Fractional Woody Cover Mapping of Texas Savanna at Landsat Scale

Xuebin Yang 1,* and Kelley A. Crews 2

1 Department of Geography, Oklahoma State University, Stillwater, OK 74078, USA
2 Department of Geography and the Environment, University of Texas at Austin, Austin, TX 78712, USA; kelley@utexas.edu
* Correspondence: yxbfuture@utexas.edu; Tel.: +1-512-983-7686

Received: 26 October 2018; Accepted: 4 January 2019; Published: 5 January 2019

Abstract: Texas savanna experienced substantial woody plant encroachment during the past several decades, resulting in habitat fragmentation and species loss. A detailed map of woody plant abundance and distribution in this area is critically needed for management purpose. This study endeavors to map the fractional woody cover of Texas savanna at Landsat scale (30 m) in an affordable way. The top of atmosphere reflectance, thermal bands, and NDVI layer of Web-Enabled Landsat Data (WELD) of 2012 were used as predictors, together with mean annual precipitation. Classification and Regression Trees (CART) were calibrated against training data of a whole range of fractional woody cover, which were derived from 1-m resolution digital orthophotos of 2012. Validation indicates a reasonable pixel level accuracy of the result fractional woody cover map, with a R-squared value of 0.45. Moreover, the result map clearly depicts the distribution of woody plants across the study area, as reflected by the orthophotos. Furthermore, this new map proves an improvement over the existing Landsat Vegetation Continuous Fields (VCF) tree cover product. The method developed here, combining remote sensing and statistical techniques, can contribute to savanna management through revealing the abundance and distribution of woody plants.

Keywords: woody cover; fractional representation; Texas savanna; Web-enabled Landsat Data; Classification and Regression Trees; Landsat Vegetation Continuous Fields

1. Introduction

Savanna ecosystems feature the coexistence of woody plant and herbaceous vegetation, and the balance between these two different life forms is crucial for the structure and function of the ecosystems [1–3]. Texas savanna, for instance, suffered habitat fragmentation and biodiversity loss due to woody plant encroachment in recent decades [4–9]. Livestock carrying capacity is also reduced as a result of the encroachment. A detailed map of woody plant abundance and distribution in this area is critically needed for management purpose.

The increasing availability of remote sensing data provides us an opportunity to map woody plant cover over large areas in savanna ecosystems. Among them, very high spatial resolution imagery can recognize individual woody plants, but such imagery is characterized by small extent and low temporal frequency and is vulnerable to cloud contamination. These characteristics limit their application in regional woody cover mapping. In contrast, medium to coarse spatial resolution imagery of high coverage and high temporal frequency would be more suitable for large scale mapping. While the medium to coarse resolution reduces data volume required to cover large areas, the high temporal frequency can offset the effect of possible cloud contamination [10].

A traditional discrete classification scheme is not appropriate for mapping woody plant cover in savanna ecosystems with medium to coarse resolution data. It is because it assumes a pure land...
cover type within each pixel, but woody plant and herbaceous vegetation may be mixed within the pixels. Fractional representation, instead, can be a good alternative method for such a circumstance, because it measures the percent coverage of each landscape component (e.g., woody plant) within every single pixel. The fractional representation is more accurate and has greater potential in research and applications [11]. Moreover, successive fractional woody cover maps are able to reveal subtle and slow woody cover change such as woody plant encroachment [12].

The fractional representation is also more ecologically relevant [13]. For instance, gradient approaches have been increasingly popular in conceptualizing and representing landscape structures [14–18]. In these models, land cover varies continuously rather than discretely, which is much closer to the reality of landscape heterogeneity [19]. These fractional representation based models can advance our understanding of the relationship between landscape patterns and ecological processes [20], and particularly have been widely applied in savanna research [21,22].

Several fractional tree cover products covering Texas savanna have been developed. For instance, the MODIS Vegetation Continuous Fields (VCF) tree cover layer depicts the proportional coverage of trees in each 250-m pixel (500-m pixel for earlier versions) on the earth’s terrestrial surface [23]. Sexton et al. [24] rescaled the 250-m MODIS tree cover product to 30-m resolution with Landsat data. Hansen et al. [25] mapped the fractional tree cover of the conterminous United States at 30-m resolution with Web-Enabled Landsat Data (WELD). In addition, the 2011 U.S. Forest Service (USFS) tree canopy product estimates the proportion of each 30-m cell covered by tree canopy [26]. Hartfield and Leeuwen [27] mapped the fractional tree cover of the Edwards Plateau (part of Texas savanna) in Texas with seasonal Landsat data. However, since the training data of these products are limited to trees above a certain height and do not explicitly consider the short woody plants (e.g., shrubs and bushes) that are common in savanna ecosystems, their accuracy in Texas savanna may be limited.

Therefore, this study endeavors to map the fractional cover of all woody plants (trees, shrubs, and bushes) in Texas savanna. In terms of observation scale, a fine scale is expected to better capture local heterogeneity (e.g., very high and very low woody cover), which may be averaged and lost in coarse scale observations [28]. Moreover, woody cover change in savanna ecosystems usually occurs at a small spatial scale [29]. A fine scale observation is thus more sensitive to woody cover change in savanna ecosystems. For these reasons, we chose an observation scale of 30 m (Landsat scale) for this study.

There have been a lot of efforts in mapping fractional woody cover in other savanna ecosystems. For instance, taking high resolution airborne LiDAR data as reference data, Urbazaey et al. [30] mapped fractional woody cover of southern African savannas with multi-temporal ALOS PALSAR L-band backscatter. Brandt et al. [31] took advantage of the phenological differences between woody plants and herbaceous vegetation in the Sahelian drylands of Africa and mapped the fractional woody cover at 1 km resolution. Specifically, they tested 10 satellite metrics representing annual Fraction of Absorbed Photosynthetically Active Radiation (FAPAR) dynamics and related them to multi-year field measurement of fractional woody cover by multiple linear regression. Gill et al. [32] mapped the woody cover density of Australian forests and other wooded lands at 30 m scale, by linking field data across the continent to Landsat TM/ETM data of 2000 to 2010. Higginbottom et al. [33] mapped the fractional woody cover of the Limpopo Province of South Africa at four different observation scales (30 m, 60 m, 90 m, and 120 m), with application of Landsat TM/ETM-based seasonal metrics and fused Landsat-PALSAR data.

Though the above research proved a success in their study areas, they are hardly repeatable in Texas savanna. It is because that they either require large amount of SAR data and LiDAR data, or assume synchronous phenological cycles of woody plants, or involve extensive field measurement, or span a long time period, all of which cannot easily be met in Texas savanna. For instance, the SAR data and LiDAR data are currently not available for Texas savanna. In addition, Texas savanna has a complex distribution of woody species with asynchronous phenological cycles along the present precipitation gradient [6,7]. It is also not appropriate to apply long time field data or remote sensing
data in Texas savanna due to the ongoing woody plant encroachment. Other than that, most land in Texas savanna is privately owned and fenced, limiting accessibility for field measurement. Therefore, a more affordable method is in need for fractional woody cover mapping in Texas savanna.

2. Materials and Methods

2.1. Study Area

Texas savanna, located in central Texas, USA (Figure 1a), borders with dry plains on its west and moist prairies and woods on its east. It has important social and environmental values. For instance, it is home to a lot of small to large ranches, which are important sources of revenues for landowners. The Edwards Plateau ecoregion bounded in Texas savanna plays a critical role in ground-water recharge for the increasingly prosperous cities around this area such as Austin and San Antonio [34].

![Figure 1. (a) The study area of Texas savanna in central Texas, USA, displayed with Landsat GeoCover Mosaics of 2000 in a natural-like rendition; (b) distribution of rain gauge stations and the interpolated mean annual precipitation surface.](image)

However, these resources and services are being threatened by woody plant encroachment in this area. For example, a lot native herbaceous species was lost due to the encroachment [8]. In addition, the increasing woody plants were found to extract more and more water from the fractured bedrock of this area and reduce aquifer recharge [35]. Given so, a map of woody plant abundance and distribution in this area is urgently needed for informed management and restoration strategy.

This savanna area is featured by a wide precipitation gradient in the west-east direction, with mean annual precipitation (1981–2010) ranging from ~360 mm to ~950 mm (Figure 1b). The thick and mostly flat bedrock in this area mainly consists of hard early Cretaceous limestone, which results in shallow (less than 10 inches) and clay rich soil [36]. The high spatial and temporal heterogeneity of resource availability and disturbance regime, together with several decades of previous encroachment, led to an uneven distribution of woody plants across this area [37]. Currently, honey mesquite (Prosopis glandulosa), Ashe juniper (Juniperus ashei), and live oak (Quercus virginiana) are the most popular woody species, interspersed by perennial and mixed C3/C4 grasses [4,7,38].

2.2. Data Processing

The annual product of Web-enabled Landsat Data (WELD) of 2012 was used as predictor in this study. It includes layers of top of atmosphere (TOA) reflectance, brightness temperature and NDVI. It is a temporally compositied product, which selects pixels with maximum NDVI value out of a full year’s geometrically corrected Level 1T Landsat data [39]. This annual maximum captures the peak point of phenological cycles of vegetation and is thus insusceptible to the different seasonal cycles of woody species [40]. In addition, this product is insusceptible to Sun-Earth distance, solar geometry,
and exoatmospheric solar irradiance, which ensures spatial and temporal comparability and makes it suitable for large area application [41,42].

With regard to the limiting role of water availability in woody plant growth in arid and semi-arid savannas [22,43], water availability was incorporated as an ancillary layer in predicting fractional woody cover in this study. Given the relatively flat topography and uniform soil type of the study area [36], mean annual precipitation (MAP) of 1981–2010 was used as the surrogate of plant water availability. Firstly, MAP data of 1981–2010 was obtained for 51 rain gauge stations within Texas savanna from Southern Regional Climate Center (Figure 1b). Then the station data was interpolated across the study area at 250 m resolution with ordinary Kriging interpolation method, which proves more reliable than other methods [44].

Training data of fractional woody cover was derived from 1 m resolution digital orthophotos of 2012, provided by the National Agriculture Imagery Program (NAIP). The very high spatial resolution, true color composite, and the imaging time of growing season ensure the ability of the orthophotos in distinguishing woody plants of various heights from background (Figure 2a). In response to the principle that training data in land cover study should be representative of the study area [25], we randomly sampled 1500 plots (500 m) across Texas savanna. Specifically, a MODIS surface reflectance layer (MOD09GA) of 500-m resolution was used for the plots sampling. That is, a total of 1500 MODIS pixels (500 m) were randomly sampled across the study area. The corresponding digital orthophoto was then clipped out for each of the sampled MODIS pixels.

![Figure 2](image-url)  
**Figure 2.** (a) 1 m resolution digital orthophoto of a 500 m plot, (b) classification result of woody plant (green) vs. the others and sampling of WELD pixels (red polygons).

We first performed an unsupervised (ISODATA) pixel-based classification of the 1 m resolution orthophoto for each of the 1500 sample plots. Specifically, the orthophoto of each sample plot was classified into 10 thematic classes, and each class was manually categorized as ‘woody plant’ or ‘the others’ through visual inspection of the true color orthophoto (Figure 2). Secondly, we systematically sampled 13 WELD pixels (30 m) within each 500 m plot in the centered horizontal and vertical direction, with a 60 m spacing (Figure 2b). It resulted in a total of 19,500 WELD pixels (13 × 1500). Thirdly, we excluded the WELD pixels that correspond to land cover types such as water, developed land, and wetlands according to the National Land Cover Database 2011, and 16,927 WELD pixels were remained. Fourthly, we calculated the fractional woody cover within each of the remained WELD pixels as the proportion of 1 m woody pixels, based on the classified binary images.

2.3. Fractional Woody Cover Mapping

Out of the 16,927 orthophoto-based measurements of fractional woody cover, a random portion of 2000 was reserved for model validation, while the other 14,927 measurements were used for model calibration. The layers of top of atmosphere reflectance, brightness temperature and NDVI of the Web-enabled Landsat Data (2012), together with the mean annual precipitation layer were prepared as
predictors. To choose an appropriate model for fractional woody cover estimation, we first examined the relationship between the orthophoto-based woody cover and each of these predictors. Among them, three typical scatterplots are displayed in Figure 3.

![Figure 3. The scatterplots of fractional woody cover vs. green band (a), thermal band (b), and NDVI (c).](image)

As revealed by the above scatterplots, the relationships between fractional woody cover and the predictors are complicated. It is probably due to the large variation of physiognomic characteristic and vegetation composition and structure along the present precipitation gradient in Texas savanna [45], which is corroborated by our fieldwork and ancillary data (maps and geolocated photographs). Moreover, the large variation could have been exaggerated by the fine observation scale (30 m). With regard to the complex relationships between response and predictor variables, machine learning could be an appropriate approach for this study [46].

Specifically, Salford Systems’ Classification and Regression Trees (CART) was applied, which is based on the algorithm by Breiman et al. [46]. CART is advantageous over other regression tree techniques and has been widely applied by remote sensing community [47,48]. Firstly, this algorithm is non-parametric and does not assume any underlying distribution of predictors. Secondly, the binary partitioning of this algorithm divides the training dataset into child nodes at a slower rate than multi-way splits and thus can reveal more complete data structures. Thirdly, the identification of “splitting” variables results from exhaustive searches of all possibilities. Fourthly, CART is capable of handling missing variables to certain extent.

To reduce random error, five CART models were developed. Each was grown out of a random portion of 80% (11,942 measurements) of the training dataset. Initially, all the variables of top of atmosphere reflectance, brightness temperature, NDVI and mean annual precipitation were input to calibrate the CART model, against the orthophoto-based measurements of fractional woody cover. However, all the five model outputs suggest that NDVI, mean annual precipitation, and brightness temperature have very minimum (less than 3% each) contribution to the models. To avoid overfitting, five new CART models were created without these three variables. Each new model results in a fractional woody cover map of Texas savanna. A final map was then derived by taking the median value out of the five maps for each pixel.

3. Results

3.1. Result

The final fractional woody cover map at 30 m scale is displayed in Figure 4. The color ramp of light green to dark green illustrates the range of fractional woody cover from 0.09 to 0.77, while the white inside the map indicates non-woody areas such as developed land and water body, according to the National Land Cover Database 2011. This map shows a high spatial coherency, particularly an increasing trend of woody plant density along the precipitation gradient, which agrees with our
reconnaissance fieldwork from roadside and state park across the study area. It also corroborates the established increasing potential woody cover pattern over the precipitation gradient [22].

This map preserves the finest spatial detail and depicts the distribution of woody plants across Texas savanna. For instance, there is a large area of very low woody cover in the southwest region. In contrast, several contiguous dense woody areas are observed in the southeast region. It is evident that the very eastern portion of the study area is interspersed by a lot of non-woody areas. This is probably due to human cultivation and development associated with the adjacent metropolitan areas of Austin and San Antonio.

The relative importance of each predictor in estimating fractional woody cover is highly consistent among the five CART models. Their average relative importance is displayed in Figure 5. It is clear that green band plays a most important role (24%) in predicting fractional woody cover, while near infrared band has the least contribution (8%). The two shortwave infrared bands contribute 18% and 16% respectively to the CART models.

Figure 5. The relative importance of each variable in predicting fractional woody cover.

3.2. Accuracy Assessment

The pixel level accuracy of the result map was assessed with the reserved 2000 orthophoto-based woody cover measurements. The scatterplot of orthophoto-based vs. estimated fractional woody cover...
is displayed in Figure 6a. The black 1:1 line is a visual aid to illustrate over- and under-estimation. A simple linear regression was performed and the output line is overlaid on the scatterplot in red. The regression is statistically significant and has an R-squared value of 0.45.

As seen in the scatterplot, the estimated fractional woody cover roughly agrees with the orthophoto-based measurement. However, while the orthophoto-based measurement can be as high as 0.99, the estimation barely reaches 0.77. In addition, the points in the scatterplot in Figure 6a fall in a limited number of horizontal lines, indicating a limited number of estimated fractional woody cover values. Other than that, the respective trend of overestimation in sparse woody area and underestimation in dense woody area is observed, which has also been reported by other similar studies [49,50].

As a comparison, we assessed the ability of the Landsat VCF tree cover product (2012) in reflecting woody plant cover at pixel level in Texas savanna, with the same validation dataset (Figure 6b). The scatterplot (Figure 6b) suggests that the Landsat tree cover product tends to underestimate woody plant cover across almost the whole range of woody cover. A simple linear regression results in an R-squared value of 0.38.

In addition to the pixel level assessment, we analyzed the ability of these two products in reflecting broad scale woody plant distribution. A random area A in the arid environment and a random area B in the mesic environment were selected, as shown in Figure 4. The subsets of the product of this study and the Landsat tree cover product in areas A and B were compared to the corresponding 1 m resolution digital orthophotos (Figure 7). In area A, the result of this study successfully captures the dense short woody plants that are in a branch-like distribution, while the Landsat tree cover product is not sensitive to these short woody plants at all. In area B, the result of this study clearly reveals the gradation of woody plant cover, while the Landsat tree cover product only captures part of the woody plants. It is suggested that the product of this study is an improvement over the Landsat tree cover product in reflecting woody plant cover in Texas savanna.
which is manifested as spectral heterogeneity of pixels, even those containing the same woody cover, and raises the estimation in the low end [40]. In addition, points in the scatterplot of the WELD product and orthophoto could be another concern, especially in consideration of the fine classification [53]. Incorporation of other metrics such as fragmentation information in future study is also the culprit of the ‘salt and pepper’ effect in the traditional pixel-based image heterogeneity is also the culprit of the ‘salt and pepper’ effect in the traditional pixel-based image.

Figure 6a are arrayed in a limited number of horizontal lines. It is a manifestation of the discontinuous character of CART-based products [48], which does not conform to reality. This discontinuity could be attributed to the limited number of terminal nodes in the CART model. Thus, caution should be exercised when applying the CART model in other savanna areas.

However, limitations exist with the CART model. For instance, the CART is implicated in the aforementioned trend of overestimation in sparse woody area and underestimation in dense woody area. It is partly because that the mean value of the training data (fractional woody cover) in each terminal node was taken as the prediction. Doing so lowers the value in the high end of fractional woody cover, and raises the estimation in the low end [40]. In addition, points in the scatterplot of Figure 6a are arrayed in a limited number of horizontal lines. It is a manifestation of the discontinuous character of CART-based products [48], which does not conform to reality. This discontinuity could be attributed to the limited number of terminal nodes in the CART model. Thus, caution should be exercised when applying the CART model in other savanna areas.

The R-squared value 0.45 of this study is not as perfect as expected. It could be attributed to several aspects. Firstly, the heterogeneity of savanna landscape should take part of the responsibility [45], which is manifested as spectral heterogeneity of pixels, even those containing the same woody cover (Figure 3). It is especially true for medium to high resolution imagery [51,52]. This spectral heterogeneity is also the culprit of the ‘salt and pepper’ effect in the traditional pixel-based image classification [53]. Incorpcion of other metrics such as fragmentation information in future study may overcome the side effect of the spectral heterogeneity [40]. Secondly, the geo-location accuracy of the WELD product and orthophoto could be another concern, especially in consideration of the fine scale observation [50]. Thirdly, Tan et al. [54] revealed that the uniform grid cells of remote sensing

4. Discussion

The accuracy assessment suggests a reasonable pixel level accuracy of the result woody cover map. More importantly, the result map exhibits great ability in depicting the broad scale woody plant distribution across Texas savanna. Its success draws on several aspects. Firstly, the availability and coherency of the WELD product is critical to this study. This radiometrically normalized, temporally compositied, and spatially mosaicked dataset represents a dramatic advance in Landsat data processing and provides a great channel for large scale landscape mapping at a fine resolution [25,39]. Secondly, the advantages of the Classification and Regression Trees (CART) ensure reliable prediction models [46,47], especially when referring to the complex nonlinear relationships between fractional woody cover and the predictors (Figure 3).
products correspond to smaller ground coverage than their source observations, which may have biased the training and validation data measurement to some extent.

There may be a balance between spatial details and pixel level accuracy in landscape components mapping. That is, a fine scale mapping can reflect more spatial details, whereas it may have a relatively low pixel level accuracy. On the contrary, a coarse scale mapping may have a higher pixel level accuracy but less spatial details. It is partly corroborated by the finding by Hansen et al. [12] that the pixel level accuracy steadily improves with observation scale from 500 m to 1000 m and 2000 m.

The finding that the green band of Landsat ETM+ data plays a most important role in predicting fractional woody cover in this savanna area differs from the finding by Hansen et al. [40] that the red band of MODIS data is most important in fractional tree cover mapping. This discrepancy suggests the unique spectral property of savanna ecosystems. Mean annual precipitation (MAP) and NDVI have very minimum contribution and were excluded out of the CART models. This is probably due to that NDVI value is a reflection of all live green vegetation rather than woody plants only. As for MAP, though it determines the maximum realizable woody cover of a given site in savanna area [22,43], the reality of woody cover is under the effect of other factors such as management history and external disturbances.

The result of this study shows a huge improvement over the Landsat tree cover product in reflecting woody plant cover in Texas savanna (Figure 7). Literature review suggests that the most significant difference underlying these two products lies in training data [24]. While the training data of the Landsat tree cover product is limited to trees above 5 m in height [23,24], the training data in this study incorporates all woody plants (trees, shrubs, and bushes). In addition, the training data of this study is more representative of Texas savanna, when comparing to the global training data of the Landsat tree cover product [23,24].

Besides depicting the abundance and distribution of woody plants, the result fractional woody cover map could be applied in the follow-up research in Texas savanna. For instance, since woody cover proves a reliable predictor of woody carbon storage in savanna ecosystems [55,56], we can estimate the woody carbon storage of Texas savanna on the basis of the result map. In addition, the fractional representation of woody plant cover has been critically needed in characterizing the two-component structure of savanna ecosystems [22,43].

5. Conclusions

In conclusion, this study presents an affordable approach to map the fractional woody cover of Texas savanna at a fine resolution. The result fractional woody cover map represents an improvement over the Landsat tree cover product in reflecting woody plant cover in Texas savanna. Continued exploration of this approach can contribute to the follow-up research in carbon stock estimation and savanna dynamics modeling. It may also contribute to best practices of landscape components mapping over large areas in other ecosystems with these and other remote sensing data. This study highlights the importance of multi-source remote sensing data and machine learning in natural resource inventory and management, as well as the importance of training data in landscape components mapping.

Author Contributions: X.Y. conducted this research and wrote the paper. K.C. supported this research and revised the paper.

Funding: This research was funded by National Science Foundation, grant number BCS-0964596.

Conflicts of Interest: The authors declare no conflict of interest.

References


37. Wilcox, B.P.; Huang, Y. Woody plant encroachment paradox: Rivers rebound as degraded grasslands convert to woodlands. *Geophys. Res. Lett.* 2010, 37. [CrossRef]


49. White, M.A.; Shaw, J.D.; Ramsey, R.D. Accuracy assessment of the vegetation continuous field tree cover product using 3954 ground plots in the south-western USA. Int. J. Remote Sens. 2005, 26, 2699–2704. [CrossRef]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).