https://doi.org/10.5194/bg-2019-194
Preprint. Discussion started: 20 June 2019

Biogeosciences Discussions

TITLE

Comparing Stability in Random Forest Models to Map Northern Great Plains Plant Communities Using 2015 and 2016 Pleiades Imagery

Jameson Brennan[a], Patricia Johnson[a], and Niall Hanan[b]

[a] South Dakota State University West River Agricultural Center 1905 N Plaza Dr. Rapid City, SD 57702

[b] Jornada Basin LTER, New Mexico State University Plant and Environmental Sciences Las Cruces, NM 88003

Corresponding author: Jameson Brennan

Email: Jameson.brennan@sdstate.edu

Second Author email: Patricia.johnson@sdstate.edu

Third Author email: nhanan@ad.nmsu.edu

https://doi.org/10.5194/bg-2019-194
Preprint. Discussion started: 20 June 2019

ABSTRACT

The use of high resolution imagery in remote sensing has the potential to improve understanding of patch level variability in plant structure and community composition that may be lost at coarser scales. Random forest (RF) is a machine learning technique that has gained considerable traction in remote sensing applications due to its ability to produce accurate classifications with highly dimensional data and relatively efficient computing times. The aim of this study was to test the ability of RF to classify five plant communities located both on and off prairie dog towns in mixed grass prairie landscapes of north central South Dakota, and assess the stability of RF models among different years. During 2015 and 2016, Pleiades satellites were tasked to image the study site for a total of five monthly collections each summer (June-October). Training polygons were mapped in 2016 for the five plant communities and used to train separate 2015 and 2016 RF models. The RF models for 2015 and 2016 were highly effective at predicting different vegetation types associated with, and remote from, prairie dog towns (misclassification rates < 5% for each plant community). However, comparisons between the predicted plant community map using the 2015 imagery and one created with the 2016 imagery indicate 6.7% of pixels on-town and 24.3% of pixels off-town changed class membership depending on the year used. Given the low model misclassification error rates, one would assume that low changes in class belonging between years. The results show that while RF models may predict with a high degree of accuracy, overlap of plant communities and inter-annual differences in rainfall may cause instability in fitted models. Researchers should be aware of similarities between target plant communities as well as issues that may arise with using single season or single year images to produce vegetation classification maps.

Keywords

Remote sensing, random forest, rangelands, plant ecology, high resolution imagery,

INTRODUCTION

Remote sensing of rangelands greatly improves our ability to study and understand complex ecological interactions across the landscape. One of the main advantages of remote sensing data is its capacity to cover wide areas, allowing assessment of plant communities at landscape level scales as compared to traditional point-based assessments (Ramoelo et al. 2015; Yu et al. 2018). Numerous studies have demonstrated the utility of remote sensing applications in monitoring rangeland condition, including mapping of vegetation communities, plant species composition, biomass estimation, and impact of grazing intensity on the landscape (Blanco et al. 2008; Franke et al. 2012). Successive images throughout a growing season may potentially help explain patterns of cattle distribution and landscape utilization across temporal scales, or capture phenological changes within the landscape to distinguish differences in warm- and cool-season grass life history, or changes associated with early brown-down in forb- versus grass-dominated communities on prairie dog towns.

Within the Northern Great Plains, black tailed prairie dog colonization is an issue of concern for livestock producers (Miller et al. 2007). Competition between prairie dogs and livestock is a major concern for land managers looking to optimize beef production while still conserving wildlife species (Augustine and Springer 2013). Prairie dogs can reduce availability of forage for livestock by directly reducing the quantity of forage available (through direct consumption, clipping plants to increase predator detection, and building soil mounds), and by

https://doi.org/10.5194/bg-2019-194
Preprint. Discussion started: 20 June 2019

changing species composition (Derner et al. 2006). Older core areas of prairie dog towns often become characterized by extensive areas of bare ground and low vegetation production, which is generally limited to annual forb and dwarf shrub species. Pastures containing extensive areas of bare ground due to prairie dog colonization may potentially depress livestock forage intake rates and ultimately beef production. Understanding the impact of prairie dogs on plant communities, and use patterns of livestock within rangelands occupied by prairie dogs requires the ability to map plant communities at landscape scales.

Advances in remote sensing technology have facilitated the mapping and assessment of a broad range of habitats at different scales (Corbane et al. 2015). For example, Schmidtlein et al. (2007) used hyperspectral imagery at 2m resolution in combination with ordination techniques to map functional plant group gradients in a Bavarian pasture. Within the Delaware Gap National Recreation Area, multiple Landsat 7 scenes were used (30m resolution) with classification tree algorithms to map forest and plant communities for the National Park Service Vegetation Mapping Program (de Colstoun et al. 2003). In Majella National Park, Italy, 4m resolution imagery was used with NDVI to map and predict grass and herbaceous biomass variability over a 200 km^2 area (Cho et al. 2007). While the focus of many of these remote sensing studies is on mapping plant communities at landscape scales to study land use changes and address conservation related issues, the utility of using thematic maps derived from high resolution satellite imagery to study plant and animal interactions has been less explored.

Several methods for accurately classifying plant communities using remote sensing techniques have been used in numerous ecological and natural resource studies. One method, random forest classification (RF), has gained considerable traction in the remote sensing community for its ability to produce accurate classifications, handle highly dimensional data, and

https://doi.org/10.5194/bg-2019-194
Preprint. Discussion started: 20 June 2019

provide efficient computing times (Belgiu and Drăguţ 2016). RF is seen as an improvement over simple classification tree analysis by reducing noise and misclassification of outliers (Laliberte et al. 2007; Nitze et al. 2015). RF is an ensemble decision tree classifier which combines bootstrap sampling to construct several individual decision trees from which a class probability is assigned (Mellor et al. 2013). RF builds each tree using a deterministic algorithm selecting a random set of variables and a random sample from the calibration data set (Ramoelo et al. 2015).

The utility of random forest algorithms has been proven in remote sensing applications. Lowe and Kulkarni (2015) showed that RF outperformed maximum likelihood, support vector machine, and neural network classification models using two Landsat scenes. Ramoelo et al. (2015) successfully used RF modeling to predict leaf nitrogen content using World-View 2 satellite images in grassland and forest communities. Similarly, Mutanga et al. (2012) concluded that RF regression modelling provided an effective methodology for variable selection and predicting biomass in wetland environments. The greatest limitation of the general use of RF has been, and continues to be, due to the lack of off-the-shelf tools for RF implementation within the most common GIS and remote sensing software packages (Hamiton 2013).

Considerable research has focused on the application of RF classification across different plant communities at various scales, however, concerns exist over the transferability of these models to different sites or between seasons. Previous research has shown that RF models have a high degree of classification accuracy at local scales, but model accuracy decreases significantly when applied to spatially separated sites, showing a lack of stability in the model (Juel et al. 2015). Other research has focused on the use of seasonality of image acquisition on improvement of RF models due to spectral differences in plant communities as a result of phenological change during a growing season. Corcoran et al. (2013) showed an improvement

https://doi.org/10.5194/bg-2019-194
Preprint. Discussion started: 20 June 2019

of RF model accuracy in classifying wetlands in northern Minnesota with the inclusion of spring Landsat 5 images across two years over a full season versus summer only, and fall only models.

Many of the plant community classification studies in remote sensing tend to focus on acquiring a single image or multiple images across a single growing season, reducing the influence of inter-annual precipitation on NDVI values (Adjorlolo et al. 2014; Beeri et al. 2007; Guo et al. 2000). Furthermore, most research studies focus solely on spectral differences in plant communities and fail to analyze community differences on the ground at the species level (de Colstoun et al. 2003; Geerken et al. 2005). While classification rates are often reported in studies, the potential overlap in plant community species is rarely explored as a potential source of error within the models. Additionally, very little research has examined how yearly differences in NDVI values across plant communities can alter classification models, especially in high resolution satellite imagery.

We conducted a large, collaborative study from 2012-2016 designed to evaluate livestock production on mixed-grass prairie pastures with varying levels of prairie dog occupation. A major goal of that study was to determine which plant communities on the pastures cattle preferred to graze, and how those preferences shifted within and between years. Plant communities on the site were categorized based on location (on- or off-town) and visually apparent dominant plant functional groups. We expected the plant communities to remain relatively stable during the study, however their signatures on satellite imagery could change within and between years as a result of the timing and magnitude of rainfall and dry periods, timing of green up, phenological progression, and other factors. The overall goal, then, was to develop maps that accurately classify plant communities based on satellite imagery collected between seasons and years. Specific objectives of this study were to 1) determine differences in

https://doi.org/10.5194/bg-2019-194
Preprint. Discussion started: 20 June 2019

the five identified plant communities based on species composition, 2) assess the utility of using a RF model with high resolution satellite imagery to classify plant communities of interest within the Northern Great Plains, and 3) determine the stability of the RF model when using subsequent years of satellite imagery with identical training data. Our ability to map and understand these plant dynamics and patterns at large scales will give researchers insight into applying RF models across years. Research from this study will allow us to better assess how plant communities drive cattle foraging behavior, and evaluate how changes throughout a growing season can cause cattle to shift behavior in response to new resources becoming available.

METHODS

Study site

The study area (45.74N, 100.65W) is located near McLaughlin, South Dakota on a northern mixed-grass prairie ecosystem. Native prairie pastures (810 ha total area) were leased from 2012-2016; pastures were continuously stocked with yearling steers from June-October of each year to achieve 50% utilization. Of the 810 ha, approximately 186 ha were occupied by black-tailed prairie dogs (*Cynomys ludovicianus*). Predominant soils at the site are clays and loams. Ecological sites, and the plant communities they support vary widely; Loamy and Clayey are the predominant Ecological Sites at the site with inclusions of Dense Clay, Shallow Clay, and Thin Claypan (Barth et al. 2014). Plant species dominating the site are largely native, including western wheatgrass (*Pascopyrum smithii* Rydb.), green needlegrass (*Nassella viridula* Trin.), and needle-and-thread (*Hesperostipa comata* Trin. & Rupr), intermixed with blue grama (*Bouteloua gracilis* Willd. Ex Kunth), buffalograss (*Bouteloua dactyloides* Nutt.), and sedges (*Carex* spp.). The most common non-native species on the site is Kentucky bluegrass (*Poa gracilis* Boivin & Love). Woody draws occupy moist drainage areas; vegetation consists primarily of bur oak

https://doi.org/10.5194/bg-2019-194
Preprint. Discussion started: 20 June 2019

(*Quercus macrocarpa* Nutt.), American plum (*Prunus americana* Marshall), and chokecherry (*Prunus virginiana* L.). These draws are frequently flanked by snowberry-dominated patches (*Symphoricarpos occidentalis* Hook.). Plant communities on areas occupied by prairie dog towns on the site are largely dominated by western wheatgrass and shortgrasses (buffalograss, blue grama, and sedges) intermixed with patches of bare ground and annual forb dominated areas. Common annual forbs on prairie dog towns include prostrate knotweed (*Polygonum aviculare* L.), fetid marigold (*Dyssodia papposa* Vent.), dwarf horseweed (*Conyza ramosissima* Cronquist), and scarlet globemallow (*Sphaeralcea coccinea* Nutt.). Mean annual rainfall at the site is 446 mm and average growing season (May through September) temperature is 15.3ºC (Mesonet).

Five plant communities of interest for our study site were identified: 1) Forb-dominated sites on prairie dog towns (PDF), 2) Grass-dominated sites on prairie dog towns (PDG), 3) Snowberry-dominated sites off-town (SNOW), 4) Cool season grass-dominated sites off-town (COOL), and 5) Warm season-dominated sites off-town (WARM).

Training sites

To facilitate classification, training site polygons were mapped for PDF, PDG, COOL, WARM, and SNOW plant communities using ArcPad for Trimble GPS units in the summer of 2016. Twenty training sites were mapped for each of the plant communities except WARM, for which only 8 sites were mapped due to the difficulty of finding homogenous stands of warm season grasses. Plant species in the Northern Great Plains are dominated by cool season species; warm season species, where they occur, are typically intermixed into stands of cool season species. Training sites for each plant community were selected from across the entire study area to capture potential site differences across research pastures. Sites were mapped in the field by

https://doi.org/10.5194/bg-2019-194
Preprint. Discussion started: 20 June 2019

walking the perimeter of the plant community patch with a Trimble GPS unit. Training polygon perimeter boundaries were always at least 3 meters interior of patch edge to minimize error introduced to the training data as a result of GPS signal noise. Identified patches were then converted into a polygon shapefile within ArcGIS to be used as training polygons for the RF classification algorithm. Within each training site polygon, three 0.25 m^2 plots were randomly located. Within each plot, percent cover by species was recorded in the summer of 2016 at the time of polygon mapping.

Plant Community Analysis

Plant community analysis was performed on vegetation data collected from the three 0.25m^2 plots measured in each training polygon. Differences between plant community compositions were determined using a Multi-Response Permutation Procedure (MRPP) with the Sorensen Bray-Curtis distance method. MRPP is a nonparametric procedure used for testing hypotheses between two or more groups (Mitchell et al. 2015). Differences in community compositions were analyzed separately between on-town groups (PD = PDF and PDG) and off-town groups (NPD = COOL, WARM, and SNOW). Although differences between all 5 plant communities are likely to occur, comparisons between on-town and off-town were not made. On-town and off-town sites were mutually exclusive from each other; for example, PDG cannot occur off-town. To analyze trends in species composition between plant community plots, Non-metric Multidimensional Scaling (NMS) ordination was used (Kruskal 1964). Only species that occurred in 3 or more plots were included in the ordination analysis. NMS analysis was conducted using the Sorensen Bray-Curtis distance method with 250 iterations and a stability criterion of 0.00001. Analysis was repeated five times to confirm ordination pattern in the data. Similarity index matrices were generated to compare plot differences between off-town plant

https://doi.org/10.5194/bg-2019-194
Preprint. Discussion started: 20 June 2019

communities and between on-town plant communities and averaged by plant community. All ordination analyses (MRPP and NMS) were performed using PC-ORD 6 software (McCune and Mefford 2002).

Imagery

During the summers of 2015 and 2016, Pleiades satellites were tasked to image the study site. Pleiades satellites, which are members of the SPOT family of satellites, are operated by AIRBUS Defense and Space. This platform was chosen due to its high spatial resolution (0.5 m pan chromatic, 2 m multispectral) and four band spectral resolution: pan chromatic (480-830 nm), red (600-720nm), green (490-610 nm), blue (430-550 nm), and near infrared (750-950 nm). Pleiades satellites were designed for commercial tasking and monitoring, allowing multiple revisits to a project site. A total of ten image collections were acquired in the summer of 2015 and 2016 (five each year) from June through October during the 1st-15th of each month (Table 1). Image collection times were chosen to correspond to the time periods when cattle were actively grazing on the site. Multispectral images were pan-sharpened and orthorectified by the image provider (Apollo Imaging Corp). Boundaries of the prairie dog town were mapped in the fall of 2015 using a handheld Trimble GPS unit. Post collection processing of the images included extracting off-town and on-town locations using the "Extract By Mask" tool in ArcGIS. Separate RF models were developed for on-town and off-town plant communities because such plant communities are mutually exclusive on the site (e.g. PDG cannot exist at off-town locations). Each monthly image collection was converted into an NDVI image using the formula:

$$NDVI = \frac{NIR-Red}{NIR+Red}$$

Random Forest model

https://doi.org/10.5194/bg-2019-194
Preprint. Discussion started: 20 June 2019

For the RF model, the Random Forest package of the Comprehensive R Archive Network (CRAN) implemented by Liaw and Wiener (2002) was utilized. Training data were constructed by stacking all satellite imagery spectral bands (Red, Blue, Green, and NIR) and NDVI bands for each month of each year (25 total dimensions per year) to create a raster stack for each year's imagery (2015 and 2016). To train the model, pixel values were extracted from the satellite imagery raster stack for each training polygon mapped in the field. The random forest models were built using 100 decision trees and default number of nodes at each split, with plant community data as the response category (WARM, COOL, SNOW, PDF, PDG) and spectral band values as the predictor. Models built for comparison include 2015 off-town, 2015 on-town, 2016 off-town, and 2016 on-town. A combined years model was also constructed using all available spectral data from 2015 and 2016 (50 dimensions).

Within the random forest package, Out of Bag (OOB) error rates were calculated by reserving one-third of the training data to test the accuracy of the predictions. Models were then used to predict class belonging for 2015 and 2016 raster stacks and the combined 2015 and 2016 stack. To assess the stability of the RF models from year to year, the "Combinatorial And" tool in ArcGIS was used to create a new raster combining plant community prediction data from 2015 and 2016. The output was used to calculate percent of pixels that were unchanged between the 2015 and 2016 model predictions and percent of change that occurred between years for plant community predictions.

Results and Discussion

MRPP pairwise comparisons were made within on-town communities (PDF vs. PDG) and within off-town communities (COOL vs. WARM vs. SNOW), but not between on- and off-town communities (Table 2). Each plant community was significantly different from all other

https://doi.org/10.5194/bg-2019-194
Preprint. Discussion started: 20 June 2019

communities within its on-town or off-town area (P < 0.001). Substantial differences are evident between off-town plant communities in the 2-D plot of the NMS ordination (final stress = 15.465, instability < 0.00001 after 98 iterations), with some overlap occurring between communities (Figure 1). The On-Town 2-D NMS ordination plot (final stress = 15.591, instability = 0.0005 after 50 iterations) also indicates substantial differences between communities, but with fairly minimal overlap (Figure 1). While there is some overlap between plant communities, in general similarities between plant communities are low, with a similarity index generated from a Sorensen (Bray-Curtis) distance matrix of 21.5 – 27.9% when comparing off-town plant communities and 15.6% when comparing PDF and PDG (Table 2).

Variable importance factor graphs indicate that NDVI training values by month tend to contribute the most to each model for both years, both on- and off-town (Figure 2). Similar results were observed by (Mishra and Crews 2014), where spectral classification features (mean NDVI or ratio NDVI) were the most significant for classifying vegetation morphology in a savanna grassland. Differences between importance of months between years within site is likely the result of interannual precipitation timing between the years, with plant communities greening up or browning down earlier or later depending on seasonal rainfall. Results from the RF model show low OOB misclassification error rates (Table 3) indicating a high degree of accuracy in the model. The lower similarity index (Table 2) for on-town communities compared to off-town communities may help explain the lower OOB classification error rates (Table 3) as well as the lower frequency of pixels changing class in the on-town communities (Table 4). OOB error rate was below 5% for all models. OOB accuracy is an unbiased estimate of the overall classification accuracy eliminating the need for cross-validation (Breiman 2001). Lawrence et al. (2006) showed OOB error rates to be reliable estimates of class accuracy for

https://doi.org/10.5194/bg-2019-194
Preprint. Discussion started: 20 June 2019

identifying invasive species. Similarly, OOB error rates have been reported to be reliable in mapping corn and soybean fields across multiple years (Zhong et al. 2014). Belgiu and Drăguţ (2016) acknowledge that the reliability of OOB error measurements needs to be further tested using a variety of datasets in different scenarios

Consistency in error rates for plant communities appears to indicate stability in the 2015 and 2016 RF models which used identical training sites on consecutive yearly satellite imagery. However, when comparing yearly predicted plant community maps, differences between community classifications are slightly more pronounced, indicating the models may not be as stable as predicted based solely on the OOB error rates. The pixels that were classified as representing one plant community in 2015 and a different one in 2016 were 24.3% of the total off-town pixels and 6.7% of total on-town pixels (Table 4). The pixels changing from COOL to SNOW and SNOW to COOL represented the highest percentage of pixels that changed plant community in off-town areas. COOL and SNOW plant communities, however, occupied the largest area on the site, and represented 70.3 and 21.0% of total pixels in 2015 and 68.5 and 25.1 % of total off-town pixels in 2016, respectively.

It is unlikely in this northern mixed-grass prairie ecosystem that all the changes in plant communities indicated by classification of pixels were real changes from one plant community type to another over one year. Such major shifts in species composition typically occur much more slowly. The results from the plant community analysis indicate training sites were chosen appropriately to account for differences in species composition on the ground, therefore apparent changes are much more likely due to factors that affect the spectral signature of the vegetation. One explanation for the difference in year to year classification could be attributed to the interannual variability of rainfall between 2015 and 2016 (Figure 3). While overall total rainfall between years was similar, differences in timing of precipitation that occurred likely affected timing of green up and dormancy for many of the cool and warm season species on the site. This, then, would create different NDVI patterns between years (Figure 4). Wehlage et al. (2016) for example, found that yearly rainfall differences resulted in large differences in NDVI and biomass measurements across two years in a dry mixed-grass prairie. Goward and Prince (1995) suggested that the relationship between NDVI and annual rainfall in any given year also depends on the previous year history of rainfall at the site, and Oesterheld et al. (2001) showed that annual above ground primary production of shortgrass communities is related to current as well as previous two years precipitation. The above average rainfall at the study site in 2015 could have added to the increase in average NDVI in 2016 when compared to 2015 through an increase in cumulative biomass or production at the site. Another possible cause for changes in plant community classifications between years is overlap of plant community species where two plant communities share a boundary. The edges of plant communities in the NGP are seldom sharp; more often there is a transition zone, where species from each community intermingle. This, along with variability in phenological development of different plants (e.g. cool season vs. warm season) associated with precipitation, as mentioned above, could result in pixels appearing to be associated with one plant community in one year and its neighboring plant community the next. It should also be noted that plant communities in the region, which are predominantly comprised of cool season grasses, often include varying levels of warm season species; and snowberry thickets often have an understory of grasses, especially near the perimeter. Thus one should expect some level of spectral mixing within each community, and the possibility that climatic factors could result in changes in NDVI values that, at least initially, might suggest apparent changes between plant communities.

https://doi.org/10.5194/bg-2019-194
Preprint. Discussion started: 20 June 2019

As noted above, one issue with using categorically classified vegetation maps is that plant communities in space are rarely mutually exclusive, and tend to change along a continuum with environmental gradients (Equihua 1990). Thus, within both on-town and off-town plant communities, transition zones are likely to account for a portion of the classification change between plant communities between years (Figure 5). Alternative approaches to mapping plant communities can be the recognition of fuzzy properties enabling a single point in space to exhibit characteristics of a number of plant communities (Duff et al. 2014; Fisher 2010). For example, Schmidtlein et al. (2007) used NMS of species data in combination with imaging spectroscopy to produce ordination maps of community structure. While fuzzy classification maps are more likely to give a better picture of plant community composition on a per pixel basis, they are also more difficult to use to draw inferences of species dominance and livestock use across landscapes.

A final RF model combining all available bands and NDVI values for 2015 and 2016 reduced error rates for all plant communities below 1% (Table 3). While we have shown that error rates may not result in more stable predictions, using all available data for a model will likely improve accuracy and result in a more accurate thematic map (Figure 6). Zhou et al. (2018) using RF models showed that using a combination of four seasons of Sentinel-1 images and a GaoFen-1 satellite winter image produced the highest classification rate of urban land cover scenes over individual seasonal images. Likewise, several other studies have reported increases in classification accuracy in RF models with the addition of combined seasonal images, hyperspectral data, LiDAR images, radar (SAR) images, and ancillary geographical data such as elevation and soil types (Corcoran et al. 2013; Pu et al. 2018; Shi et al. 2018; Xia et al. 2018; Yu et al. 2018). RF models have the ability to handle highly dimensional correlated data, and data

https://doi.org/10.5194/bg-2019-194
Preprint. Discussion started: 20 June 2019

Biogeosciences Discussions

combined from multiple different data sources across different temporal scales. The internal information provided by the model, such as variable importance, can be a useful tool for researchers to select features of greatest importance to reduce computation times in the instance of large datasets. At the size of our study area (810 ha) and a maximum of 50 variables, the combined 2015-16 data model only slightly added to computation time, but not enough to warrant feature trimming from the dataset. Variable importance plots from the combined data model also indicate that different months between years contribute highly to the classification accuracy between models. For example June 2016 NDVI and October 2015 NDVI were the most important for classification of the data based on the variable importance plot from the combined years' model.

Conclusions

Stability of models is important when applying similar techniques across different sites, plant communities, and in this case years. Differences in year-to-year NDVI values may alter classification results; those differences may be even more pronounced if only one or two satellite imagery scenes are used from a single year. One of the main benefits to RF classification in remote sensing is the relatively fast computing time (Belgiu and Drăguţ 2016), and, given the availability of free satellite imagery, researchers would be prudent to include multiple images across years and seasons in their model to improve accuracy. Furthermore, while the desired outcome is often to produce thematic maps, recognizing that plant communities rarely exist in discrete communities is important when trying to interpret remotely sensed classification maps. This is likely to be magnified as pixel size increases, resulting in less "pure" vegetation structure in the classified pixel. Further work should examine the reliability of OOB error rates across different scenarios, and the influence of year and timing of image acquisition on classification results.

Acknowledgements

We would like to acknowledge and thank the U.S. Department of Agriculture (Grant Number 2011-68004-30052) for funding this research as well as North Dakota State University. We would also like to thank the McLaughlin family for providing access to the land the research was conducted.

References

Adjorlolo, C., Mutanga, O., & Cho, M.A.: Estimation of Canopy Nitrogen Concentration Across C3 and C4 Grasslands Using WorldView-2 Multispectral Data, Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7, 4385-4392, 2014.

Augustine, D.J., & Springer, T.L.: Competition and facilitation between a native and a domestic herbivore: trade-offs between forage quantity and quality, *Ecological Applications*, *23*, 850-863, 2013.

Barth, C.J., Liebig, M.A., Hendrickson, J.R., Sedivec, K.K., & Halvorson, G.: Soil Change Induced by Prairie Dogs across Three Ecological Sites, Soil Science Society of America Journal, 78, 2054-2060, 2014.

Beeri, O., Phillips, R., Hendrickson, J., Frank, A.B., & Kronberg, S.: Estimating forage quantity and quality using aerial hyperspectral imagery for northern mixed-grass prairie, Remote Sensing of Environment, 110, 216-225, 2007.

Belgiu, M., & Drăguţ, L.: Random forest in remote sensing: A review of applications and future directions, Isprs Journal of Photogrammetry and Remote Sensing, 114, 24-31, 2016.

Blanco, L.J., Aguilera, M.O., Paruelo, J.M., & Biurrun, F.N.: Grazing effect on NDVI across an aridity gradient in Argentina, Journal of Arid Environments, 72, 764-776, 2008.

Breiman, L.: Random Forest, Machine Learning, 45, 5-32, 2001.

Cho, M.A., Skidmore, A., Corsi, F., van Wieren, S.E., & Sobhan, I.: Estimation of green grass/herb biomass from airborne hyperspectral imagery using spectral indices and partial least squares regression,
International Journal of Applied Earth Observation and Geoinformation, 9, 414-424, 2007.

Corbane, C., Lang, S., Pipkins, K., Alleaume, S., Deshayes, M., Milian, V.E.G., Strasser, T., Vanden Borre, J., Spanhove, T., & Foerster, M.: Remote sensing for mapping natural habitats and their conservation status - New opportunities and challenges, International Journal of Applied Earth Observation and Geoinformation, 37, 7-16, 2015.

Corcoran, J., Knight, J., & Gallant, A.: Influence of Multi-Source and Multi-Temporal Remotely Sensed and Ancillary Data on the Accuracy of Random Forest Classification of Wetlands in Northern Minnesota, Remote Sensing, 5, 3212, 2013.

de Colstoun, E.C.B., Story, M.H., Thompson, C., Commisso, K., Smith, T.G., & Irons, J.R.: National Park vegetation mapping using multitemporal Landsat 7 data and a decision tree classifier, Remote Sensing of Environment, 85, 316-327, 2003.

Derner, J.D., Detling, J.K., & Antolin, M.F.: Are livestock weight gains affected by black-tailed prairie dogs?, *Frontiers in Ecology and the Environment, 4*, 459-464, 2006.

Duff, T.J., Bell, T.L., & York, A.: Recognising fuzzy vegetation pattern: the spatial prediction of floristically defined fuzzy communities using species distribution modelling methods, Journal of Vegetation Science, 25, 323-337, 2014.

Equihua, M.: Fuzzy Clustering of Ecological Data, Journal of Ecology, 78, 519-534, 1990.

Fisher, P.F.: Remote sensing of land cover classes as type 2 fuzzy sets, Remote Sensing of Environment, 114, 309-321, 2010.

Franke, J., Keuck, V., & Siegert, F.: Assessment of grassland use intensity by remote sensing to support conservation schemes, Journal for Nature Conservation, 20, 125-134, 2012.

Geerken, R., Zaitchik, B., & Evans, J.P.: Classifying rangeland vegetation type and coverage from NDVI time series using Fourier Filtered Cycle Similarity, International Journal of Remote Sensing, 26, 5535-5554, 2005.

Goward, S.N., & Prince, S.D.: Transient Effects of Climate on Vegetation Dynamics: Satellite Observations, Journal of Biogeography, 22, 549-564, 1995.

Guo, X., Price, K.P., & Stiles, J.M.: Modeling Biophysical Factors for Grasslands in Eastern Kansas Using Landsat TM Data, Transactions of the Kansas Academy of Science (1903-), 103, 122-138, 2000.

Hamiton, G.: Random Forest: The Landscape Toolbox, Jornada Experimental Range In: Jornada Experimental Range, 2013.

Juel, A., Groom, G.B., Svenning, J.-C., & Ejrnæs, R.: Spatial application of Random Forest models for fine- scale coastal vegetation classification using object based analysis of aerial orthophoto and DEM data, International Journal of Applied Earth Observation and Geoinformation, 42, 106-114, 2015.

Kruskal, J.B.: Nonmetric multidimensional scaling: A numerical method, Psychometrika, 29, 115-129, 1964.

Laliberte, A.S., Fredrickson, E.L., & Rango, A.: Combining decision trees with hierarchical object-oriented image analysis for mapping arid rangelands, Photogrammetric Engineering and Remote Sensing, 73, 197-207, 2007.

Lawrence, R.L., Wood, S.D., & Sheley, R.L.: Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (randomForest), Remote Sensing of Environment, 100, 356-362, 2006.

Liaw, A., & Wiener, M.: Classification and Regression by Random Forest, R News, 2, 18-22, 2002.

Lowe, B., & Kulkarni, A.: Multispectral Image Analysis Using Random Forest, International Journal on Soft Computing, 6, 1-14, 2015.

McCune, B., & Mefford, M.: PC-ORD for windows, Multivariate analysis of ecological data, Version, 6, 2002.

Mellor, A., Haywood, A., Stone, C., & Jones, S.: The Performance of Random Forests in an Operational Setting for Large Area Sclerophyll Forest Classification, Remote Sensing, 5, 2838-2856, 2013.

Mesonet, S. South Dakota Mesonet Annual Weather Summary. In SDSU (Ed.). http://climate.sdstate.edu/awdn/archive/annualarchive.asp?station=MHT&yr=2015&Submit=Submit

Miller, B.J., Reading, R.P., Biggins, D.E., Detling, J.K., Forrest, S.C., Hoogland, J.L., Javersak, J., Miller, S.D., Proctor, J., Truett, J., & Uresk, D.W.: Prairie dogs: An ecological review and current biopolitics, *Journal of Wildlife Management*, *71*, 2801-2810, 2007.

Mishra, N.B., & Crews, K.A.: Mapping vegetation morphology types in a dry savanna ecosystem: integrating hierarchical object-based image analysis with Random Forest, International Journal of Remote Sensing, 35, 1175-1198, 2014.

Mitchell, S., Palmquist, K., Cohen, S., & Christensen, N.L.: Patterns of vegetation composition and diversity in pine-dominated ecosystems of the Outer Coastal Plain of North Carolina: Implications for ecosystem restoration, Forest Ecology and Management, 356, 64-73, 2015.

Mutanga, O., Adam, E., & Cho, M.A.: High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm, International Journal of Applied Earth Observation and Geoinformation, 18, 399-406, 2012.

https://doi.org/10.5194/bg-2019-194
Preprint. Discussion started: 20 June 2019

Nitze, I., Barrett, B., & Cawkwell, F.: Temporal optimisation of image acquisition for land cover classification with Random Forest and MODIS time-series, International Journal of Applied Earth Observation and Geoinformation, 34, 136-146, 2015.

Oesterheld, M., Loreti, J., Semmartin, M., & Sala, O.E.: Inter-annual variation in primary production of a semi-arid grassland related to previous-year production, Journal of Vegetation Science, 12, 137-142, 2001.

Pu, R.L., Landry, S., & Yu, Q.Y.: Assessing the potential of multi-seasonal high resolution Pleiades satellite imagery for mapping urban tree species, International Journal of Applied Earth Observation and Geoinformation, 71, 144-158, 2018.

Ramoelo, A., Cho, M.A., Mathieu, R., Madonsela, S., van de Kerchove, R., Kaszta, Z., & Wolff, E.: Monitoring grass nutrients and biomass as indicators of rangeland quality and quantity using random forest modelling and World View-2 data, International Journal of Applied Earth Observation and Geoinformation, 43, 43-54, 2015.

Schmidtlein, S., Zimmermann, P., Schüpferling, R., & Weiß, C.: Mapping the floristic continuum: Ordination space position estimated from imaging spectroscopy, Journal of Vegetation Science, 18, 131-140, 2007.

Shi, Y.F., Wang, T.J., Skidmore, A.K., & Heurich, M.: Important LiDAR metrics for discriminating foresttree species in Central Europe, Isprs Journal of Photogrammetry and Remote Sensing, 137, 163-174, 2018.

Wehlage, D.C., Gamon, J.A., Thayer, D., & Hildebrand, D.V.: Interannual Variability in Dry Mixed-Grass Prairie Yield: A Comparison of MODIS, SPOT, and Field Measurements, Remote Sensing, 8, 19, 2016.

Xia, J.S., Yokoya, N., & Iwasaki, A.: Fusion of Hyperspectral and LiDAR Data With a Novel Ensemble Classifier, Ieee Geoscience and Remote Sensing Letters, 15, 957-961, 2018.

Yu, Y., Li, M.Z., & Fu, Y.: Forest type identification by random forest classification combined with SPOT and multitemporal SAR data, Journal of Forestry Research, 29, 1407-1414, 2018.

Zhong, L., Gong, P., & Biging, G.S.: Efficient corn and soybean mapping with temporal extendability: A multi-year experiment using Landsat imagery, Remote Sensing of Environment, 140, 1-13, 2014.

Zhou, T., Zhao, M.F., Sun, C.L., & Pan, J.J.: Exploring the Impact of Seasonality on Urban Land-Cover Mapping Using Multi-Season Sentinel-1A and GF-1 WFV Images in a Subtropical Monsoon-Climate Region, Isprs International Journal of Geo-Information, 7, 2018.

https://doi.org/10.5194/bg-2019-194
Preprint. Discussion started: 20 June 2019

Biogeosciences Discussions
Open Access
EGU

Tables and Figures

Table 1. Acquisition dates of Pleiades satellite imagery tasked for each month (June – October) in 2015 and 2016.

2015 Dates of Acquisition	2016 Dates of Acquisition
6/1/2015	6/5/2016
7/9/2015	7/2/2016
8/4/2015	8/2/2016
9/1/2015	9/11/2016
10/8/2015	10/1/2016

Table 2: Similarity index (Sorensen (Bray-Curtis) distance method) values averaged by plot across plant communities.

Community[1]	Similarity Index (%)
COOL vs. SNOW	27.9
COOL vs. WARM	27.6
SNOW vs. WARM	21.5
PDG vs. PDF	15.6

[1]Plant communities on prairie dog towns are grass-dominated (PDG) and forb-dominated (PDF); plant communities in off-town areas are cool season grass-dominated (COOL), warm season grass-dominated (WARM), and snowberry-dominated (SNOW).

https://doi.org/10.5194/bg-2019-194
Preprint. Discussion started: 20 June 2019

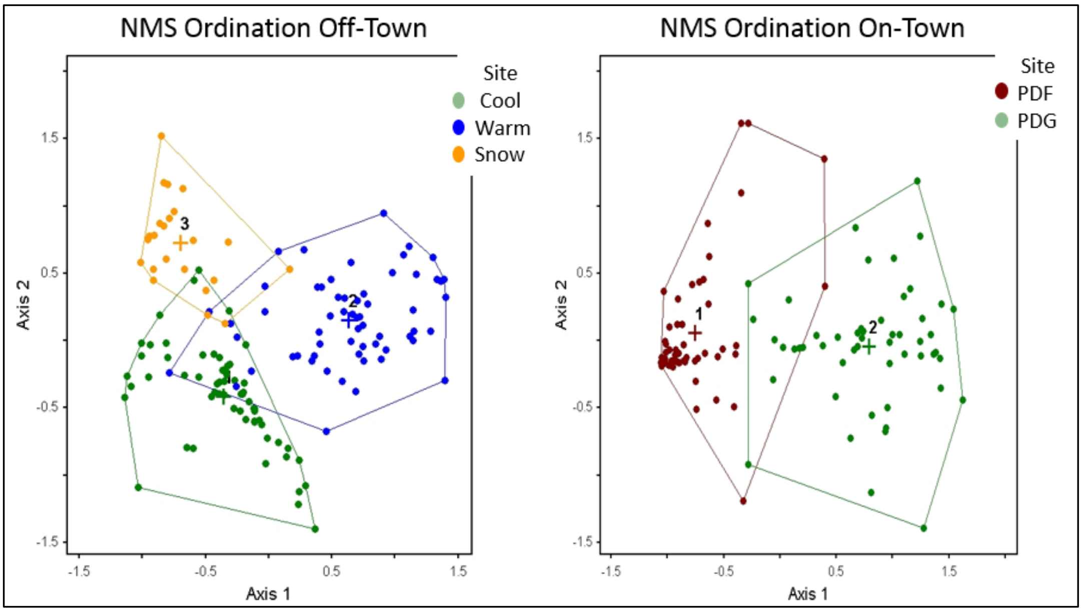

Figure 1. NMS ordination plots for plant communities located on and off of prairie dog towns, based on plant cover by species data collected in 2016 on the study site in north central South Dakota. Plant communities on prairie dog towns are grass-dominated (PDG) and forb-dominated (PDF); plant communities in off-town areas are cool season grass-dominated (COOL), warm season grass-dominated (WARM), and snowberry-dominated (SNOW).

https://doi.org/10.5194/bg-2019-194
Preprint. Discussion started: 20 June 2019

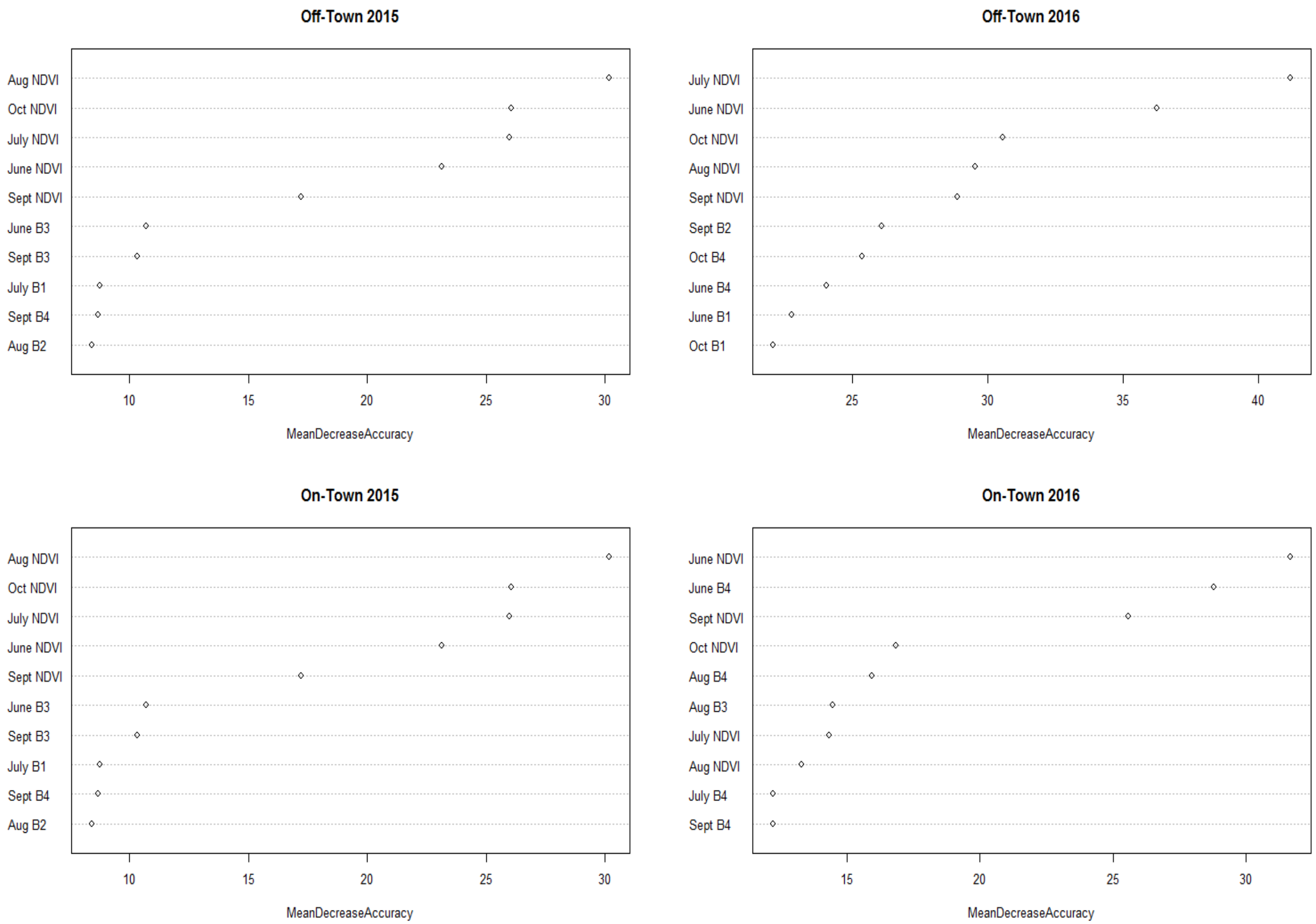

Figure 2: Variable importance reported as mean decrease in accuracy. Ten most important variables are shown, with B1 and B4 corresponding to spectral bands 1 and 4 respectively from Pleiades image. Variable importance is determined by the model output as the decrease in accuracy due to the exclusion of that variable during the out of bag error calculation process. Higher mean decrease in accuracy variables are more important in classifying the data.

https://doi.org/10.5194/bg-2019-194
Preprint. Discussion started: 20 June 2019

Biogeosciences Discussions
Open Access EGU

Table 3: Out of Bag misclassification error rates (%) for each plant community for 2015, 2016, and combined year random forest models.

Plant Community[1]	2015 Model	2016 Model	2015-2016 Combined Model
COOL	0.20%	0.20%	0.03%
SNOW	2%	2%	0.60%
WARM	3%	5%	0.70%
PDG	0.30%	0.20%	0.07%
PDF	0.90%	0.70%	0.30%

[1] Plant communities on prairie dog towns are grass-dominated (PDG) and forb-dominated (PDF); plant communities in off-town areas are cool season grass-dominated (COOL), warm season grass-dominated (WARM), and snowberry-dominated (SNOW).

Table 4: Percent of pixels within each area (prairie dog town and off-town) for each plant community that remain unchanged and are changed between class belonging between 2015 and 2016 models.

Community Location	Transitions[1]	Percent of Total Area Pixels
Prairie Dog Town	Unchanged Pixels	93.3
	PDG ↔ PDF	6.7
Off-Town	Unchanged Pixels	75.7
	COOL ↔ SNOW	14.1
	COOL ↔ WARM	6.7
	SNOW ↔ WARM	3.5

[1]Plant communities on prairie dog towns are grass-dominated (PDG) and forb-dominated (PDF); plant communities in off-town areas are cool season grass-dominated (COOL), warm season grass-dominated (WARM), and snowberry-dominated (SNOW).

https://doi.org/10.5194/bg-2019-194
Preprint. Discussion started: 20 June 2019

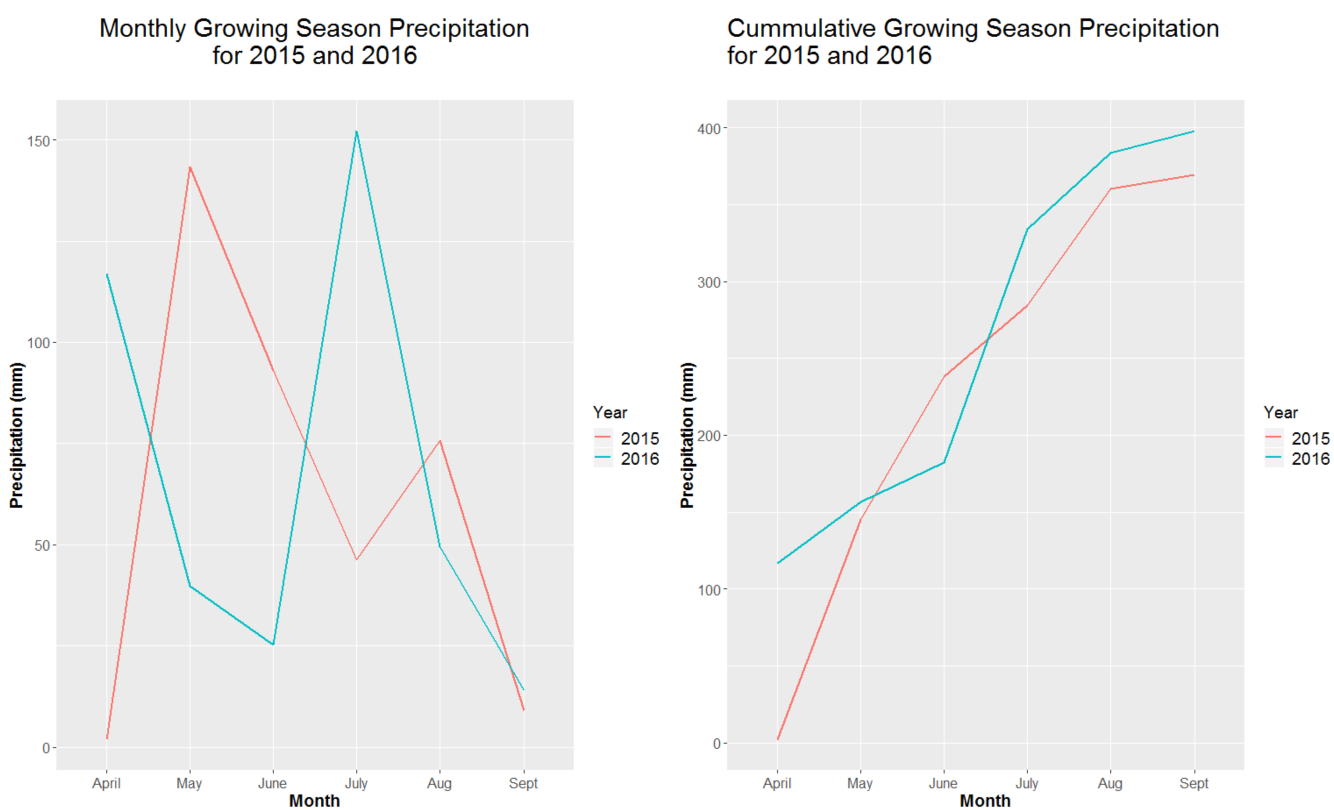

Figure 3: Monthly and cumulative growing season precipitation patterns for 2015 and 2016 recorded at a weather station located on the study area in north central SD (45.737296 N, -100.657540 W)(South Dakota Mesonet 2018).

https://doi.org/10.5194/bg-2019-194
Preprint. Discussion started: 20 June 2019

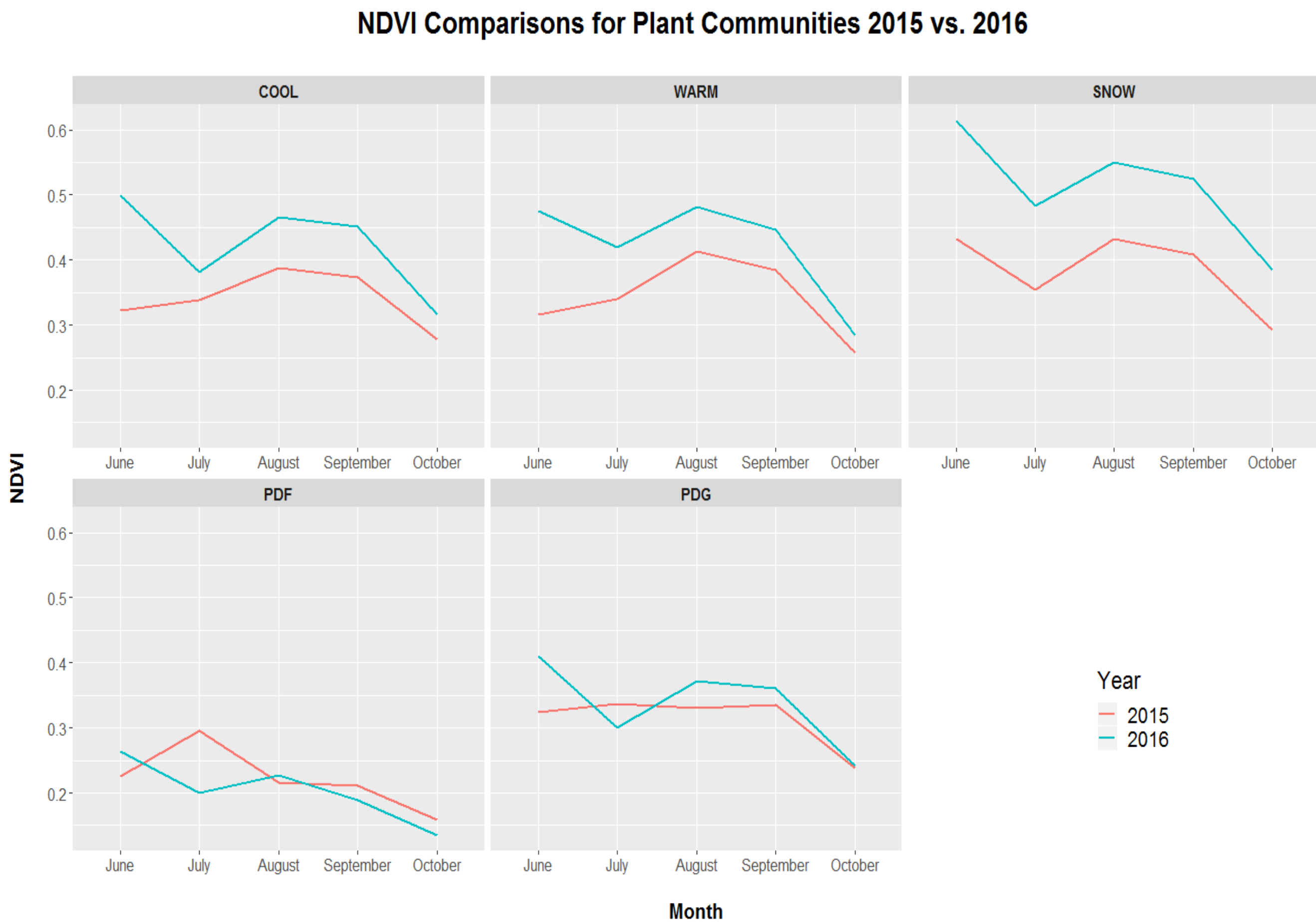

Figure 4: Comparison of mean monthly NDVI for training polygons in five plant communities on the study site in north central SD. Plant communities on prairie dog towns are grass-dominated (PDG) and forb-dominated (PDF); plant communities in off-town areas are cool season grass-dominated (COOL), warm season grass-dominated (WARM), and snowberry-dominated (SNOW).

https://doi.org/10.5194/bg-2019-194
Preprint. Discussion started: 20 June 2019

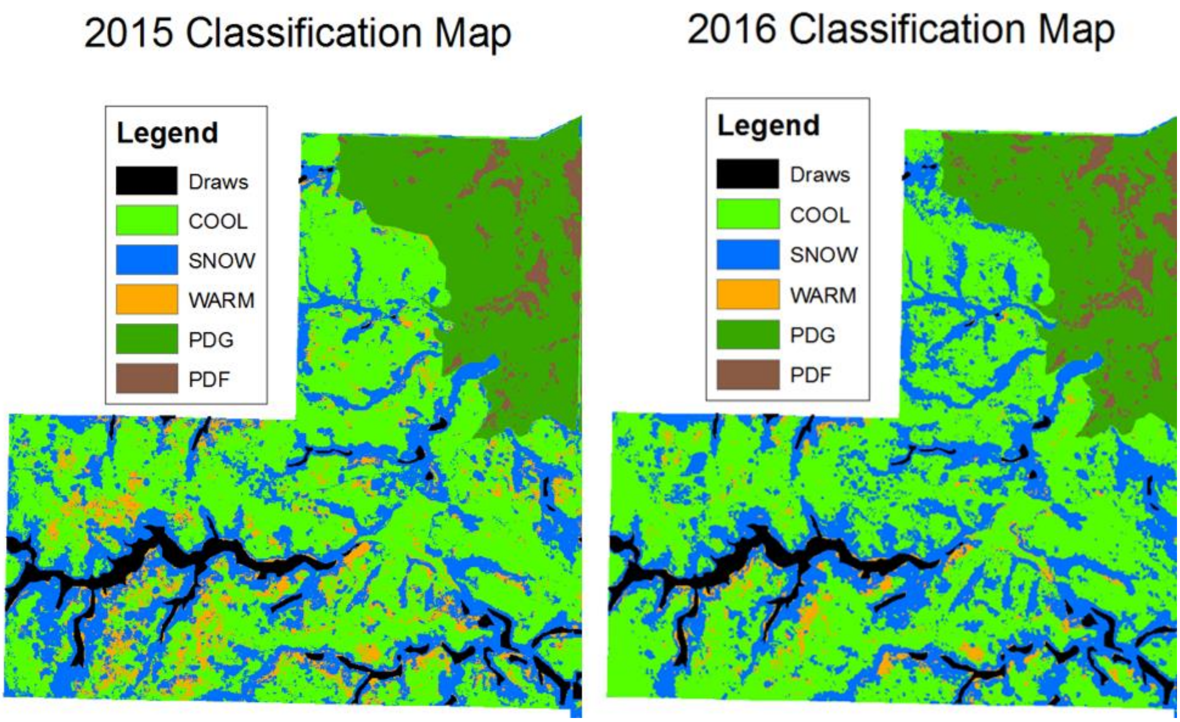

Figure 5: Random forest classification maps from 2015 and 2016 of one pasture in the study area in north central South Dakota. Plant communities on prairie dog towns are grass-dominated (PDG) and forb-dominated (PDF); plant communities in off-town areas are cool season grass-dominated (COOL), warm season grass-dominated (WARM), and snowberry-dominated (SNOW).

https://doi.org/10.5194/bg-2019-194
Preprint. Discussion started: 20 June 2019

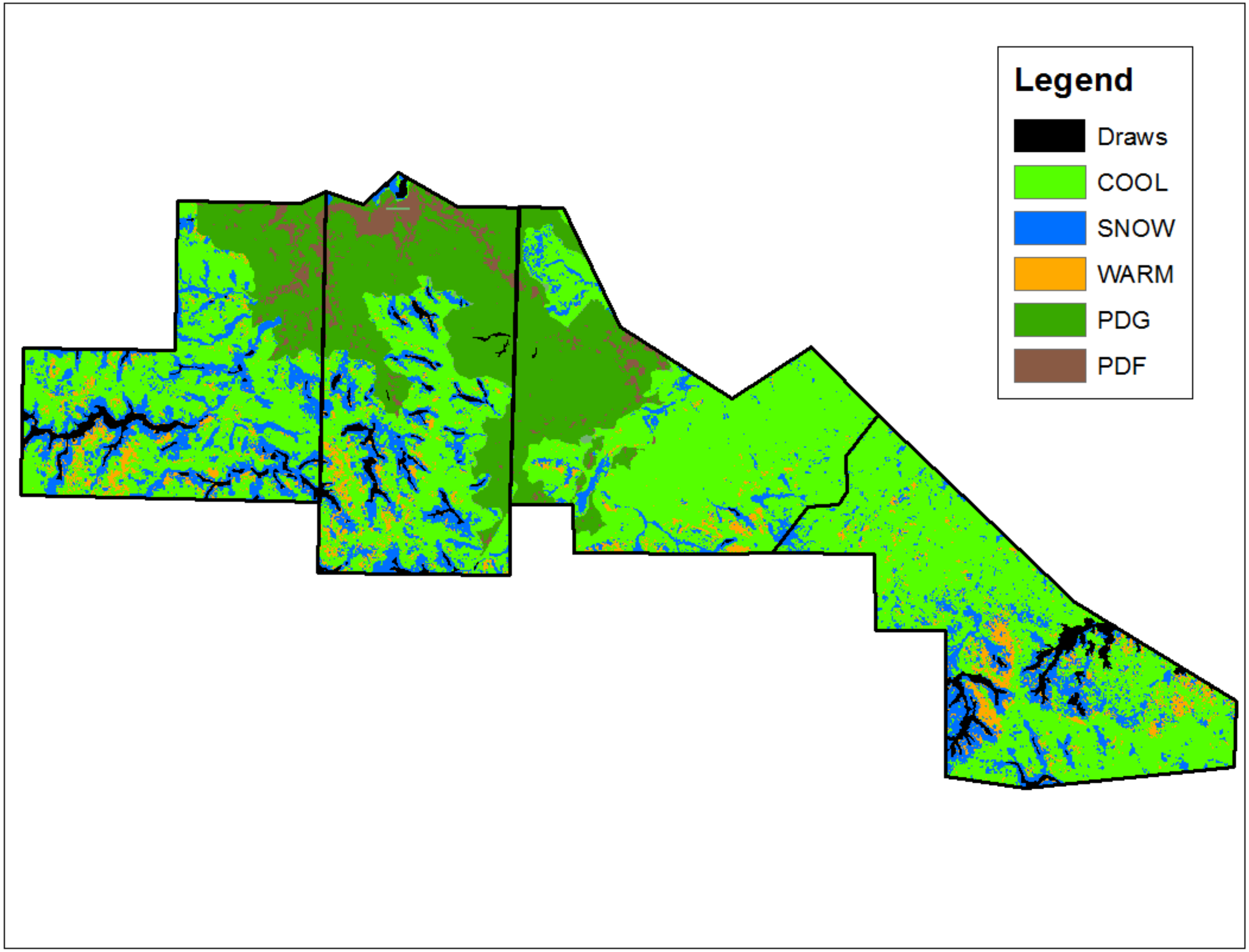

Figure 6: Final random forest generated thematic map of the entire study site in north central South Dakota produced from the combined 2015-2016 imagery data. Plant communities on prairie dog towns are grass-dominated (PDG) and forb-dominated (PDF); plant communities in off-town areas are cool season grass-dominated (COOL), warm season grass-dominated (WARM), and snowberry-dominated (SNOW).