TITLE

Comparing Stability in Random Forest Models to Map Northern Great Plains Plant Communities in Pastures Occupied by Prairie Dogs Using 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

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, assess the stability of RF models among different years, and determine the utility of utilizing remote sensing techniques to identity prairie dog colony extent. 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 RF models. Both the 2015 and 2016 RF models had low (1%) out of bag error rates. However, comparisons between the predicted plant community maps using the 2015 imagery and one created with the 2016 imagery indicate over 32.9% of pixels changed plant community class between 2015 and 2016. 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. A final RF model combining both 2015 and 2016 data yielded the lowest error rates, and was also highly accurate in determining prairie dog colony boundaries.

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. As technology advances, monitoring of rangelands via remote sensing platforms will facilitate research products freely available to land managers (Browning et al. 2015). 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). Additionally, successive images throughout a growing season may potentially capture phenological changes associated with differences in C-3 and C-4 plant species composition (Goodin and Henebry 1997).

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 normalized difference vegetation index (NDVI) to map and predict grass and herbaceous biomass variability over a 200 km^2 area (Cho et al. 2007).

Many 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 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 demonstrated in remote sensing applications at multiple scales. Lowe and Kulkarni (2015) showed that RF was effective at producing highly accurate classification maps using two Landsat scenes (30m resolution). Ramoelo et al. (2015) successfully used RF modeling to predict leaf nitrogen content using World-View 2 satellite images (2m resolution) 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 using high resolution satellite imagery (2m).

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, across seasons, or years. For example, Juel et al. 2015 showed that RF models have a high degree of classification accuracy for classifying fine scale coastal vegetation using digital elevation maps and high resolution orthophoto imagery, but model accuracy decreased significantly when applied to spatially separated sites, showing a lack of stability in the model. Corcoran et al. (2013) showed an improvement of RF model accuracy in classifying wetlands in northern Minnesota with the inclusion Landsat 5 images across two years using full season data versus summer only, and fall only models. Jones et al. 2018 demonstrated the utility of using RF modeling with 30m Landsat data to monitor rangeland cover across the western United States over a 33 year period. Results of these studies suggest the scale and seasonality of the imagery may play an important role in the stability of RF models to accurately map plant communities. While classification rates are often reported in studies, the potential overlap in plant community composition is rarely explored as a potential source of error within the models. 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). Lastly, 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, very little research has examined the impacts of animal species on plant community composition, and how this might affect classification accuracy.

Within the mixed grass prairie ecosystem, 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 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. The ability to accurately map prairie dog colonies using remote sensing will help improve our understanding of the impact of prairie dogs on plant communities, and help inform land management decisions within rangelands occupied by prairie dogs.

A large collaborative study from 2012-2016 was conducted to evaluate livestock production on mixed-grass prairie pastures with varying levels of prairie dog occupation. A major goal of the larger 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. Thus, plant community as defined for this study was a collection of species within an area of a relatively uniform composition different from neighboring patches. Differences in neighboring patches were evident by differences in dominant functional group (forb vs grass) or differences in photosynthetic pathways (C3 vs C4 grasses). 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, timing of green up, phenological progression, and other factors. The overall goal of this paper, then, was to develop maps that accurately classify plant communities based on satellite imagery collected between years. Specific objectives of this study were to 1) determine differences in 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 a mixed grass prairie ecosystem containing prairie dogs, 3) determine the stability of the RF model when using subsequent years of satellite imagery with identical training data, and 4) determine the ability of high resolution satellite imagery to accurately classify prairie dog towns. Our ability to map and understand these plant communities' at large scales will give researchers insight into applying RF models across years using high resolution imagery. Research from this study will allow us to better assess how prairie dogs drive changes in plant communities, and provide a new tool to map the extent of prairie dog colonization on the landscape to better inform land management decisions.

METHODS

Study site

The study area (45.74N, 100.65W) was 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 were clays and loams. Ecological sites, and the plant communities they support vary widely; Loamy and Clayey were 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 were 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 was Kentucky bluegrass (*Poa pratensis* Boivin & Love). Woody draws occupied moist drainage areas; vegetation consists primarily of bur oak (*Quercus macrocarpa* Nutt.), American plum (*Prunus americana* Marshall), and chokecherry (*Prunus virginiana* L.). These draws were frequently flanked by snowberry-dominated patches (*Symphoricarpos occidentalis* Hook.). Plant communities on areas occupied by prairie dog towns on the site were 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 included prostrate knotweed (*Polygonum aviculare* L.), fetid marigold (*Dyssodia papposa* Vent.), dwarf horseweed (*Conyza ramosissima* Cronquist), and scarlet globemallow (*Sphaeralcea coccinea* Nutt.). A weather station has been maintained on site from May 2013 operated by South Dakota Mesonet. Mean annual rainfall at the site is 446 mm and average growing season (May through September) temperature is 15.3ºC (South Dakota Climate and Weather 2017).

Five plant communities of interest for our study site were identified: 1) Forb-dominated sites on prairie dog towns (On-Forb), 2) Grass-dominated sites on prairie dog towns (On-Grass), 3) Snowberry-dominated sites off-town (Off-Snow), 4) Cool season grass-dominated sites off-town (Off-Cool), and 5) Warm season-dominated sites off-town (Off-Warm). An additional plant community labeled 'Draws' was delineated visually within ArcGIS software due to difficulty in mapping these areas in the field. As mentioned prior, these areas are dominated by bur oak, chokecherry, and American plum, and occupied lower lying drainage areas on the site.

Training sites

To facilitate classification, training site polygons were mapped for On-Forb, On-Grass, Off-Cool, Off-Warm, and Off-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 Off-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 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 by tossing plot frames into the area of interest to determine sampling area. 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 for all plant communities, and pairwise comparisons generated. 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 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). Each monthly image collection was converted into an NDVI image. Areas delineated as Draws were removed from the analysis area. In addition, boundaries of the prairie dog town were mapped using a handheld Trimble GPS unit to compare predicted colony location with ground truth location.

Random Forest model

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 200 decision trees and default number of nodes at each split (sqrt(n)), with plant community data as the response category (On-Grass, On-Forb, Off-Cool, Off-Warm, and Off-Snow) and spectral band values as the predictor. Models were checked for error stabilization, for all models error rates stabilized around 50 trees. Yearly models (2015 and 2016) were built for output comparison. 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 using the ‘predict’ function within program R. To assess the stability of the RF models from year to year, the “crosstab” function in the raster package in program R was used to calculate the number of pixels that changed class from 2015 to 2016. The output was used to calculate percent of pixels that were unchanged from 2015 to 2016 model predictions and percent of pixel change that occurred between years for plant community predictions.

Results and Discussion

Plant Community

MRPP pairwise comparisons results showed a significant difference between all plant communities ($P < 0.001$). Differences are evident between plant communities in the 2-D plot of the NMS ordination (final stress = 20.01, instability < 0.00001 after 66 iterations), with some overlap occurring between communities (Figure 1). Plant communities on-town and off-town are clustered at opposite ends of the ordination plot, with the greatest distance being between On-Forb and Off-Snow. Archer et al. (1987) showed in a detrended correspondence analysis of plant communities ranging from uncolonized, 2 years post colonization, and 4-6 years post colonization, that uncolonized sites were clustered at one extreme and the 4-6 year sites at the other extreme. Interestingly, Off-Warm and On-Grass communities are clustered closer in ordination space. Plant communities shifts on-town towards those dominated by shortgrass species have been documented (Agnew et al. 1986; Koford 1958), and is probably attributable to the high grazing resistance of the C4 species blue grama and buffalograss (Derner et al. 2006). While there is some overlap between plant communities, in general similarities between plant communities are low (< 29%), with the greatest similarity index differences generated from a Sorensen (Bray-Curtis) distance matrix occurring between the On-Forb communities and the off-town communities (Table 2). Based on how plant communities were selected in this study, we expected plant community composition to be distinct between groups. Though plant communities are defined by dominant functional group in this study, the amount of overlap occurring demonstrates that other functional groups and species exist within these distinct patches, which may be a potential source of error in classification models.

Random Forest Model Results

Results from the RF model show low OOB misclassification error rates for each individual plant community (Table 3) indicating a high degree of accuracy in the model. Overall the OOB model error rates were 0.9% and 1.12% for the 2015 and 2016 model respectively. 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 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. Overall a total of 67.04% pixels remained unchanged in their plant community classification from 2015 to 2016 (Table 4). The pixels changing from On-Grass to Off-Cool represented the highest percentage of pixels that changed plant community classification at 7.28%. Johnson-Nistler et al. (2004) observed up to 7 times more standing dead forage present on uncolonized sites compared to colonized areas, mainly attributed to prairie dogs clipping vegetation which greatly reduced the amount of grasses that reached maturity. Areas either less maintained on-town or grazed by cattle repeatable off-town may show up similar spectral signatures. Additionally, On-Grass and Off-Cool plant communities have western wheatgrass as a dominant species, and similarity in species dominance between these communities may explain yearly shifts in predictions. Of the pixels that changed classification between years, 15.13 were on-town to off-town transitions, 2.26 were on-town to on-town transitions, and 15.57 were off-town to off-town plant community transitions.

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. In the absence of a major disturbance event, 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 2). 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 3). 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.

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 4). 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. The variable importance graph of the combined model indicates that NDVI variables contribute the most to the model over individual bands (Figure 5). 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. The variable importance plot from the combined data model also indicates that different months between years contribute highly to the classification accuracy. Of the ten most important variables in the model, 6 were from 2015 and 4 from 2016, suggesting additional data in the model is likely to yield greater classification accuracy. Zhou et al. (2018) using RF models showed that 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 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.

Remote Sensing Prairie Dog Colonies

Visual comparison of the predicted on-town plant communities versus off-town plant communities show a clearly defined boundary between areas colonized by prairie dogs and areas not colonized (Figure 6). Previous research has demonstrated that colonization by prairie dogs and subsequent increases in grazing pressure can result in significant differences between on-and off-town plant community composition and production (Coppock et al., 1983; Winter et al. 2002; Johnson-Nistler et al. 2004; Geaumont et al. 2019). The results of our study demonstrate that these differences are significant enough to be identified using remote sensing techniques. Interestingly, a considerable portion of the area misclassified as on-town is from a previously colonized area that had been poisoned in 2013, suggesting that, at least spectrally, these areas still resemble plant communities similar to those actively colonized. Additionally, most other pixels misclassified as on-town are likely drainage areas with high bare ground off-town, whose variability was not captured in the dataset. Results from mapping colony boundaries with a hand held GPS device estimated the colony to be 276 ha in 2012 to 186 ha in 2015. Total colony acreage estimated from summing the pixel area occupied by the On-Grass and On-Forb community pixels from the combined 2015-2016 RF model was 246 ha. As mentioned prior, the higher area estimate in from the RF model is likely the result of transition areas controlled two years prior. One prior study had sought identify prairie dog colonies using 30m Landsat imagery, however concluded that the scale was too course for accurately measuring prairie dog towns (Wolbrink et al. 2002). High resolution satellite imagery used in this study appears capable at capturing fine scale transitions that occur between plant communities along the on-town off-town gradient.

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, and the addition of two years' worth of data likely resulted in improved classification accuracy. 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 decreases, resulting in fine scale predictions which may be more susceptible to plant community transitions. Lastly, results from this study indicate that plant community changes induced by prairie dogs are significant enough to be detected via remote sensing techniques. Land managers looking to optimize rangeland health on pastures occupied by prairie dogs may potentially utilize high resolution imagery to monitor colony size and make recommendations of appropriate stocking rates based on extent of colonization.

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

Agnew, W.D., W. Uresk, D., & Hansen, M.R.: Flora and Fauna Associated with Prairie Dog Colonies and Adjacent Ungrazed Mixed-Grass Prairie in Western South Dakota, Journal of Range Management, *39*, 135-139, 1986.

Archer, S., Garrett, M.G., & Detling, J.K.: Rates of Vegetation Change Associated with Prairie Dog (Cynomys ludovicianus) Grazing in North American Mixed-Grass Prairie, *Vegetatio, 72,* 159-166, 1987.

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.

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.

Browning, D. M., A. Rango, J. W. Karl, C. M. Laney, E. R. Vivoni, & Tweedie, C. E.: Emerging technological and cultural shifts advancing drylands research and management, Frontiers in Ecology and the Environment, 13,52-60, 2015.

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.

Coppock, D.L., Ellis J. E., Detling J. K., & Dyer, M.I.: Plant-herbivore interactions in a North American mixed-grass prairie: Responses of bison to modification of vegetation by prairie dogs, Oecologia, 56,10-15,1983.

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.

Geaumont, B.A., Hovick, T.J., Limb, R.F., Mack, W.M., Lipinski, A.R., & Sedivec, K.K.: Plant and Bird Community Dynamics in Mixed-Grass Prairie Grazed by Native and Domestic Herbivores, Rangeland Ecology & Management, 72, 374-384, 2019.

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.

Goodin, D. G., & Henebry, G. M.: A technique for monitoring ecological disturbance in tallgrass prairie using seasonal NDVI trajectories and a discriminant function mixture model, Remote Sensing of Environment, 61, 270-278, 1997.

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

Johnson-Nistler, C. M., B. F. Sowell, H. W. Sherwood, & Wambolt, C. L.: Black-tailed prairie dog effects on Montana's mixed-grass prairie, Journal of Range Management, 57,641-648, 2004.

Jones, M. O., B. W. Allred, D. E. Naugle, J. D. Maestas, P. Donnelly, L. J. Metz, J. Karl, R. Smith, B. Bestelmeyer, C. Boyd, J. D. Kerby, & McIver, J. D.: Innovation in rangeland monitoring: annual, 30 m, plant functional type percent cover maps for U.S. rangelands, 1984–2017, Ecosphere 9 (9), 2018.

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.

Koford, C.B.: Prairie dogs, whitefaces, and blue grama, Wildlife Society, 1958.

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.

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.

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.

South Dakota Climate and Weather, South Dakota Climate and Weather Mesonet at SDSTATE, Available at: https://climate.sdstate.edu/archive/, Accessed 1st Dec 2017

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.

Winter, S. L., J. F. Cully, & Pontius, J. S.: Vegetation of prairie dog colonies and non-colonized short-grass prairie, Journal of Range Management, 55, 502-508, 2002.

Wolbrink, G. A., Jenks, J. A., Hubbard D. E., & Klaver, R. W.: Feasibility of using Landsat satellite imagery to monitor black-tailed prairie dog towns in western South Dakota, Final report to South Dakota Department of Game, Fish and Parks, 2002.

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.

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 commuinities.

Community Comparison[1]	Similiarity Index (%)
Off-Cool vs. Off-Snow	28.2
Off-Cool vs. Off-Warm	27.8
Off-Cool vs. On-PDG	27.7
Off-Snow vs. Off-Warm	21.6
On-PDG vs. On-PDF	17.8
Off-Snow vs. On-PDG	17.3
Off-Warm vs. On-PDG	17.3
Off-Cool vs. On-PDF	7.9
Off-Snow vs. On-PDF	6.2
Off-Warm vs. On-PDF	6.2

[1] Plant communities on prairie dog towns are grass-dominated (On-Grass) and forb-dominated (On-Forb); plant communities in off-town areas are cool season grass-dominated (Off-Cool), warm season grass-dominated (Off-Warm), and snowberry-dominated (Off-Snow).

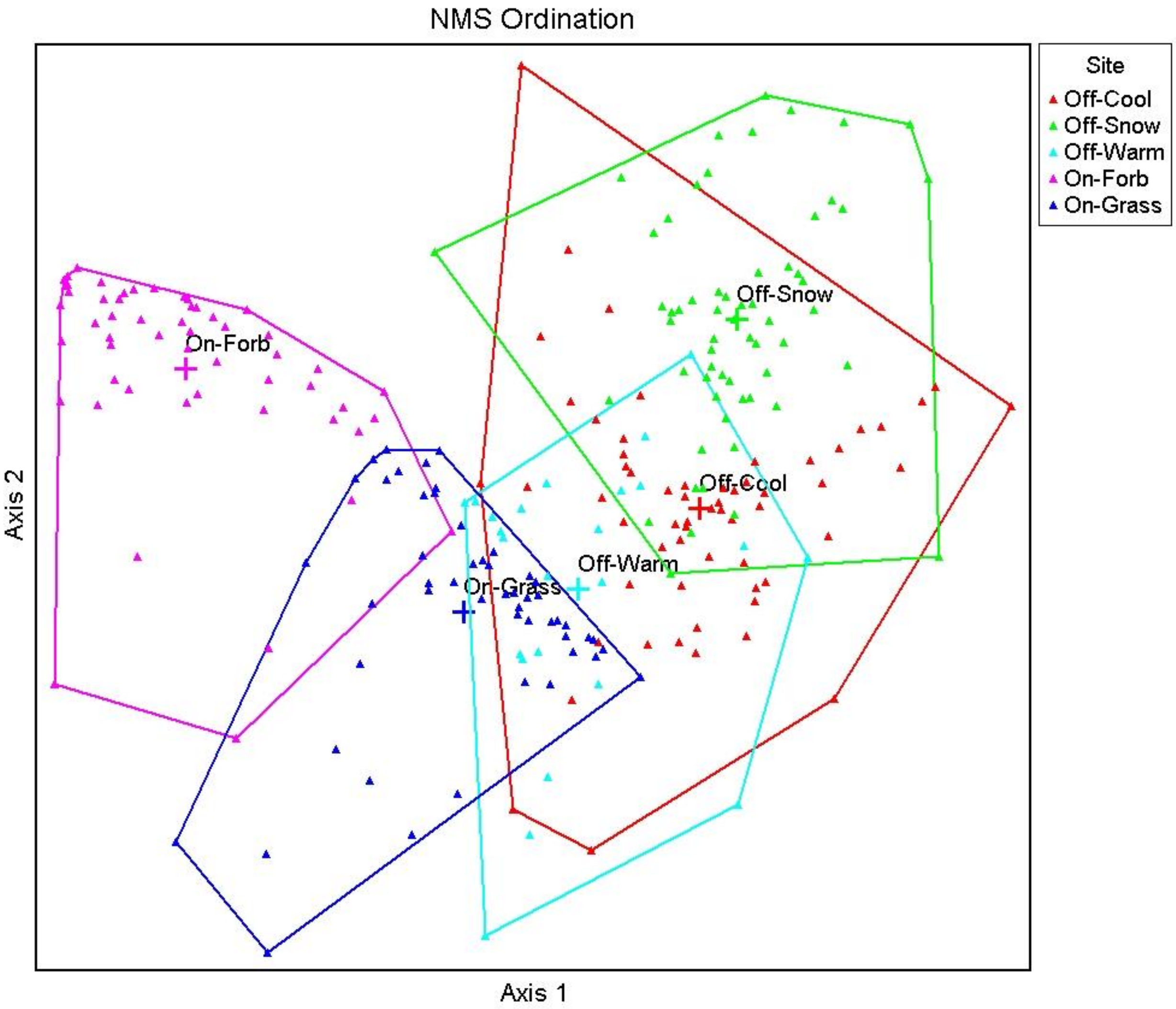

Figure 1. NMS ordination plots for plant communities located on and off of prairie dog towns, based on plant cover by species data collected in 2016 on the study site in north central South Dakota. Plant communities on prairie dog towns are grass-dominated (On-Grass) and forb-dominated (On-Forb); plant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

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
Off-Cool	0.20%	0.40%	0.04%
Off-Snow	2.2%	1.9%	0.69%
Off-Warm	3.2%	5.3%	0.73%
On-Grass	0.40%	0.60%	0.09%
On-Forb	0.60%	0.70%	0.19%

[1] Plant communities on prairie dog towns are grass-dominated (On-Grass) and forb-dominated (On-Forb); plant communities in off-town areas are cool season grass-dominated (Off-Cool), warm season grass-dominated (Off-Warm), and snowberry-dominated (Off-Snow).

Table 4: Percent of pixels within each plant community that remain unchanged and that changed class belonging between 2015 and 2016 models.

Transition	2015 PC[1]	2016 PC	Total Pixels	Percent of Total Pixels
Unchanged Pixels	Off-Cool	Off-Cool	9712857	31.03
	On-PDG	On-PDG	6427817	20.54
	Off-Snow	Off-Snow	3401264	10.87
	On-PDF	On-PDF	887151	2.83
	Off-Warm	Off-Warm	555635	1.78
Changed Pixels	On-PDG	Off-Cool	2278390	7.28
	Off-Cool	Off-Snow	1468042	4.69
	Off-Cool	On-PDG	1262373	4.03
	Off-Snow	Off-Cool	1174565	3.75
	Off-Warm	Off-Cool	729511	2.33
	Off-Cool	Off-Warm	716503	2.29
	Off-Warm	Off-Snow	629212	2.01
	On-PDG	Off-Snow	626695	2.00
	On-PDG	On-PDF	362417	1.16
	On-PDF	On-PDG	343774	1.10
	Off-Snow	On-PDG	281061	0.90
	Off-Snow	Off-Warm	155213	0.50
	On-PDG	Off-Warm	82450	0.26
	On-PDF	Off-Cool	72758	0.23
	Off-Cool	On-PDF	69188	0.22
	Off-Warm	On-PDG	43132	0.14
	On-PDF	Off-Snow	19575	0.06
	Off-Warm	On-PDF	573	0.00
	On-PDF	Off-Warm	314	0.00
	Off-Snow	On-PDF	17	0.00

[1] Plant communities (PC) on prairie dog towns are grass-dominated (On-Grass) and forb-dominated (On-Forb); plant communities in off-town areas are cool season grass-dominated (Off-Cool), warm season grass-dominated (Off-Warm), and snowberry-dominated (Off-Snow).

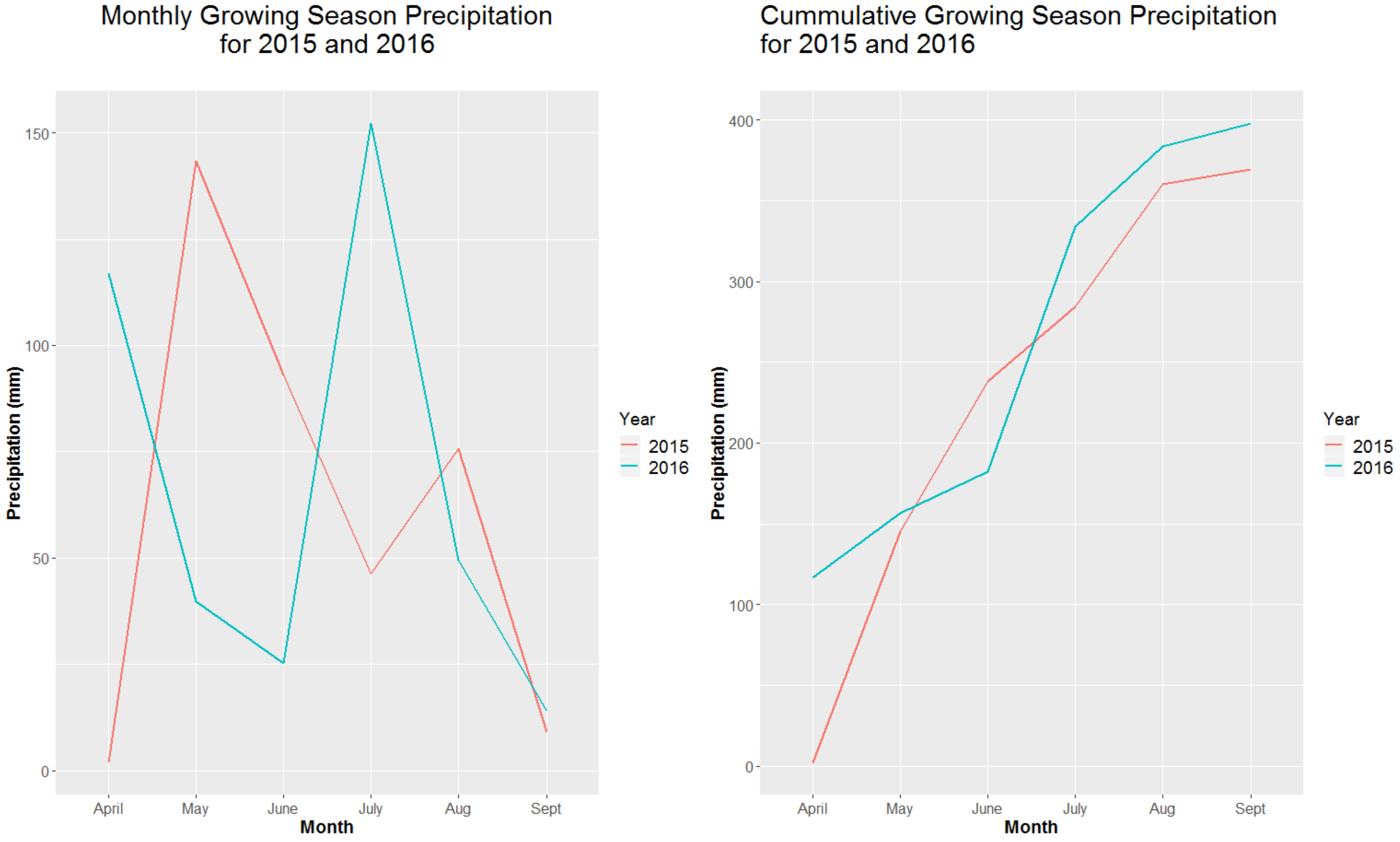

Figure 2: Monthly and cumulative growing season precipitation patterns for 2015 and 2016 recorded at a weather station located on the study area in north central SD (45.737296 N, -100.657540 W)(South Dakota Mesonet 2018).

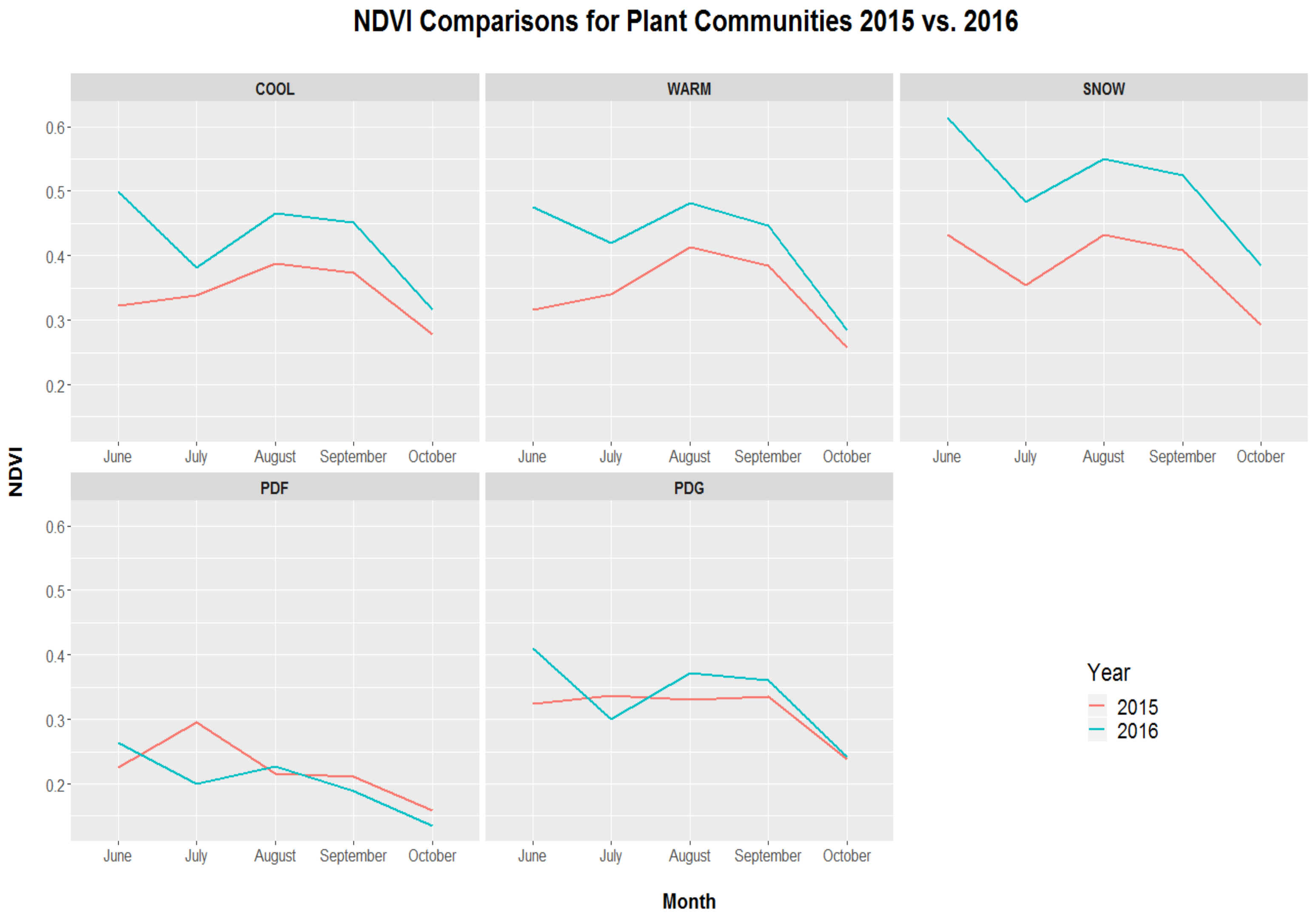

Figure 3: Comparison of mean monthly NDVI for training polygons in five plant communities on the study site in north central SD. Plant communities on prairie dog towns are grass-dominated (PDG) and forb-dominated (PDF); plant communities in off-town areas are cool season grass-dominated (COOL), warm season grass-dominated (WARM), and snowberry-dominated (SNOW).

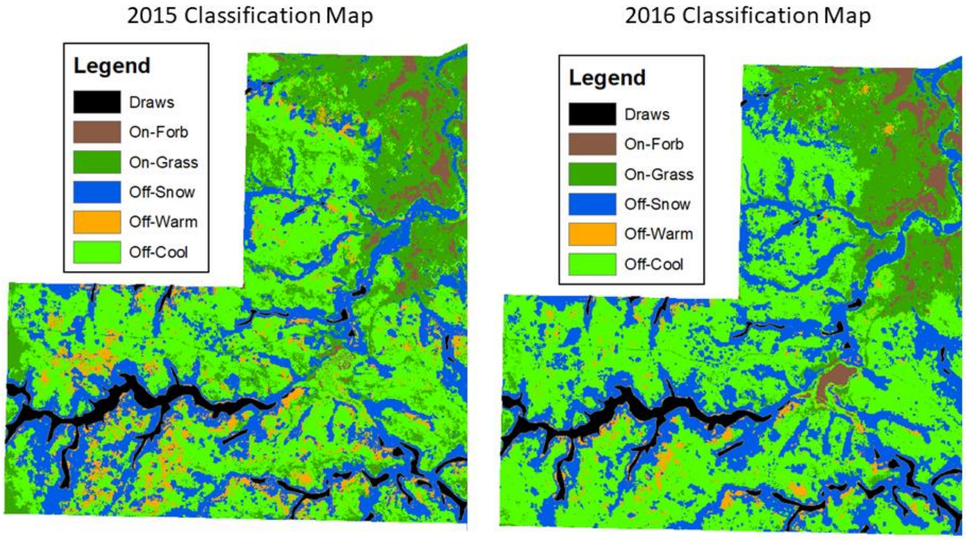

Figure 4: Random forest classification maps from 2015 and 2016 of one pasture in the study area in north central South Dakota. Plant communities on prairie dog towns are grass-dominated (On-Grass) and forb-dominated (On-Forb); plant communities in off-town areas are cool season grass-dominated (Off-Cool), warm season grass-dominated (Off-Warm), and snowberry-dominated (Off-Snow).

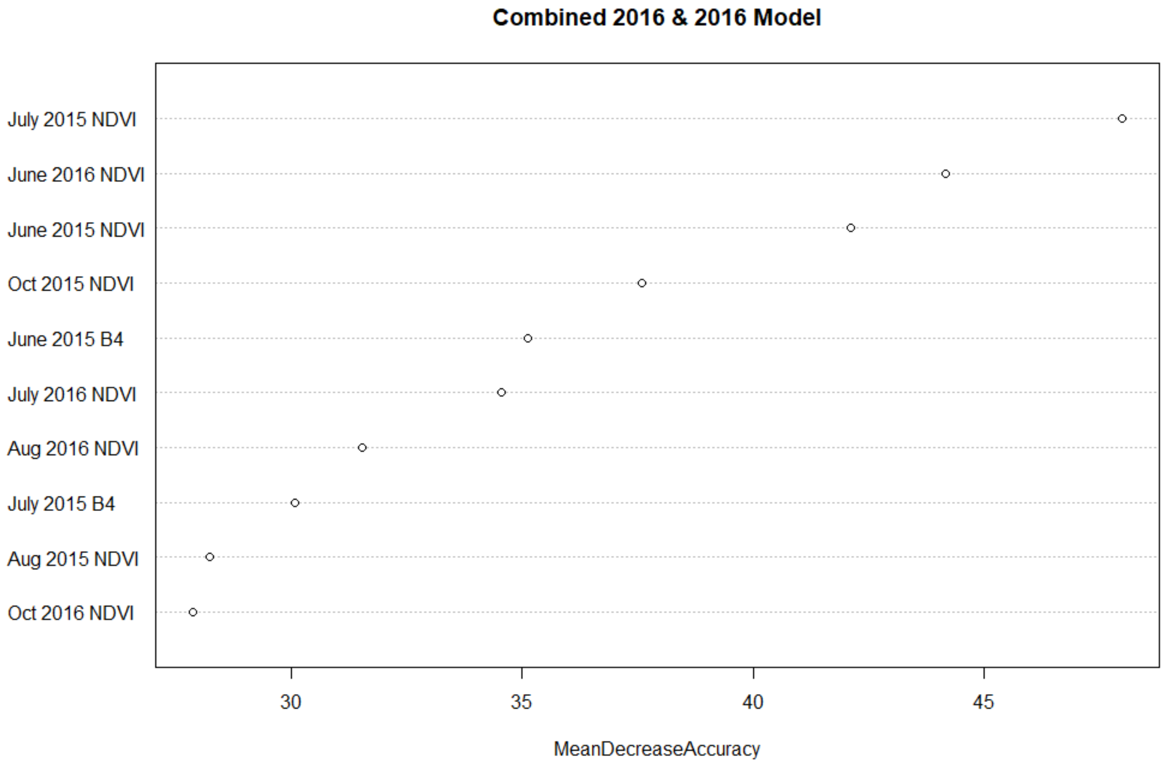

Figure 5: Variable importance reported as mean decrease in accuracy. Ten most important variables are shown, with B1 and B4 corresponding to spectral bands 1 and 4 respectively from Pleiades image. Variable importance is determined by the model output as the decrease in accuracy due to the exclusion of that variable during the out of bag error calculation process. Higher mean decrease in accuracy variables are more important in classifying the data.

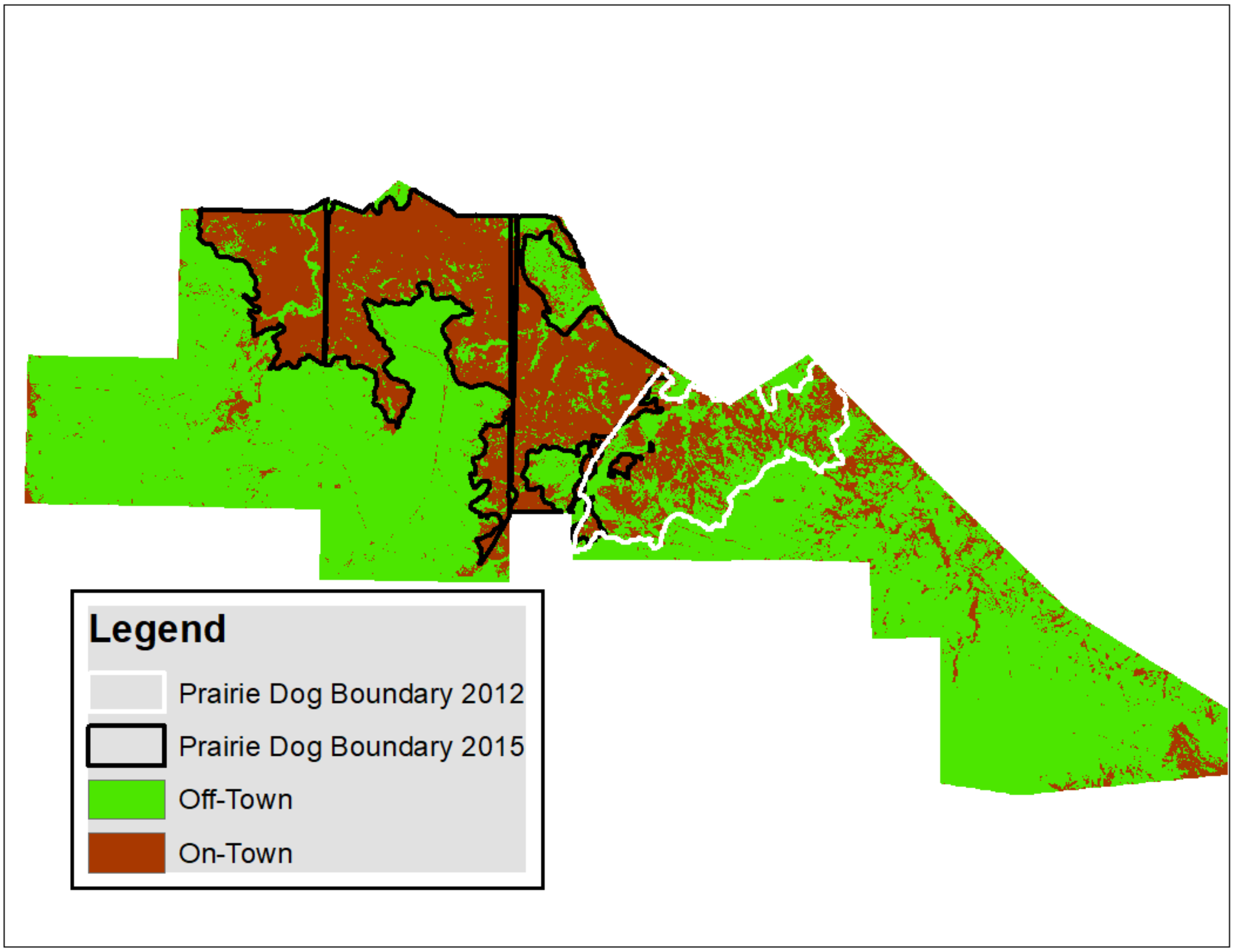

Figure 6: Random forest classification map created from predictions from the combined 2015 and 2016 models. Off-town areas were created by combining the predicted off-town plant communities (Off-Cool, Off-Warm, and Off-Snow) and on-town plant communities (On-Grass and On-Forb). The prairie dog boundary was mapped using a handheld GPS unit, the outlined 2012 prairie dog boundary was former prairie dog colony poisoned in 2013.