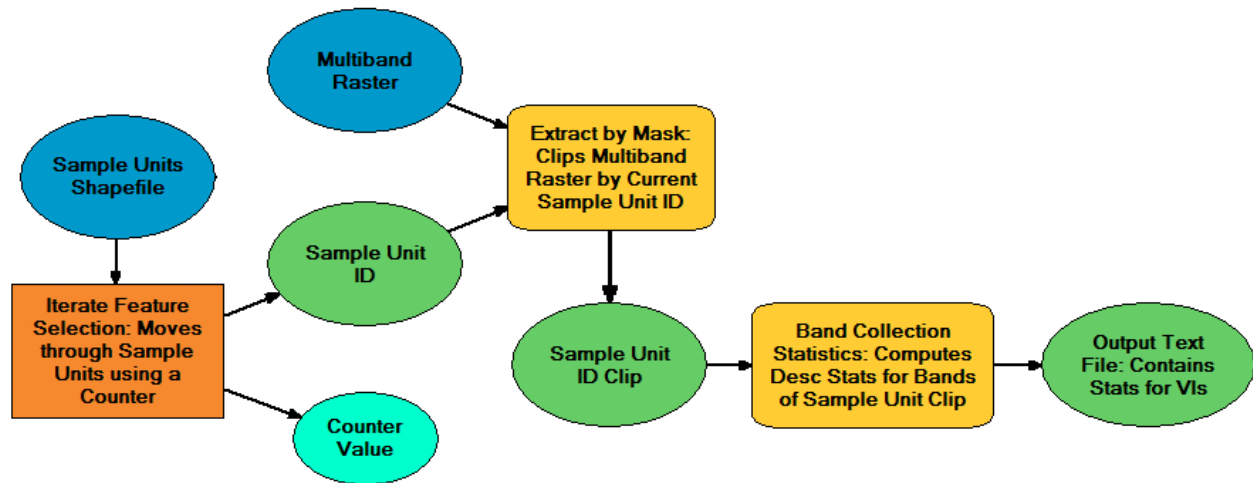


Figure 3. ModelBuilder workflow for extracting descriptive statistics for bands of the sample units.

The ModelBuilder workflow implements Iterate Feature Selection to move through the 105 sample units, and integrates two tools, Extract by Mask and Band Collection Statistics, to focus on a particular sample unit and to export descriptive statistics for each band to a file, respectively. This workflow was used for each of the flights.

Multiple Linear Regression for Climate Metrics

Changes in the derived metrics and corresponding changes in temperature and precipitation were integrated in a spatial-temporal model using a script written in R 3.5.1 (64-bit). First, the *read.table* function was used to import sample observation statistics from individual text files. Using a loop and the *rbind* function, the VI means for each sample unit were merged into a single table sorted by plot number. Subsequently, tables for each flight were

appended using the *rbind* function so that repeated measures were sorted by flight (1-3) and observation number (1-105). Column names were added manually to the combined table using the *colnames* function (See Appendix).

Next, four climate metrics were chosen from a series of climate variables measured by the Wright county weather station at Mountain Grove, MO. The daily weather data was downloaded from the historical agriculture weather database for a span of two weeks prior to each flight (MHAWD, 2018). The mean values of the four chosen metrics were calculated in Microsoft Office Excel 2013 for each of the three flights and transferred to a tab-delimited text file for use as response variables for multiple linear regression. Summary statistics for the chosen climate metrics, as measured daily for fourteen consecutive days prior to each flight, are displayed in Table 2.

In the R script, the *spectable* function was used to select a single climate metric, and the *regsubsets()* function from the *leaps* package was used to perform a best subsets regression. Since stepwise regressions only investigate a fraction of the possible models in highly dimensional data, exhaustive regression was used instead (R Development Core Team, 2018; Goerndt et al., 2010, 2011; McGaughey, 2008; see Appendix). Furthermore, an exhaustive approach isn't biased by the order of predictor variables when performing multiple linear regression. This is important since the data is arranged in such a way that neighboring columns (VIs) in the R data frame used for regression are not intentionally related. That is, the VIs selected for analysis may or may not rely on the same spectral region for their calculation; therefore, any covariance between their values is circumstantial with the alphabetical ordering of VI names.

Table 2. Summary statistics for select climate metrics used in linear modeling.

Flight 1							
	Mean	Median	StDev	s²	SE	Min	Max
Total Precipitation (in)	0.2	0.01	0.44	0.19	0.12	0	1.3
Maximum Air Temperature (°F)	77.16	78.85	5.83	34.02	1.56	61.6	84
Average Temperature (°F)	66.04	67.6	5.28	27.88	1.41	52.8	71.4
Total Solar Radiation (MJ/M²)	18.76	18.14	3.9	15.21	1.04	13.8	25.8
Flight 2							
	Mean	Median	StDev	s²	SE	Min	Max
Total Precipitation (in)	0.05	0	0.17	0.03	0.05	0	0.6
Maximum Air Temperature (°F)	91.05	92.15	4.56	20.76	1.22	83.2	96.9
Average Temperature (°F)	79.41	79.6	4.11	16.87	1.1	70.8	85
Total Solar Radiation (MJ/M²)	21.02	23.36	4.82	23.23	1.29	9.5	26
Flight 3							
	Mean	Median	StDev	s²	SE	Min	Max
Total Precipitation (in)	0	0	0	0	0	0	0
Maximum Air Temperature (°F)	78.39	78.5	5.19	26.95	1.39	69	88.4
Average Temperature (°F)	66.12	65.75	5.25	27.53	1.4	56	77
Total Solar Radiation (MJ/M²)	18.89	19.56	3.39	11.49	0.91	10.73	22.76

*StDev, standard deviation; s², sample variance; SE, standard error

The 30 VIs selected for analysis are a list of both broadband and narrowband indices (Table 3). Broadband VIs are the most standardized measures of the overall amount and quality of photosynthetic material in vegetation. The simplest type of index, broadband VIs compare

reflectance measurements from red and near-infrared (NIR) regions. Even with sensors of lower spectral resolution, broadband VIs are effective for many remote sensing applications. Similar to the broadband VIs, narrowband VIs are also formulated to measure the amount and quality of photosynthetic material in vegetation. However, most narrowband VIs are calculated with greater resolution and focus on reflectance measurements for the red edge portion of the spectral reflectance curve. Used to describe the sloped region of the vegetation reflectance curve from 690nm to 740nm, the red edge is the result of the shift from chlorophyll absorption in the high red wavelengths to the lower wavelengths of NIR region. Since changes in this region are directly linked to changes in chlorophyll absorption in the plant, the red edge is a particularly suitable region for monitoring changes in overall plant health. Furthermore, narrowband VIs are more sensitive to slight deviations in plant health than broadband VIs, especially for areas of dense vegetation where broadband measures may saturate (Narrowband Greenness, 2018).

With 30 VIs, or predictor variables p , there are 2^p possible models. More complex models will often fit better and produce a smaller residual sum of squares (RSS) at the expense of using several predictor variables. Thus, the best model balances fit with size. Bayesian Information Criterion (BIC) penalizes larger models more heavily compared to other criterion, such as Akaike Information Criterion (AIC). Since BIC fluctuates only in the comparison of differently-sized models, the results are independent of the choice of complexity. BIC was used to choose the top three models for regression with each of four climate variables, response variables y_1 , y_2 , y_3 , and y_4 . BIC is defined as

$$\text{BIC} = n \log \frac{\text{RSS}}{n} + p \log (n),$$

where n is number of observations. AIC may choose too complex a model, despite n (Akaike, 1973). While BIC does not have this problem if n is sufficient, it still may choose too simple a

Table 3. Derived vegetation indices of the multi-band hyperspectral scene.

Band	Index Name	Index Type	
		Broadband	Narrowband
B1	Anthocyanin Reflectance Index 1 (ARI1)		*
B2	Anthocyanin Reflectance Index 2 (ARI2)		*
B3	Carotenoid Reflectance Index 1 (CRI1)		*
B4	Carotenoid Reflectance Index 2 (CRI2)		*
B5	Green Atmospherically Resistant Index (GARI)	*	
B6	Green Difference Vegetation Index (GDVI)	*	
B7	Green Normalized Difference Vegetation Index (GNDVI)	*	
B8	Green Ratio Vegetation Index (GRVI)	*	
B9	Leaf Area Index (LAI)	*	
B10	Modified Chlorophyll Absorption Ratio Index (MCARI)		*
B11	Modified Chlorophyll Absorption Ratio Index 2 (MCARI2)		*
B12	Modified Red Edge Normalized Difference Vegetation Index (MRENDVI)		*
B13	Modified Red Edge Simple Ratio (MRESR)		*
B14	Modified Simple Ratio (MSR)	*	
B15	Modified Triangular Vegetation Index (MTVI)		*
B16	Modified Triangular Vegetation Index - Improved (MTVI2)		*
B17	Normalized Difference Vegetation Index (NDVI)	*	
B18	Photochemical Reflectance Index (PRI)		*
B19	Plant Senescence Reflectance Index (PSRI)		*
B20	Red Edge Normalized Difference Vegetation Index (RENDVI)		*
B21	Red Edge Position Index (REPI)		*
B22	Red Green Ratio Index (RGRI)		*
B23	Simple Ratio (SR)	*	
B24	Soil Adjusted Vegetation Index (SAVI)	*	
B25	Structure Insensitive Pigment Index (SIPI)		*
B26	Transformed Chlorophyll Absorption Reflectance Index (TCARI)		*
B27	Transformed Difference Vegetation Index (TDVI)	*	
B28	Vogelmann Red Edge Index 1 (VREI1)		*
B29	Vogelmann Red Edge Index 1 (VREI2)		*
B30	Water Band Index (WBI)		*

model for any given n . BIC closely resembles K-fold cross-validation and AIC closely resembles leave-one-out cross-validation. AIC is defined as

$$AIC = n \log \frac{RSS}{n} + 2p.$$

The equation for BIC is simply the equation for AIC whose penalty term has been multiplied by the factor $\frac{1}{2} \log(n)$ in preference of less complex models (Schwarz, 1978). Another popular criterion is Mallow's C_p (Mallows, 1973). Mallow's C_p has been shown to be equivalent to AIC in the case of Gaussian linear regression (Boisbunon, 2013) and was used as proxy to AIC for plotting purposes. Even so, AIC values were reported alongside BIC values for a comparison of ranking. Mallow's C_p is defined as

$$C_p = \frac{RSS}{\sigma^2} + 2p - n,$$

where σ is the variance for the model using all predictors.

The regression models can be represented by the following equation:

$$\hat{Y}_i = \beta_0 + \beta_1 x_B + \dots + \beta_p x_b,$$

where $i = 1, 2, 3, 4$ for the climate variables total precipitation ($i = 1$), maximum air temperature ($i = 2$), average temperature ($i = 3$), and total solar radiation ($i = 4$); β_0 is the intercept of the regression equation; and $b = 1, 2, 3, \dots, 30$ for each of 30 VIs.

In the R script, the number of subsets of each size to record, *nbest*, was set to 5. The maximum size of the subsets allowed for examination, *nvmax*, was also set to 5. In effect, the *regsubset* function returned the best models with up to five allowed predictor variables. BIC and Mallow's C_p values for each model were plotted. AIC, BIC, and Mallow's C_p , were used to collectively rank the regression models for each seasonal climate metric. The highest ranked model for each metric was selected for further analysis. Summary statistics were calculated and the residual values were plotted against the predicted values for each of the selected models.

Finally, a bootstrap analysis of the three multivariate regression models was performed to discern the relative importances of individual predictors and their contribution to R-squared using the *boot.relimp()* and *booteval.relimp()* functions of the *relaimpo* package for R. This helps prevent the over-interpretation of differences. The parameter for the number of bootstrap resamples was left at its default value of 1000 (R Development Core Team, 2018; see Appendix). Six relative importance metrics were chosen for the bootstrapping procedure. The name, brief description, and advantages and disadvantages of each metric are in Table 4.

The difficulty in decomposing R-squared for regression models with correlated predictors stems from the fact that each arrangement of predictors leads to a different decomposition of the model sum of squares. Although the use of individual metrics may be ineffective in determining relative importance, combinations of several metrics can provide insights regarding the contributions of predictors (Groemping, 2006). For this reason, the ranking of predictor contributions by all six relative importance metrics were considered.

Table 4. Relative importance metrics for bootstrap analysis.

Metric	Description	Advantages and Disadvantages
<i>first</i>	Calculates R-squared when the predictor is the only predictor in the model	Only uses direct effect of predictor Does not decompose R-squared into predictor contributions
<i>last</i>	Calculates the increase in R-squared when the predictor is added to the model with all other predictors	Does not use direct effect of predictor Does not decompose R-squared into predictor contributions
<i>betasq</i>	Calculates the squared standardized coefficient using the empirical variances of the predictor and response	Marginal use of direct effect of predictor Does not provide a natural decomposition of R-squared
<i>pratt</i>	Calculates the product of the standardized coefficient and marginal correlation	Use is greatly limited by nature of the data Provides a semi-natural decomposition of R-squared
<i>lmq</i>	Calculates sequential R-squared values, but accounts for dependence on ordering by averaging over orderings using unweighted averages	Successfully decomposes R-squared into contributions that sum to the total R-squared Contribution may be overestimated for a predictor if correlated with other highly-contributing predictor(s)
<i>pmvd</i>	Calculates sequential R-squared values, but accounts for dependence on ordering by averaging over orderings using weighted averages with data-dependent weights	Successfully decomposes R-squared into contributions that sum to the total R-squared Variability in estimated contributions of predictors dependent upon the nature of the data.

*For more details related to relative importance metrics, see Groemping (2006).

RESULTS

Criterion-based Model Selection

The highest ranked regression model for total precipitation was (Figures 4, 5):

$$\hat{Y}_1 = \beta_0 + \beta_1 x_{B7},$$

The highest ranked regression model for maximum air temperature, average temperature, and total solar radiation based on AIC, BIC, and C_p were (Table 5; Figures 4, 5):

$$\hat{Y}_2 = \beta_0 + \beta_1 x_{B6} + \beta_2 x_{B16} + \beta_3 x_{B20} + \beta_4 x_{B24} + \beta_5 x_{B27},$$

$$\hat{Y}_3 = \beta_0 + \beta_1 x_{B6} + \beta_2 x_{B16} + \beta_3 x_{B20} + \beta_4 x_{B24} + \beta_5 x_{B27}, \text{ and}$$

$$\hat{Y}_4 = \beta_0 + \beta_1 x_{B6} + \beta_2 x_{B16} + \beta_3 x_{B20} + \beta_4 x_{B24} + \beta_5 x_{B27}, \text{ respectively.}$$

For each climate metric, the ranking results using different criteria were equivalent. Regardless of the models ranked highest by AIC, BIC, and Mallow's C_p , GDVI, MTVI, NDVI, and VREI2 appeared as predictors in several middle-ranked regression models for all climate metrics. (Figures 4, 5)

Table 5. Criterion values for highest ranked regression models.

Model	df	AIC	BIC
y_1	3	-670.29	-659.03
y_2	7	529.53	555.80
y_3	7	-30.120	-3.8519
y_4	7	-823.31	-797.04

*df, degrees of freedom

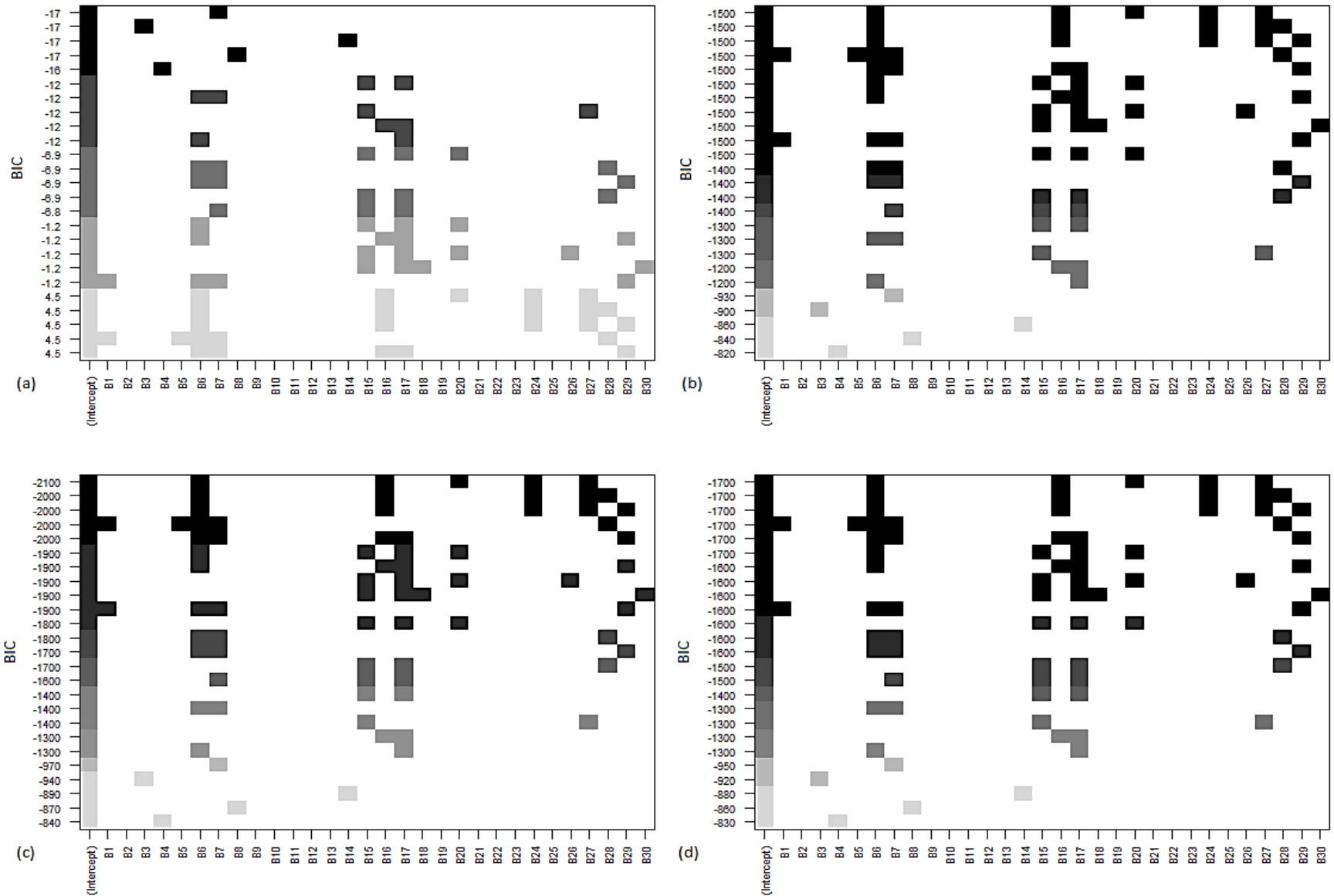


Figure 4. BIC ranking of regression models for (a) total precipitation, (b) maximum air temperature, (c) average temperature, and (d) total solar radiation.

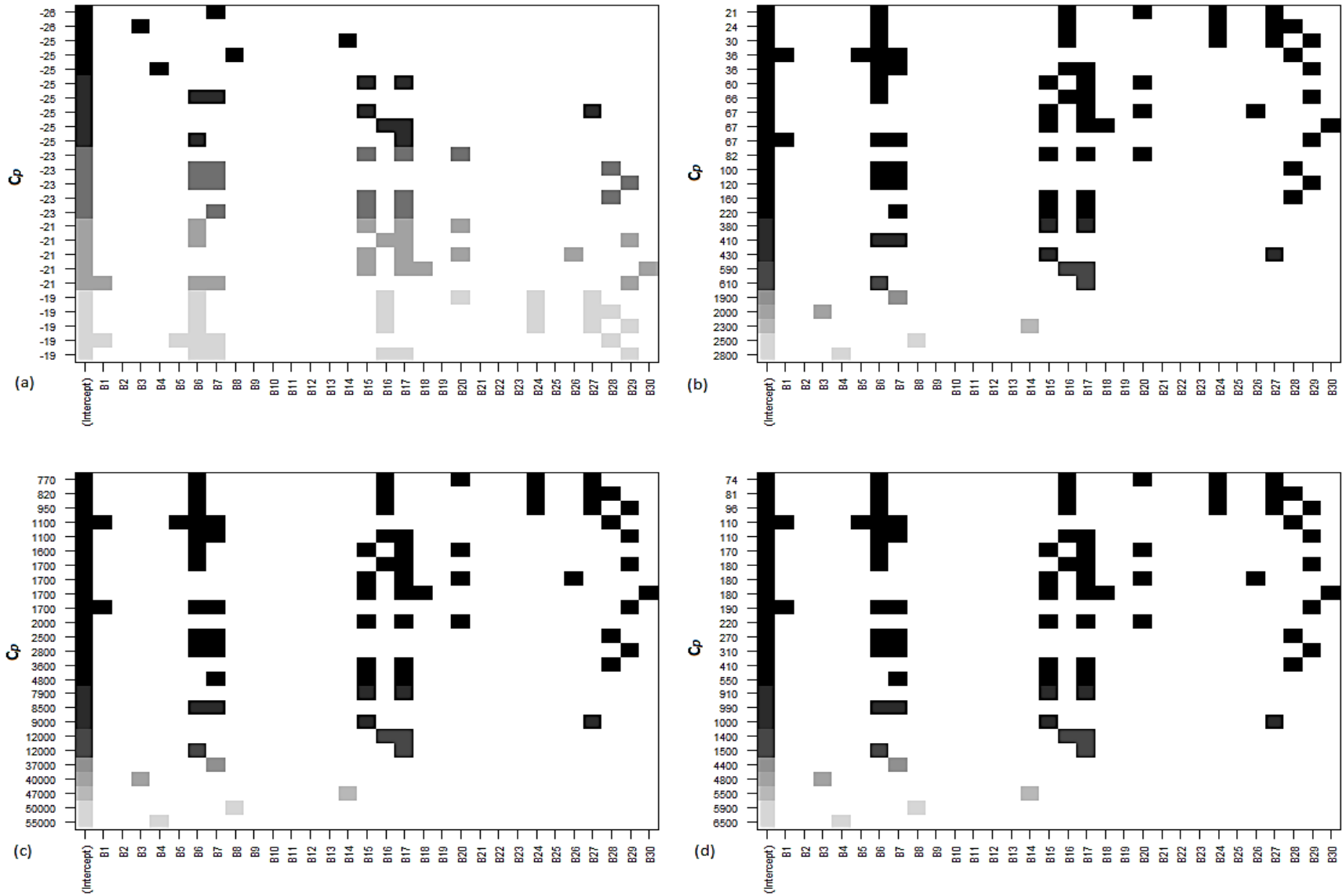


Figure 5. Mallow's C_p ranking of regression models for (a) total precipitation, (b) maximum air temperature, (c) average temperature, and (d) total solar radiation.

Model Statistics

The highest ranked model for total precipitation as identified by criterion-based model selection was with the predictor variable GNDVI (Table 6). While the model yielded a p value of $9.959\text{e-}08$ and RMSE value of 0.083 inches, the adjusted R-squared value for the model was only 0.087, indicating poor fit of the data to the regression line (Table 7).

The highest ranked model for maximum air temperature as identified by criterion-based model selection was with the predictor variables GDVI, MTVI2, RENDVI, SAVI, and TDVI (Table 8). The model yielded a p value of less than $2.2\text{e-}16$, RMSE value of 0.55°F , and adjusted R-squared value of 0.99, indicating good fit of the data to the regression line (Table 9).

The highest ranked model for average temperature as identified by criterion-based model selection was with the predictor variables GDVI, MTVI2, RENDVI, SAVI, and TDVI (Table 10). The model yielded a p value of less than $2.2\text{e-}16$, RMSE value of 0.23°F , and adjusted R-squared value very close to one, indicating good fit of the data to the regression line (Table 11).

The highest ranked model for total solar radiation as identified by criterion-based model selection was with the predictor variables GDVI, MTVI2, RENDVI, SAVI, and TDVI (Table 12). The model yielded a p value of less than $2.2\text{e-}16$, RMSE value of 0.064°F , and adjusted R-squared value very close to one, indicating good fit of the data to the regression line (Table 13).

The residuals were plotted against the predicted values for the highest ranked model of each climate metric. With only three flights, or sample groups, the residuals values for each of the models are clearly aggregated into discrete clusters at the three predicted means (Figure 6). It is clear, for the highest ranked regression model for maximum air temperature (Figure 6b), average air temperature (Figure 6c), and total solar radiation (Figure 6d), that the models are relatively unbiased. That is, variation for these models can be attributed primarily to the random

variation in the residual values between sample units (i.e., within-group variance). In contrast, the highest ranked model for total precipitation exhibits significantly greater relative variation in the predicted values for each sample group (i.e., between-group variance; Figure 6a), indicating relatively low model strength. Due to the discrete nature of the data resulting from a low number of sample groups, patterns in the distribution of error terms cannot be inferred from the results with confidence. However, it should be noted that the distribution of error for the highest ranked model for average temperature (Figure 6c) appears slightly more homoscedastic in comparison with all other models (Figure 6, a, b, and d), based on the distribution of residual groupings around the zero line.

Table 6. Regression model for Green Normalized Difference Vegetation Index (GNDVI) and total precipitation.

Predictors	Estimate	SE	T value	Pr(> t)
Intercept	0.11	< 0.01	16.57	< 2e-16
GNDVI	-0.09	0.02	-5.46	9.96e-08

*SE, standard error

Table 7. Summary statistics for regression of Green Normalized Difference Vegetation Index (GNDVI) and total precipitation.

df	Residual SE	RMSE	Multiple R²	Adj R²	F statistic	P value
313	0.083	0.083	0.087	0.084	29.76	9.959e-08

*df, degrees of freedom; RMSE, root mean squared error (in degrees Fahrenheit)

Table 8. Regression model for $\text{GDVI} \times \text{MTVI2} \times \text{RENDVI} \times \text{SAVI} \times \text{TDVI}$ and maximum air temperature.

Predictors	Estimate	SE	T value	Pr(> t)
Intercept	39.82	1.37	29.11	< 2e-16
GDVI	-2.70	0.10	-26.57	< 2e-16
MTVI2	-13.41	0.44	-30.57	< 2e-16
RENDVI	-13.29	1.32	-10.10	< 2e-16
SAVI	14.79	0.65	22.76	< 2e-16
TDVI	52.83	1.89	27.92	< 2e-16

*SE, standard error

Table 9. Summary statistics for regression of $\text{GDVI} \times \text{MTVI2} \times \text{RENDVI} \times \text{SAVI} \times \text{TDVI}$ and maximum air temperature.

df	Residual SE	RMSE	Multiple R²	Adj R²	F statistic	P value
309	0.554	0.549	0.992	0.992	8031	< 2.2e-16

*df, degrees of freedom; RMSE, root mean squared error (in degrees Fahrenheit)

Table 10. Regression model for $\text{GDVI} \times \text{MTVI2} \times \text{RENDVI} \times \text{SAVI} \times \text{TDVI}$ and average temperature.

Predictors	Estimate	SE	T value	Pr(> t)
Intercept	27.95	0.56	49.69	< 2e-16
GDVI	-2.71	0.04	-64.88	< 2e-16
MTVI2	-13.47	0.18	-74.64	< 2e-16
RENDVI	-13.35	0.54	-24.67	< 2e-16
SAVI	14.85	0.27	55.57	< 2e-16
TDVI	53.06	0.78	68.17	< 2e-16

*SE, standard error

Table 11. Summary statistics for regression of GDVI \times MTVI2 \times RENDVI \times SAVI \times TDVI and average temperature.

df	Residual SE	RMSE	Multiple R²	Adj R²	F statistic	P value
309	0.228	0.226	0.999	0.999	4.787e04	< 2.2e-16

*df, degrees of freedom; RMSE, root mean squared error (in degrees Fahrenheit)

Table 12. Regression model for GDVI \times MTVI2 \times RENDVI \times SAVI \times TDVI and total solar radiation.

Predictors	Estimate	SE	T value	Pr(> t)
Intercept	12.54	0.16	78.51	< 2e-16
GDVI	-0.45	0.01	-37.64	< 2e-16
MTVI2	-2.22	0.05	-43.31	< 2e-16
RENDVI	-2.20	0.15	-14.31	< 2e-16
SAVI	2.45	0.08	32.24	< 2e-16
TDVI	8.74	0.22	39.55	< 2e-16

*SE, standard error

Table 13. Summary statistics for regression of GDVI \times MTVI2 \times RENDVI \times SAVI \times TDVI and total solar radiation.

df	Residual SE	RMSE	Multiple R²	Adj R²	F statistic	P value
309	0.065	0.064	0.996	0.996	1.61e+04	< 2.2e-16

*df, degrees of freedom; RMSE, root mean squared error (in degrees Fahrenheit)

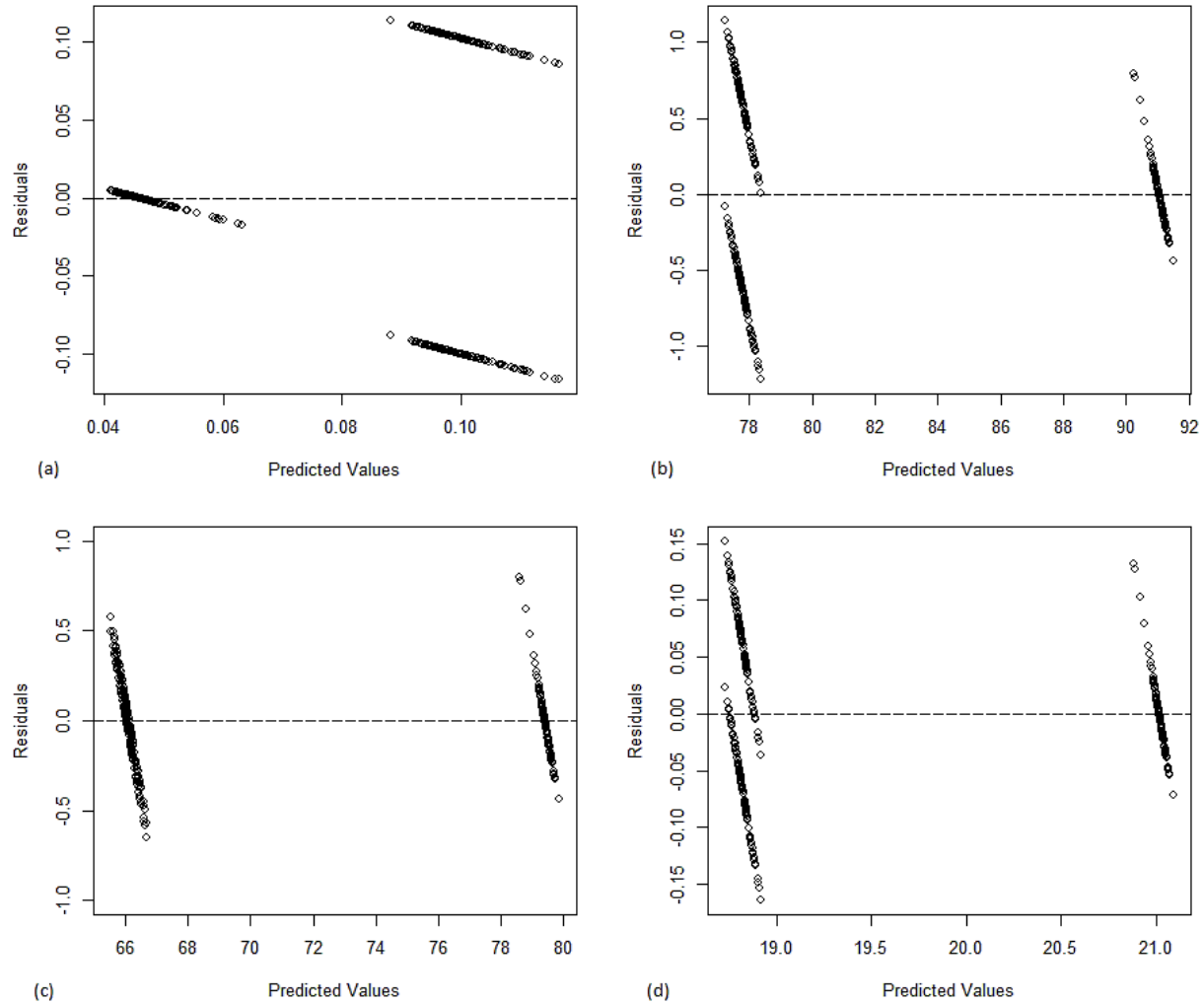


Figure 6. Residual plots for highest ranked models for (a) total precipitation against GDVI, (b) maximum air temperature against $GDVI \times MTVI2 \times RENDVI \times SAVI \times TDVI$, (c) average temperature against $GDVI \times MTVI2 \times RENDVI \times SAVI \times TDVI$, and (d) total solar radiation against $GDVI \times MTVI2 \times RENDVI \times SAVI \times TDVI$.

Ranking Predictors for Select Models

The contribution of predictors to the highest ranked model for maximum air temperature (Figure 7), average air temperature (Figure 8), and total solar radiation (Figure 9) was determined using six relative importance metrics (Table 4). TDVI (B27), appears to have the greatest contribution to R-squared for the highest ranked model for maximum air temperature according to all relative importance metrics except *last* and *first* (Figure 7, c and d). The

contribution of predictors to the highest ranked model for average temperature (Figure 8) and total solar radiation (Figure 9) produced identical ordering results for all relative importance metrics.

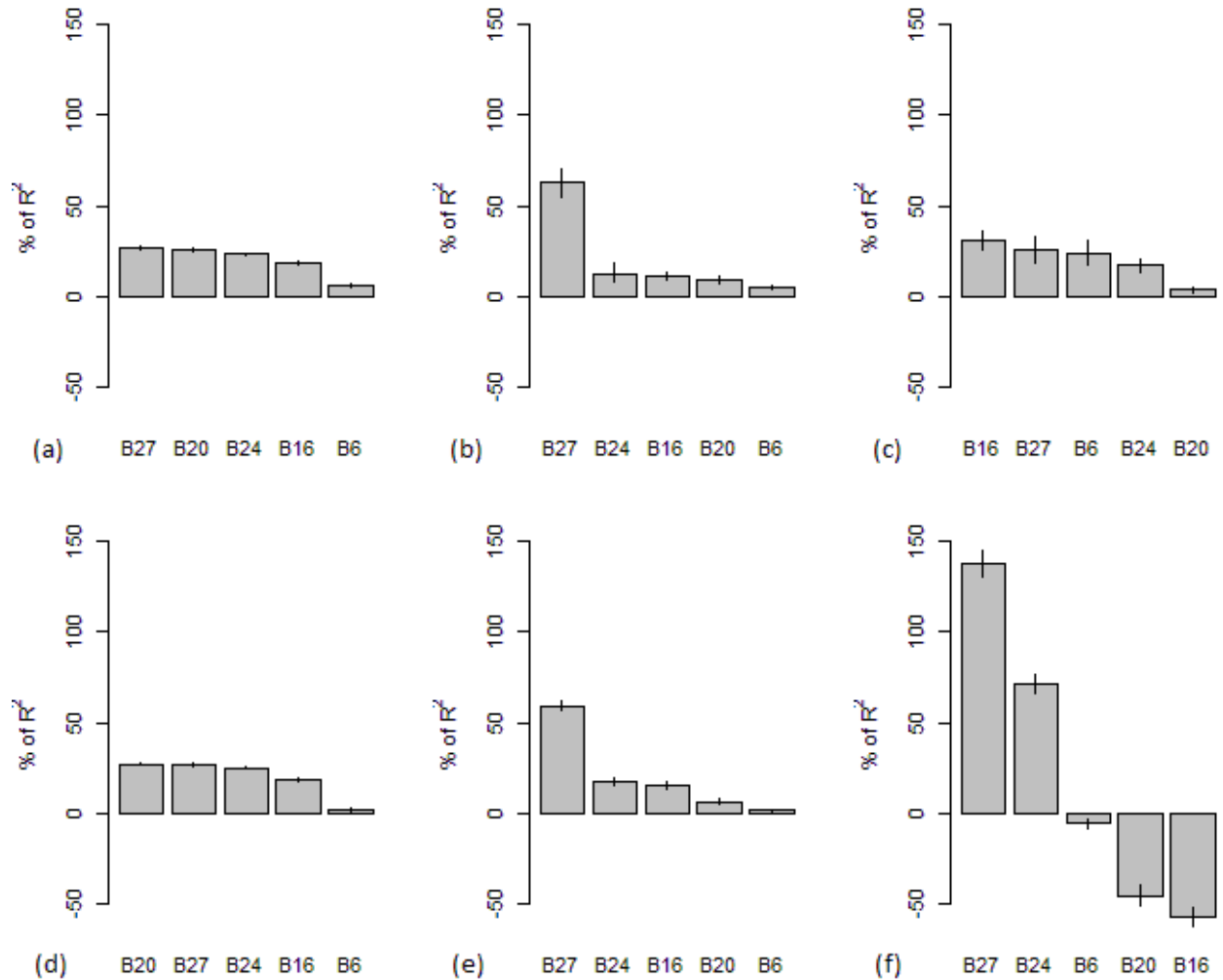


Figure 7. Contribution of predictors to the highest ranked model for maximum air temperature using relative importance metrics (a) *lmg*, (b) *pmvd*, (c) *last*, (d) *first*, (e) *betasq*, and (f) *pratt*. Performed using a 95% bootstrap confidence interval. Metrics normalized to sum 100%.

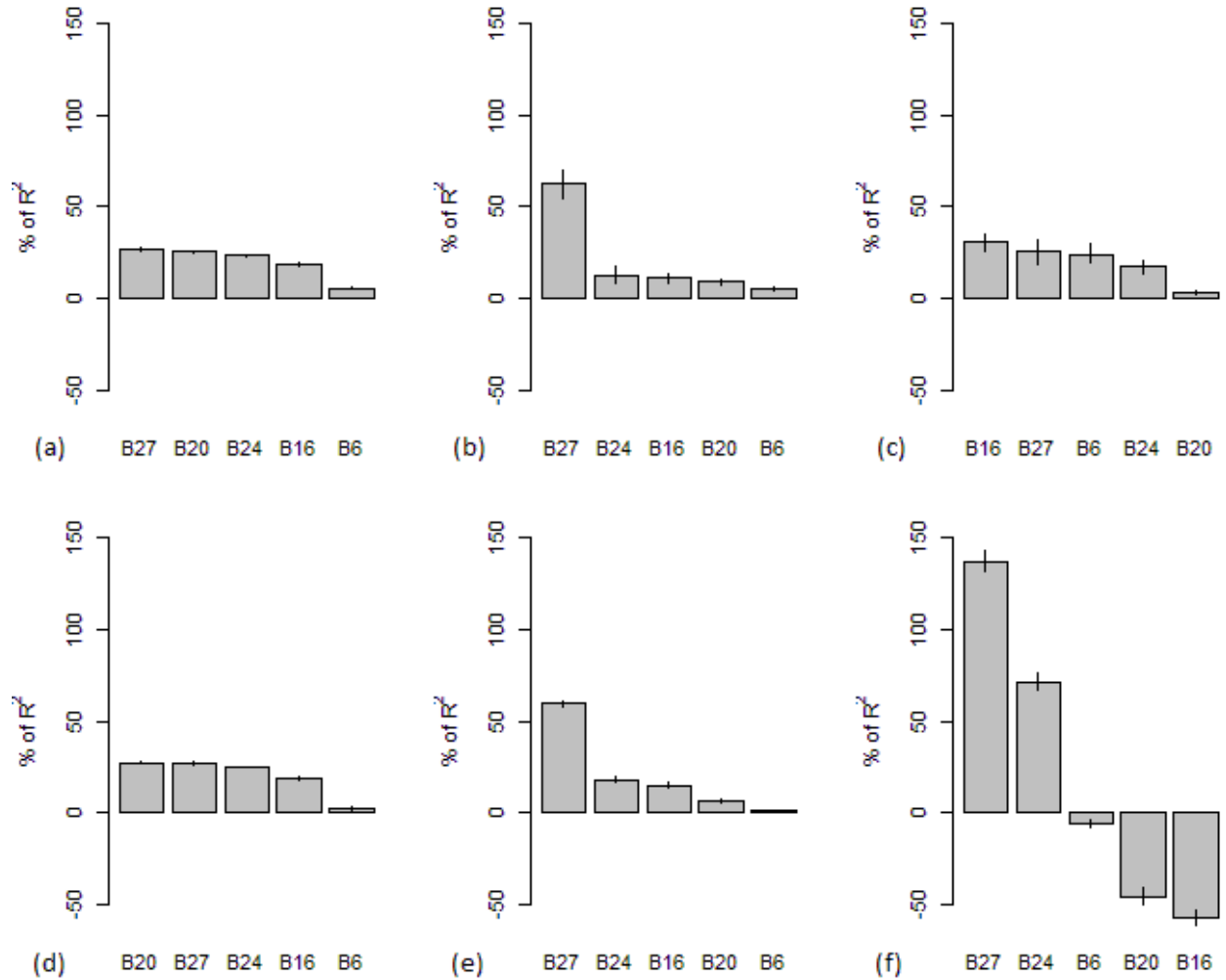


Figure 8. Contribution of predictors to the highest ranked model for average temperature using relative importance metrics (a) *lmg*, (b) *pmvd*, (c) *last*, (d) *first*, (e) *betasq*, and (f) *pratt*. Performed using a 95% bootstrap confidence interval. Metrics normalized to sum 100%.

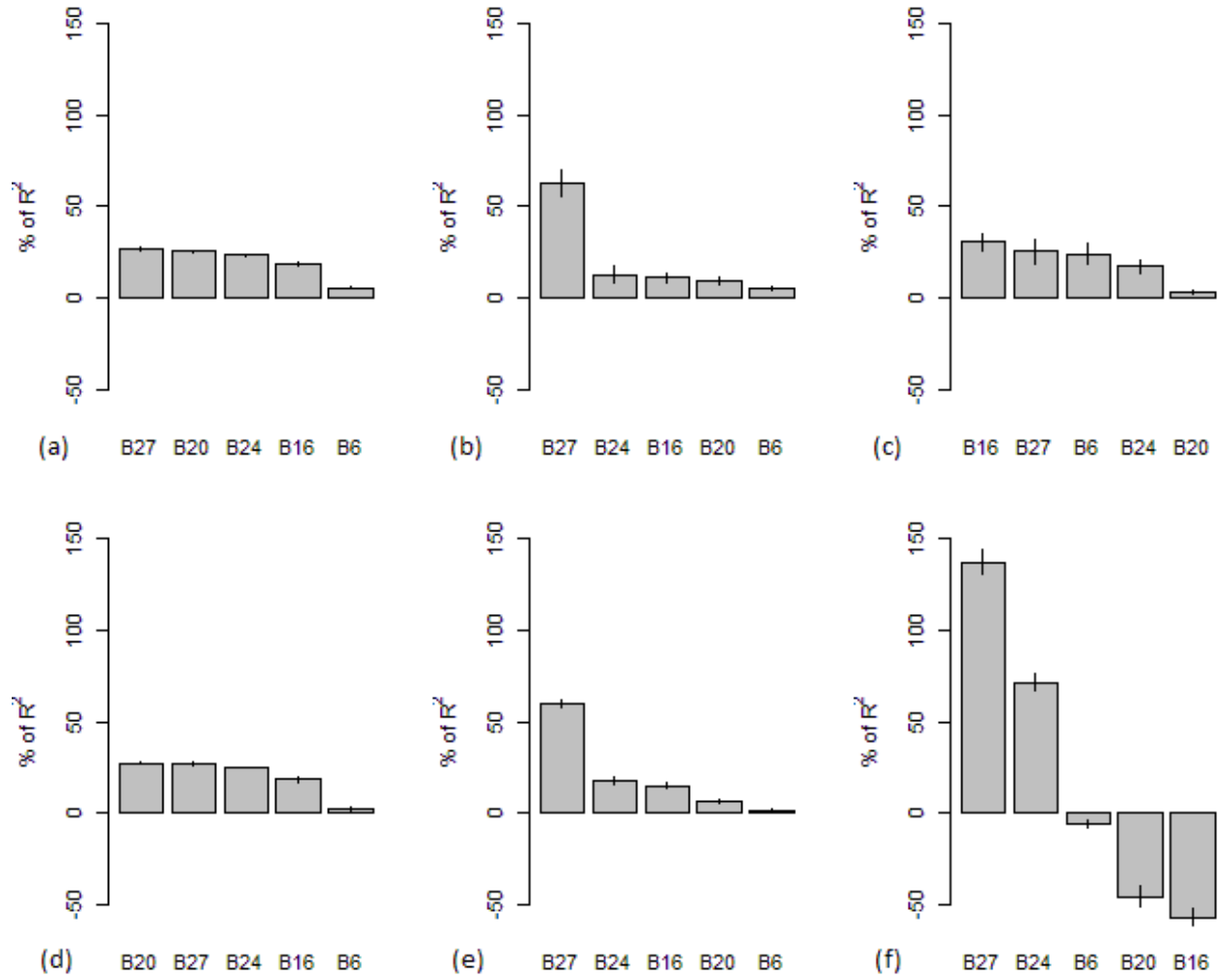


Figure 9. Contribution of predictors to the highest ranked model for total solar radiation using relative importance metrics (a) *lmg*, (b) *pmvd*, (c) *last*, (d) *first*, (e) *betasq*, and (f) *pratt*. Performed using a 95% bootstrap confidence interval. Metrics normalized to sum 100%.

DISCUSSION

Limitations

The two most notable limitations that shaped the methodology of this study were (1) the difficulties in geo- and ortho-rectifying the hyperspectral data and (2) the small number of sample groups for linear modeling. The complications in geo- and ortho-rectifying the hyperspectral data stemmed directly from the inability of the proprietary software to properly remove the distorting effects of tilt and terrain relief. For this reason, a manual transformation of the imagery using image features was necessary before sampling was possible. While the number of sample units was reasonably sufficient, the number of sample groups was rather low. Originally, there was intent to use the hyperspectral data collected for the previous year, effectively doubling the number of sample groups and increasing the predictive power of the linear models. However, disabilities of the sensor software led to workarounds which were impractical given the time constraints of the study, ultimately leading to the exclusion of this data from the analysis.

Due to the proprietary nature of the sensor data, correction of the imagery using the manufacturer's software was the only practical option, despite its limitations. The proprietary nature stems largely from the use of a nonstandard unit of time for image timestamps. By using nonstandard time units, the software failed to account for data gaps (i.e., periodic instances of data not being written to the camera storage) resulting in the alignment of collected data with incorrect timestamps. This required the construction of surrogate data for gaps so that the software could process the images. The use of standard GPS time units and incorporation of a simple detection algorithm for these gaps would serve well toward resolving this issue.

Another key inadequacy of the software that ultimately affected the end product was failure to support the orthorectification process with genuine elevation data. SpectralView uses height above the reference ellipsoid (GRS80, in this case) to approximate the earth's surface, rather than orthometric height. Orthometric height is the height above the reference geoid, which is determined by the earth's gravity and approximated by mean sea level. Geoid models provide an irregular, yet arguably more precise estimation of elevation than ellipsoid models, which assume that earth's surface is smooth. Geoid models would provide a better representation of actual elevation and likely counter many failings of the software's orthorectification process. The effect of this inadequacy resulted in an improper scaling of images in the lateral direction and required the orthorectification of cubes on an individual basis.

Aside from the time inconvenience, individually orthorectifying cubes using the software revealed other faults. For each image, the value of a software parameter labeled *altitude offset* had to be roughly approximated until the scaling of the orthorectified image was sufficient for mosaicking. While the range of values for this adjustment factor closely resembled the difference between the ellipsoid and geoid height, the fact that this parameter must be approximated by the user demonstrates the inefficiency of the software's orthorectification protocol. Once orthorectified, the *multi-ortho* tool was used to mosaic the images.

Misleading in its name, the *multi-ortho* tool does not orthorectify multiple images. Instead it attempts to join already orthorectified images for the purpose of creating a spatially correct hyperspectral scene. Previously orthorectified images are displayed in a window and the position of each must be adjusted manually so that an adequate product may be achieved. Despite the effects of user-error in guessing at orthorectification parameters, this is somewhat manageable for a very small number of sufficiently orthorectified images (5 or less). However,

the problem of mosaicking becomes evident when attempting to orthorectify moderate to large numbers of images for a usable hyperspectral scene. Resolving the issues with the geo- and ortho-rectification process would likely eliminate the need for user-guided alignment of images in the mosaicking phase and result in a better end product. This tool was removed entirely in later versions of the software.

Notwithstanding the problems faced using the sensor software, tools in ArcMap made it possible to remedy some of the negative effects of poor geo- and ortho-rectification in the hyperspectral scene. After calculating VIs for the poorly rectified scene in ENVI, ArcMap used user-defined tie points to warp and realign the hyperspectral scene to the spatially accurate RGB orthomosaic. Resolving some of the unfortunate effects of substandard image rectification using ArcMap is what made developing a sampling structure attainable for this study.

It is likely that the primary customer base for Headwall's Nano-Hyperspec sensor value spectral resolution more than spatial resolution, rendering further development of geo- and ortho-rectification processes in their software unnecessary. With many spatially small-scale research studies, the total number of images required for analysis and the amount of terrain relief for the study area is relatively low. This allows insufficiencies in rectification software to have a lesser effect on intermediate and end products. Those of Headwall's customer base who are involved in this type of research are likely more accepting of a loosely integrated Global Navigation Satellite System (GNSS) and inertial measurement unit (IMU) due to a lack of necessity for larger-scale spatial accuracy.

Conclusions

Despite the independent variables included in the highest ranked models for climate metrics having low p values, the highest ranked model for total precipitation was not strong in terms of prediction and had an adjusted R-squared value that was close to zero, indicating a poor fit of the data to the regression line. In contrast, the other regression models with the other climate variables produced much higher adjusted R-squared values, indicating a better fit of the data to the regression line, and less biased residual plots. Moreover, the top ranked regression models for each of the other response variables were the same. This is likely due to the fact that maximum air temperature, average air temperature, and total solar radiation are highly correlated. Therefore, it is expected that changes in these variables would correspond to changes in many of the same VIs.

It should be noted that, for the two-week window by which the mean total precipitation value was calculated, there was very little rainfall. While causation cannot be established to explain this result, it may be that the effect of temperature or the combined effect of temperature and other climate metrics on stress for the trees within the study area (as measured by VI values) is greater than the effect of precipitation on stress. In all, models with the other climate metrics proved stronger than models with total precipitation.

The associations between these VIs and seasonal climate variable means are directly related to the way in which these VIs are calculated; that is, the equations used in their calculation may utilize many of the same spectral regions, some using the same wavelengths. GARI, MTVI2, RENDVI, SAVI, and TDVI appear in the top three ranked models for maximum air temperature, average temperature, and total solar radiation. The equations used to calculate these VIs are listed below:

$$GARI = \frac{NIR - [Green - \gamma(Blue - Red)]}{NIR + [Green - \gamma(Blue - Red)]}$$

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

$$RENDVI = \frac{\rho_{750} - \rho_{705}}{\rho_{750} + \rho_{705}}$$

$$SAVI = \frac{1.5 \cdot (NIR - Red)}{NIR + Red + 0.5}$$

$$TDVI = \sqrt{0.5 + \frac{NIR - Red}{NIR + Red}}$$

In order to understand how RENDVI and TDVI are calculated, we must first understand Difference Vegetation Index (DVI). DVI is the difference between the NIR and either the red or blue wavelength. It can provide a generalized depiction of vegetation health, as NIR wavelengths are reflected much more than visible wavelengths. DVI does well to differentiate vegetation from soil, but it does not consider the effects of atmosphere and shadow on spectral values (Tucker, 1979). NDVI is simply a normalized form of DVI. NDVI is the difference between the NIR and red wavelengths divided by the sum of the NIR and red wavelengths. This normalization permits comparison between collections and at different times. A modified form of NDVI is TDVI. Also a broadband VI, TDVI does not saturate like SAVI and NDVI and has proven useful when vegetation subjacent the canopy is sparse, as in urban landscapes (Bannari et al., 2002). Unlike NDVI, RENDVI uses bands along the red edge rather than reflectance peaks, exploiting senescence, as well as minor transformations in the canopy (Gitelson, 1994; Sims and Gamon, 2002). The equation for calculating GARI shares many similarities to NDVI, except that GARI is less sensitive to atmospheric effects and more sensitive to a wide range of chlorophyll concentrations (Gitelson, 1996).

SAVI is a broadband index similar to NDVI, yet it suppresses the effects of bare soil. SAVI uses an adjustment factor that serves as a proxy to vegetation density. The value of this factor changes with the abundance of vegetation for the study area. A value of 0.5 for the adjustment factor was suggested as suitable for first-order decay soil conditions. The recommended use of this index is in areas where soil is clearly visible through the canopy (Huete, 1988).

To understand MTVI2, we must first understand Triangular Vegetation Index (TVI). TVI is the calculated area of an imaginary triangle whose three vertices are defined by the spectral values of green peak reflectance, minimum chlorophyll absorption, and the NIR shoulder. As photosynthetic activity and the abundance of leaf matter increases, the area of the triangle increases, making TVI good for approximating green LAI, but sensitive to increased chlorophyll activity in conditions of dense canopy (Broge and Leblanc, 2000). The narrowband MTVI replaces the 750nm wavelength in the TVI equation with 800nm, whose reflectance is affected by more by variations in leaf and canopy structures than chlorophyll activity. However, MTVI2, credited as a better predictor of green LAI, accounts for the spectral effects of subjacent soil, preserves LAI sensitivity, and is resistant to fluctuations in chlorophyll activity (Haboudane et al., 2004).

Based on the calculation of six different relative importance metrics, it is evident that TDVI is the primary contributor to the highest ranked model for maximum air temperature, average temperature, and total solar radiation. Only the relative importance metrics *first* and *last* produced contrary results. Even though *first* ranks TDVI as the second-highest contributing predictor, both methods are limited in their inclusion of the direct effect of the predictor, in that *first* only uses the direct effect of the predictor and *last* does not consider the direct effect at all.

Future Recommendations

In consideration of the limitations of the sensor software, it is recommended, at least for studies requiring spectral remote sensing of a moderate- to large- scaled area, that spatial accuracy be given priority over higher spectral resolution. While high-resolution hyperspectral sensor may be useful when the wavelengths of interest are numerous or not known, a multispectral sensor may be sufficiently effective for the purpose of modeling the relationship between seasonal climate variation and several broadband VIs.

Furthermore, the cost of UAV-based multispectral remote sensing is more practical than hyperspectral remote sensing for the majority of researchers. That is not to say that hyperspectral data may not be useful on a large scale, as the high spectral resolution could potentially provide advantage in band selection testing and the differentiation of species, among other applications, but that careful analysis of the quality of the sensor and its software is recommended before choosing the appropriate system. The cohesiveness of all parts of the system is arguably more important than any individual component. However, preference may vary greatly depending on the nature of the study and its research objectives.

Future studies exploring the relationship between seasonal climate events and VIs should seek to improve statistical power by modifying the experimental design in several ways, including but not limited to: (1) increasing the temporal scale of the study, (2) increasing the size of the study area, (3) exploring various sampling structures and validation techniques, (4) incorporating other seasonal climate measures (e.g., total estimated evapotranspiration), and (5) incorporating custom or other stress-related VIs for inclusion in modeling (e.g., NDWI, TVDI, MSI, SRWI, PRI, EVI, etc.) It may also prove useful to develop mixed models, incorporating

both VIs and biophysical properties of the stand. It follows that more studies are required to discern the best VI for detecting a response to seasonal climate stress.

Beneficiaries

The developed model will be used to predict the vegetation response of local hardwoods to water stress under seasonal changes in Southern Missouri. This type of modeling should be of particular interest to local landowners, researchers, and other natural resource organizations such as the Center for Agroforestry at the University of Missouri, the Missouri Department of Conservation (MDC), the United States Forestry Service (USFS), and the United States Department of Agriculture (USDA). While the methodology of processing spectral data for moderate- to large- scale operations has yet to be streamlined for all sensors and applications, the implications of such studies may reveal VIs, or combinations of VIs, that are highly correlated with seasonal climate metrics. In the short term, this type of research could aid relevant parties in detecting plant stress by exploring and using specific models. Even more, this type of research could aid in analyzing the effects of long-term climate change.

REFERENCES

- AgiSoft. (2017). PhotoScan Professional (Software). Available at <http://www.agisoft.com/downloads/installer/>
- Akaike, H. Information Theory and an Extension of the Maximum Likelihood Principle. *Springer Series in Statistics*, 199-213. doi:10.1007/978-1-4612-1694-0_15 Springer-Verlag.
- Bannari, A., Asalhi, H., & Teillet, P.M. (2002) Transformed difference vegetation index (TDVI) for vegetation cover mapping. Toronto, Ontario, Canada: *Institute of Electrical and Electronics Engineers*. doi:10.1109/IGARSS.2002.1026867
- Berni, J., Zarco-Tejada, P.J., Suarez, L. & Fereres, E. (2009). Thermal and Narrowband Multispectral Remote Sensing for Vegetation Monitoring From an Unmanned Aerial Vehicle. *IEEE Transactions on Geoscience and Remote Sensing*, 47(3), 722-738. doi:10.1109/TGRS.2008.2010457
- Bohlin, J., Wallerman, J. & Fransson, J.E.S. (2012). Forest variable estimation using photogrammetric matching of digital aerial images in combination with a high-resolution DEM. *Scandinavian Journal of Forest Research*, 27(7), 692-699. doi:10.1080/02827581.2012.686625
- Boisbunon, A., Canu, S., Fourdrinier, D., et al. (2013). AIC, Cp and estimators of loss for elliptically symmetric distributions. Available at <https://hal.archives-ouvertes.fr/hal-00851206/document>.
- Breidenbach, J. & Astrup, R. (2012). Small area estimation of forest attributes in the Norwegian National Forest Inventory. *European Journal of Forest Research*, 131(4), 1255-1267. doi:10.1007/s10342-012-0596-7
- Broge, N. & Leblanc, E. (2000). Comparing Prediction Power and Stability of Broadband and Hyperspectral Vegetation Indices for Estimation of Green Leaf Area and Canopy Chlorophyll Density. *Remote Sensing of Environment*, 76: 156-172.
- Burkart, A., Cogliati, S., Schickling, A., et al. (2014). A Novel UAV-Based Ultra-Light Weight Spectrometer for Field Spectroscopy. *IEEE Sensors Journal*, 14:62–67.
- Dandois, J., Olano, M., & Ellis, E. (2015). Optimal Altitude, Overlap, and Weather Conditions for Computer Vision UAV Estimates of Forest Structure. *Remote Sensing*, 7(10), 13895-13920. doi:10.3390/rs71013895

- Dash, J., Pearse, G., & Watt, M. (2018). UAV Multispectral Imagery Can Complement Satellite Data for Monitoring Forest Health. *Remote Sensing*, 10(8), 1216. doi:10.3390/rs10081216
- DeWitt, J.D., Warner, T.A., Chirico, P.G., et al. (2017). Creating high-resolution bare-earth digital elevation models (DEMs) from stereo imagery in an area of densely vegetated deciduous forest using combinations of procedures designed for lidar point cloud filtering. *GIScience & Remote Sensing*, 54(4), 552-572. doi:10.1080/15481603.2017.1295514
- ESRI. (2018). ArcGIS Desktop: Release 10, Version 5.1 (Software). Redlands, CA: Environmental Systems Research Institute. Available at <https://www.esri.com/en-us/arcgis/products/index>.
- Exelis Visual Information Solutions (EVIS). (2018). ENVI (Software) Boulder, Colorado: Exelis Visual Information Solutions. Available at <https://www.harrisgeospatial.com/Software-Technology/ENVI>.
- Gini, R., Passoni, D., Pinto, L., et al. (2012). Aerial Images from an UAV System: 3d Modeling and Tree Species Classification in a Park Area. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 39-B1, 361-366. doi:10.5194/isprsarchives-XXXIX-B1-361-2012
- Gitelson, A., and M. Merzlyak. (1994). Spectral Reflectance Changes Associated with Autumn Senescence of *Aesculus Hippocastanum* L. and *Acer Platanoides* L. Leaves. *Journal of Plant Physiology*, 143: 286-292.
- Gitelson, A.A., Kaufman, Y.J. & Merzlyak, M.N. (1996). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment*, 58(3), 289-298. doi:10.1016/S0034-4257(96)00072-7
- Gobakken, T., Bollandsås, O.M., & Næsset, E. (2015). Comparing biophysical forest characteristics estimated from photogrammetric matching of aerial images and airborne laser scanning data. *Scandinavian Journal of Forest Research*, 30(1), 73-86. doi:10.1080/02827581.2014.961954
- Goerndt, M.E., Monleon, V. J., & Temesgen, H. (2010). Relating Forest Attributes with Area- and Tree-Based Light Detection and Ranging Metrics for Western Oregon. *Western Journal of Applied Forestry*, 25(3), 105–111.
- Goerndt, M.E., Monleon, V.J., & Temesgen, H. (2011). A comparison of small-area estimation techniques to estimate selected stand attributes using LiDAR-derived auxiliary variables. *Canadian Journal of Forest Research*, 41(6), 1189-1201. doi:10.1139/x11-033
- Groemping, U. (2006). Relative Importance for Linear Regression in R: The Package relaimpo. *Journal of Statistical Software*, 17(1), 1 - 27. doi:10.18637/jss.v017.i01.

- Haboudane, D. (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. doi:10.1016/j.rse.2003.12.013
- Headwall Photonics. SpectralView. (2018). Bolton, Massachusetts: Headwall Photonics. Available at <http://www.headwallphotonics.com/software>.
- Hruska, R., Mitchell, J., Anderson, M., et al. (2012). Radiometric and Geometric Analysis of Hyperspectral Imagery Acquired from an Unmanned Aerial Vehicle Remote Sensing, 4(9), 2736-2752. doi:10.3390/rs4092736
- Ishimura, A., Shimizu, Y., Rahimzadeh-Bajgiran, P., et al. (2011). Remote sensing of Japanese beech forest decline using an improved Temperature Vegetation Dryness Index (iTVDI) iForest - Biogeosciences and Forestry, 4(5), 195-199. doi:10.3832/ifor0592-004
- Latifi, H., Fassnacht, F. & Koch, B. (2012). Forest structure modeling with combined airborne hyperspectral and LiDAR data Remote Sensing of Environment, 121, 10-25. doi:10.1016/j.rse.2012.01.015
- Lucieer, A., Malenovsky, Z., Veness, T., et al. (2014). HyperUAS-Imaging Spectroscopy from a Multirotor Unmanned Aircraft System. *Journal of Field Robotics*, 31(4), 571-590. doi:10.1002/rob.21508
- Magnussen, S., Næsset, E., Kändler, G., et al. (2016). A functional regression model for inventories supported by aerial laser scanner data or photogrammetric point clouds. *Remote Sensing of Environment*, 184, 496-505. doi:10.1016/j.rse.2016.07.035
- Mallows, C. L. (1973). Some Comments on CP. *Technometrics*. 15 (4): 661–675. doi:10.2307/1267380. JSTOR 1267380.
- Marx, A & Kleinschmit, B. (2017). Sensitivity analysis of RapidEye spectral bands and derived vegetation indices for insect defoliation detection in pure Scots pine stands. *iForest - Biogeosciences and Forestry*, 10(4), 659-668. doi:10.3832/ifor1727-010
- McGaughey, R. (2008). FUSION/LDV: Software for LiDAR data analysis and visualization. *USDA Forest Service*. FUSION v2.65. Available at https://w3.ual.es/GruposInv/ProyectoCostas/FUSION_manual.pdf
- Missouri Historical Agricultural Weather Database (MHAWD). (2018). Commercial Agriculture Automated Weather Station Network. University of Missouri, Extension. Retrieved from http://agebb.missouri.edu/weather/history/index.asp?station_prefix=mtg

- Mohd Razali, S., Marin Atucha, A.A., Nuruddin, A.A., et al. (2016). Monitoring vegetation drought using MODIS remote sensing indices for natural forest and plantation areas. *Journal of Spatial Science*, 61(1), 157-172. doi:10.1080/14498596.2015.1084247
- Moon, H. & Choi, M. (2015). Dryness Indices Based on Remotely Sensed Vegetation and Land Surface Temperature for Evaluating the Soil Moisture Status in Cropland-Forest-Dominant Watersheds. *Terrestrial, Atmospheric and Oceanic Sciences*, 26(5), 599. doi:10.3319/TAO.2015.04.22.01(Hy)
- Nackaerts, K., Delauré, B., Everaerts, J., et al. (2010). Evaluation of a lightweight UAS-prototype for hyperspectral imaging. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 38 (Part 5). Commission V Symposium, Newcastle upon Tyne, UK.
- Narrowband Greenness. (2018). Retrieved from <https://www.harrisgeospatial.com/docs/NarrowbandGreenness.html>
- Nilsson, H.E. (1995). Remote sensing and image analysis in plant pathology Canadian Journal of Plant Pathology, 17(2), 154-166. doi:10.1080/07060669509500707
- Peng, D., Hu, Y. & Li, Z. (2016). Spectral Reflectance and Vegetation Index Changes in Deciduous Forest Foliage Following Tree Removal: Potential for Deforestation Monitoring Journal of Applied Spectroscopy, 83(2), 330-337. doi:10.1007/s10812-016-0291-4
- Puliti, S., Ørka, H., Gobakken, T., et al. (2015). Inventory of Small Forest Areas Using an Unmanned Aerial System. *Remote Sensing*, 7(8), 9632-9654. doi:10.3390/rs70809632
- R Development Core Team. (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2), 461-464. Retrieved from <http://www.jstor.org/stable/2958889>
- Sims, D. & Gamon, J. (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: 337-354. doi.org/10.1016/S0034-4257(02)00010-X
- Sripada, R.P. (2005). Determining In-Season Nitrogen Requirements for Corn Using Aerial Color-Infrared Photography. Ph.D. dissertation. North Carolina State University.
- Tao, W., Lei, Y., & Mooney, P. (2011). Dense point cloud extraction from UAV captured images in forest area. *IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services*, Fuzhou, China, 389-392. doi: 10.1109/ICSDM.2011.5969071

- Tucker, C.J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8(2): 127-150. doi:10.1016/0034-4257(79)90013-0
- Wallace, L., Musk, R., & Lucieer, A. (2014). An Assessment of the Repeatability of Automatic Forest Inventory Metrics Derived From UAV-Borne Laser Scanning Data. *IEEE Transactions on Geoscience and Remote Sensing*, 52(11), 7160-7169. doi:10.1109/TGRS.2014.2308208
- Weber, G. & Jorg, E. (1991). Errors in disease assessment – A survey. *Phytopathology*, 81(10):1238.
- Zarco-Tejada, P.J., González-Dugo, V., & Berni, J.A.J. (2012). Fluorescence, temperature and narrow-band indices acquired from a UAV platform for water stress detection using a micro-hyperspectral imager and a thermal camera *Remote Sensing of Environment*, 117, 322-337. doi:10.1016/j.rse.2011.10.007
- Zhou, Y., Ju, W., Sun, X., et al. (2014). Close relationship between spectral vegetation indices and V_{cmax} in deciduous and mixed forests *Tellus: Series B, Chemical and Physical Meteorology*, 66(0), 1-16. doi:10.3402/tellusb.v66.23279 (2014)

APPENDIX

```
## START OF CODE-----

# FLIGHT 1-----

# Prepare list of text files for flight
setwd("E:/Data/Journagan/9_2017_Stats")
filelist = list.files(pattern = ".*.txt")
dat = lapply(filelist, FUN=read.table)

# Define variables before loop, setting default dataset, column, and counter
values
timelmdat = dat[[1]]
timelmdat = t(timelmdat[,4])
count = 2

# Loops through each text file, appending the data to a new, combined dataset
for(i in 1:104){
newdat = dat[[count]]
newdat = newdat[,4]
newdat = t(newdat)
timelmdat = rbind(timelmdat, newdat)
count = count + 1
}

# Convert matrix created by the for loop back into a data frame
timelmdat = as.data.frame(timelmdat)

# Append Plot and Flight columns
timelmdat$V33=1

# FLIGHT 2-----

# Prepare list of text files for flight
setwd("E:/Data/Journagan/7_2017_Stats")
filelist = list.files(pattern = ".*.txt")
dat = lapply(filelist, FUN=read.table)

# Define variables before loop, setting default dataset, column, and counter
values
time2dat = dat[[1]]
time2dat = t(time2dat[,4])
count = 2

# Loops through each text file, appending the data to a new, combined dataset
for(i in 1:104){
newdat = dat[[count]]
newdat = newdat[,4]
newdat = t(newdat)
time2dat = rbind(time2dat, newdat)
count = count + 1
}
```

```

# Convert matrix created by the for loop back into a data frame
time2dat = as.data.frame(time2dat)

# Append Plot and Flight columns
time2dat$V33=2

# FLIGHT 3-----

# Prepare list of text files for flight
setwd("E:/Data/Journagan/9_2017_Stats")
filelist = list.files(pattern = ".*.txt")
dat = lapply(filelist, FUN=read.table)

# Define variables before loop, setting default dataset, column, and counter
values
time3dat = dat[[1]]
time3dat = t(time3dat[,4])
count = 2

# Loops through each text file, appending the data to a new, combined dataset
for(i in 1:104){
newdat = dat[[count]]
newdat = newdat[,4]
newdat = t(newdat)
time3dat = rbind(time3dat, newdat)
count = count + 1
}

# Convert matrix created by the for loop back into a data frame
time3dat = as.data.frame(time3dat)

# Append Plot and Flight columns
time3dat$V33=3

# APPEND FLIGHTS & DEFINE COLUMN NAMES-----

flightdat = rbind(time1dat, time2dat, time3dat)
colnames(flightdat) =
c("B1", "B2", "B3", "B4", "B5", "B6", "B7", "B8", "B9", "B10", "B11", "B12", "B13", "B14",
"B15", "B16", "B17", "B18", "B19", "B20", "B21", "B22", "B23", "B24", "B25", "B26", "B27",
"B28", "B29", "B30", "B31", "B32", "Flight")

# MERGE FLIGHT CONDITIONS WITH DATA-----

setwd("E:/Data/Journagan/R")
fltcon = read.delim("Flight_Conditions.txt", header = TRUE, sep = "\t", dec =
".")
spectable = merge(flightdat, fltcon, by="Flight")

# REDEFINE FUNCTION IN LEAPS PACKAGE-----

library(leaps)

plot.regsubsets2 =
function(x, labels = obj$xnames, main = NULL, scale = c("bic", "Cp", "adjr2",
"r2"), col = gray(seq(0, 0.9, length = 10)), mar = c(10, 5, 6, 3) + 0.1, ...)
{

```

```

obj = x
lsum = summary(obj)
par(mar = mar)
nmodels = length(lsum$rsq)
np = obj$np
propscale = FALSE
sscale = pmatch(scale[1], c("bic", "Cp", "adjr2", "r2"), nomatch = 0)
if (sscale == 0)
  stop(paste("Unrecognised scale=", scale))
if (propscale)
  stop(paste("Proportional scaling only for probabilities"))
yscale = switch(sscale, lsum$bic, lsum$cp, lsum$adjr2, lsum$rsq)
up = switch(sscale, -1, -1, 1, 1)
index = order(yscale * up)
colorscale = switch(sscale, yscale, yscale, -log(pmax(yscale,
  1e-04)), -log(pmax(yscale, 1e-04)))
image(z = t(ifelse(lsum$which[index, ], colorscale[index],
  NA + max(colorscale) * 1.5)), xaxt = "n", yaxt = "n",
  x = (1:np), y = 1:nmodels, xlab = "", ylab = scale[1],
  col = col)
laspar = par("las")
on.exit(par(las = laspar))
par(las = 2)
axis(1, at = 1:np, labels = labels, ...)
axis(2, at = 1:nmodels, labels = signif(yscale[index], 2), ...)
if (!is.null(main))
  title(main = main)
box()
invisible(NULL)
}

# TOTPRECIP CODE-----

# Perform Exhaustive Subsets Regression
spectable_precip = spectable[, -c(1, 32, 33, 35:37)]
regsub_precip = regsubsets(TotPrecip~., data=spectable_precip, nbest=5,
nvmax=5, method="exhaustive")
summary = summary(regsub_precip)

# MAXAIRTEMP CODE-----

# Perform Exhaustive Subsets Regression
spectable_maxtemp = spectable[, -c(1, 32:34, 36, 37)]
regsub_maxtemp = regsubsets(MaxAirTemp~., data=spectable_maxtemp, nbest=5,
nvmax=5, method="exhaustive")
summary = summary(regsub_maxtemp)

# AVGTEMP CODE-----

# Perform Exhaustive Subsets Regression
spectable_avgtemp = spectable[, -c(1, 32:35, 37)]
regsub_avgtemp = regsubsets(AvgTemp~., data=spectable_avgtemp, nbest=5,
nvmax=5, method="exhaustive")
summary = summary(regsub_avgtemp)

# TOTSOLARRAD CODE-----

```

```

# Perform Exhaustive Subsets Regression
spectable_solrad = spectable[,-c(1,32:36)]
regsub_solrad = regsubsets(TotSolarRad~., data=spectable_solrad, nbest=5,
nvmax=5, method="exhaustive")
summary = summary(regsub_solrad)

# CRITERION BASED MODEL SELECTION-----

#Plot BIC
dev.new()
par(mfrow=c(2,2))
plot.regsubsets2(regsub_precip, scale = "bic", cex.axis=0.75, cex.lab=0.75)
plot.regsubsets2(regsub_maxtemp, scale = "bic", cex.axis=0.75, cex.lab=0.75)
plot.regsubsets2(regsub_avgtemp, scale = "bic", cex.axis=0.75, cex.lab=0.75)
plot.regsubsets2(regsub_solrad, scale = "bic", cex.axis=0.75, cex.lab=0.75)

#Plot AdjR2
dev.new()
par(mfrow=c(2,2))
plot.regsubsets2(regsub_precip, scale = "adjr2", cex.axis=0.75, cex.lab=0.75)
plot.regsubsets2(regsub_maxtemp, scale = "adjr2", cex.axis=0.75,
cex.lab=0.75)
plot.regsubsets2(regsub_avgtemp, scale = "adjr2", cex.axis=0.75,
cex.lab=0.75)
plot.regsubsets2(regsub_solrad, scale = "adjr2", cex.axis=0.75, cex.lab=0.75)

#Plot Mallow CP as proxy to AIC
dev.new()
par(mfrow=c(2,2))
plot.regsubsets2(regsub_precip, scale = "Cp", cex.axis=0.75, cex.lab=0.75)
plot.regsubsets2(regsub_maxtemp, scale = "Cp", cex.axis=0.75, cex.lab=0.75)
plot.regsubsets2(regsub_avgtemp, scale = "Cp", cex.axis=0.75, cex.lab=0.75)
plot.regsubsets2(regsub_solrad, scale = "Cp", cex.axis=0.75, cex.lab=0.75)

# TOTPRECIP MODEL ANALYSIS-----

# Report summary statistics for top 3 ranked models

totprecip_mod1 = lm(TotPrecip ~ B7, data = spectable_precip)
summary(totprecip_mod1)

totprecip_mod2 = lm(TotPrecip ~ B3, data = spectable_precip)
summary(totprecip_mod2)

totprecip_mod3 = lm(TotPrecip ~ B14, data = spectable_precip)
summary(totprecip_mod3)

# MAXAIRTEMP MODEL ANALYSIS-----

# Report summary statistics for top 3 ranked models

maxairtemp_mod1 = lm(MaxAirTemp ~ B6+B16+B20+B24+B27, data =
spectable_maxtemp)
summary(maxairtemp_mod1)

maxairtemp_mod2 = lm(MaxAirTemp ~ B6+B16+B24+B27+B28, data =
spectable_maxtemp)

```

```

summary(maxairtemp_mod2)

maxairtemp_mod3 = lm(MaxAirTemp ~ B6+B16+B24+B27+B29, data =
spectable_maxtemp)
summary(maxairtemp_mod3)

# AVGAIRTEMP MODEL ANALYSIS-----

# Report summary statistics for top 3 ranked models

avgtemp_mod1 = lm(AvgTemp ~ B6+B16+B20+B24+B27, data = spectable_avgtemp)
summary(avgtemp_mod1)

avgtemp_mod2 = lm(AvgTemp ~ B6+B16+B24+B27+B28, data = spectable_avgtemp)
summary(avgtemp_mod2)

avgtemp_mod3 = lm(AvgTemp ~ B6+B16+B24+B27+B29, data = spectable_avgtemp)
summary(avgtemp_mod3)

# TOTSOLRAD MODEL ANALYSIS-----

# Report summary statistics for top 3 ranked models

solrad_mod1 = lm(TotSolarRad ~ B6+B16+B20+B24+B27, data = spectable_solrad)
summary(solrad_mod1)

solrad_mod2 = lm(TotSolarRad ~ B6+B16+B24+B27+B28, data = spectable_solrad)
summary(solrad_mod2)

solrad_mod3 = lm(TotSolarRad ~ B6+B16+B24+B27+B29, data = spectable_solrad)
summary(solrad_mod3)

# REPORT AIC AND BIC VALUES FOR TOP 3 RANKED MODELS-----

AIC(totprecip_mod1, totprecip_mod2, totprecip_mod3)
AIC(maxairtemp_mod1, maxairtemp_mod2, maxairtemp_mod3)
AIC(avgtemp_mod1, avgtemp_mod2, avgtemp_mod3)
AIC(solrad_mod1, solrad_mod2, solrad_mod3)

BIC(totprecip_mod1, totprecip_mod2, totprecip_mod3)
BIC(maxairtemp_mod1, maxairtemp_mod2, maxairtemp_mod3)
BIC(avgtemp_mod1, avgtemp_mod2, avgtemp_mod3)
BIC(solrad_mod1, solrad_mod2, solrad_mod3)

# REPORT RMSE FOR TOP MODELS-----

paste("TotPrecip Mod1 RMSE: ", sqrt(mean(residuals(totprecip_mod1)^2)))
paste("MaxAirTemp Mod1 RMSE:", sqrt(mean(residuals(maxairtemp_mod1)^2)))
paste("AvgAirTemp Mod1 RMSE:", sqrt(mean(residuals(avgtemp_mod1)^2)))
paste("TotSolRad Mod1 RMSE:", sqrt(mean(residuals(solrad_mod1)^2)))

# PLOT REGRESSION AND RESIDUALS-----

dev.new()
par(mfrow=c(2,2))

# --FOR TOTPRECIP--

```



```

plot(predict(totprecip_mod1), residuals(totprecip_mod1), ylab="Residuals", xlab=
"Predicted Values")
abline(0,0, lwd=1, lty=5)

# --FOR MAXAIRTEMP--
plot(predict(maxairtemp_mod1), residuals(maxairtemp_mod1), ylab="Residuals", xla
b="Predicted Values")
abline(0,0, lwd=1, lty=5)

# --FOR AVGTEMP--
plot(predict(avgtemp_mod1), residuals(avgtemp_mod1), ylab="Residuals", xlab="Pre
dicted Values", ylim=c(-1,1))
abline(0,0, lwd=1, lty=5)

# --FOR TOTSOLRAD--
plot(predict(solrad_mod1), residuals(solrad_mod1), ylab="Residuals", xlab="Predi
cted Values")
abline(0,0, lwd=1, lty=5)

# PLOT RELATIVE IMPORTANCE FOR MODELS WITH RELAIMPO PACKAGE-----

library(relaimpo)
# lmg - is the R2 contribution averaged over orderings among predictors
# last - is each variables contribution when included last, also sometimes
called usefulness
# first - is each variables contribution when included first, which is just
the squared covariance between y and the variable
# pratt - is the product of the standardized coefficient and the correlation
# betasq - is the squared standardized coefficient

# --FOR MAXAIRTEMP--
maxtemp_bootstrap = boot.relimp(maxairtemp_mod1, b = 1000, type =
c("lmg", "last", "first", "pratt", "betasq", "pmvd"), rank = TRUE, diff = TRUE,
rela = TRUE)
dev.new()
par(mfrow=c(3,3))
plot(booteval.relimp(maxtemp_bootstrap, sort=TRUE), mai=c(1,1,1,1))

# --FOR AVGTEMP--
avgtemp_bootstrap = boot.relimp(avgtemp_mod1, b = 1000, type =
c("lmg", "last", "first", "pratt", "betasq", "pmvd"), rank = TRUE, diff = TRUE,
rela = TRUE)
dev.new()
par(mfrow=c(3,3))
plot(booteval.relimp(avgtemp_bootstrap, sort=TRUE))

# --FOR TOTSOLRAD--
solrad_bootstrap = boot.relimp(solrad_mod1, b = 1000, type =
c("lmg", "last", "first", "pratt", "betasq", "pmvd"), rank = TRUE, diff = TRUE,
rela = TRUE)
dev.new()
par(mfrow=c(3,3))
plot(booteval.relimp(solrad_bootstrap, sort=TRUE))

# END OF CODE-----

```