Mapping forest in the southern Great Plains with ALOS-2 PALSAR-2 and Landsat 7/8 data

Xuebin Yang a,*, Xiangming Xiao a, Yuanwei Qin a, Jie Wang b, Kevin Neal c

a Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK 73019, United States
b College of Grassland Science and Technology, China Agricultural University, Beijing 100094, China
c Department of Geography and Environmental Sustainability, University of Oklahoma, Norman, OK 73019, United States

A R T I C L E   I N F O

Keywords:
Forest
Evergreen forest
Woody plant encroachment
Southern Great Plains
ALOS-2 PALSAR-2
Landsat
Google Earth Engine

A B S T R A C T

Southern Great Plains (SGP) of the United States, comprising the states Kansas, Oklahoma, and Texas, spans diverse climatic regions. In recent decades, woody plant continues to expand and form forest (above 2 m in height) across the SGP. However, our knowledge of the forest amount and distribution in this region is very limited. This study aims to map forest, especially evergreen forest (above 2 m in height) in the SGP for the time period of 2015–2017. Annual mosaic data of HH and HV polarization backscattering (25 m) from Phased Arrayed L-band Synthetic Aperture Radar-2 (PALSAR-2) aboard Advanced Land Observing Satellite-2 (ALOS-2), along with their difference (HH-HV) and ratio (HH/HV) were utilized. With the four bands (HH, HV, difference, ratio) of 2017, decision rules of forest were developed based on 30 randomly selected forest plots (as of 2017) across the study area. With the decision rules, a PALSAR-2 based forest map was created for each year from 2015 to 2017. Then an annual maximum normalized difference vegetation index (NDVI) threshold of 0.5, derived from Landsat 8 data of 2017 and the 30 forest plots, was used to filter potential commission error of rough surface and building in each PALSAR-2 based forest map. After that, a forest map circa 2016 was generated, in which each forest pixel was identified as such at least twice during 2015 and 2017. Lastly, a threshold of seasonal NDVI change (0.3) was derived to extract evergreen forest out of the forest map circa 2016. Accuracy assessment for the result forest map suggests a user’s accuracy of 99.2% and a producer’s accuracy of 88.7% for forest. Accuracy assessment for the evergreen forest map suggests a user’s accuracy of 97.3% and a producer’s accuracy of 90.5% for evergreen forest. The result forest map, especially the evergreen forest map, paves the way for follow-up studies on forest resource and woody plant encroachment in the SGP.

1. Introduction

Forest cover and its distribution are very important indicators of biodiversity, climate change, carbon and water cycles at regional to global scale (Foley et al. 2005; Hansen et al., 2016). As for the southern Great Plains (SGP) of the United States, which comprises the three states Kansas, Oklahoma, and Texas, forest area has been expanding mainly due to the long-term woody plant encroachment (Barger et al., 2011; Wine and Zou 2012). Juniper (Juniperus), oak (Quercus), and mesquite (Prosopis glandulosa) are the primary encroaching species in the SGP. The increase of mesquite and oak trees across Texas and the concomitant formation of closed-canopy largely lowered biodiversity (Ansley et al. 2001; Diamond and True 2008). The encroachment of junipers in the Edwards Plateau of central Texas caused habitat fragmentation and herbaceous species loss (Alofs and Fowler 2010, 2013). In addition, the rapid expansion of junipers across Oklahoma and Kansas in recent decades severely threatened the present prairies in terms of forage and livestock productivity (Knapp et al. 2008; Twidwell et al. 2016).

The above recognized encroachment and consequence further underscore the necessity of forest map for the SGP, not only for forest resource inventarisation but also for pertinent encroachment management strategy (Bucini and Hanan 2007). It has to be noted that much of the forest in the SGP does not reach 5 m high (Scholtz et al., 2018), a criterion adopted by the widely applied forest definition of the United Nations Food and Agriculture Organization (FAO) (FAO Statistics 2010). One reason is because many of the encroaching trees in the SGP are at early life stages (e.g. juvenile) (Hughes et al. 2006). The other reason is that the stature of a lot of trees, especially in southwest SGP, is limited by...
the climate, fire, soil type and tree species (Simpson 1999; Scholtz et al. 2018). As such, this study adopts the Australian forest definition of an area with tree height above 2 m and tree crown cover over 20% (Hnatiuk et al. 2003). However, the existing forest products covering the SGP such as Landsat-based global forest map (Hansen et al. 2013) and radar-based global forest/non-forest maps (Shimada et al. 2014) are targeted at trees above 5 m in height, which makes them unsuitable to track forest area in the SGP. Therefore, the first objective of this study is to develop a forest map for the SGP, following the Australian forest definition.

Different methods have been developed with optical remote sensing data for local- to regional-scale woody plant mapping, which mainly take advantage of the phenological discrepancy between woody plant and herbaceous vegetation (Yang 2019). Brandt et al. (2016) mapped woody plants by identifying their photosynthetic activity through seasonal metrics of fraction of absorbed photosynthetically active radiation (FAPAR) in the Sahelian drylands of Africa. Higginsbottom et al. (2018) proved that the Landsat metrics of dry season are most useful in mapping woody cover in semi-arid savannas of South Africa. The phenology of woody plants in these study areas is relatively uniform and distinctive from that of herbaceous vegetation. This prerequisite, however, does not hold on in the SGP, where phenophases of trees vary considerably between evergreen and deciduous species, and differ a lot from the southern-most subtropical climate to northern-most mild temperate climate (Wilcox et al. 2018). Other than that, optical remote sensing data of medium to coarse resolution has very limited capability in capturing sparse tree cover (Hansen et al. 2005; Montesano et al. 2009; Yang and Crews 2019).

The increasing availability of synthetic aperture radar (SAR) data provides us an alternative means to map forest in the SGP (Woodhouse 2017; Qin et al. 2021). While microwaves from SAR sensors can penetrate cloud and interact with different parts of trees (e.g. leaves, branches, trunks) according to their wavelength, polarization and incidence angle, these emitted signals are much less responsive to non-woody vegetation (Lucas et al. 2004; Urbazaev et al. 2015). The recorded SAR backscatter intensity, mainly determined by tree canopy structure and moisture, can therefore be utilized to identify forest out of other land cover types (Raney, 1998). Among various SAR data, those of longer wavelength such as L- and P-band (>15 cm) have greater penetration into trees and more interaction with branches and trunks, and are consequently preferred for woody plant structure modeling (e.g. woody cover, aboveground biomass) (Lucas et al. 2004; Naidoo et al. 2016; Liu et al. 2021).

The L-band of Advanced Land Observation Satellites’ Phased Array L-band Synthetic Aperture Radar-1/2 (ALOS-1/2 PALSAR-1/2) is featured by a wavelength of 23.6 cm and four possible polarizations (HH, HV, VH, VV) (Rosenqvist et al. 2007). It has been very popular in large scale forest mapping (trees taller than 5 m). Dong et al. (2012) generated a forest map for Mainland Southeast Asia with three layers (HV, HH-HV, HH/HV) derived from PALSAR-1 data of 2009, Shimada et al. (2014) created annual global forest maps of 2007 to 2010, with PALSAR-1 data and region-specific thresholds of HV gamma-naught ($\gamma$). Reihe et al. (2018) incorporated PALSAR-2 data in forest deforestation monitoring in tropical area. On the other side, the PALSAR data has been successfully applied in mapping woody plant shorter than 5 m in savannas. Urbazaev et al. (2015) demonstrated strong sensitivity of PALSAR-1 L-band signal to woody cover in southern African savannas. Naidoo et al. (2016) proved the advantage of PALSAR-1 L-band over Landsat 5 data in retrieving savanna woody cover. As such, the PALSAR data shows promising capability in identifying trees both above and below 5 m in height. Nevertheless, it has been rarely used to map trees both taller and shorter than 5 m simultaneously. Given so, this study will explore its potential to capture the forest above 2 m in height in the SGP.

The second objective of this study is to extract evergreen forest out of the result forest map. The consideration is double fold. Firstly, evergreen forest differs a lot from other forests in terms of carbon storage, water use efficiency, and climate adaptability (Wu et al. 2015). And evergreen forest (e.g. junipers, pine trees) is a crucial component in the SGP (Lyons et al. 2009; Wang et al. 2018; Yang and Crews 2020). Secondly, evergreen forest (e.g. junipers) has been expanding rapidly across the SGP, especially northward in recent decades (Twidwell et al. 2016). An accurate evergreen forest map can help disentangle the complex encroachment pattern.

Various approaches have been developed to map evergreen forest. Xiao et al. (2006, 2009) compared the intra-annual profile of land surface water index (LSWI) and enhanced vegetation index (EVI) of different land cover types. It was found that the LSWI of evergreen forest stays above zero throughout the whole year, while EVI of evergreen forest is always greater than 0.2 (Wu et al. 2009). This criterion has been widely applied in generating evergreen forest map (Sheldon et al. 2012; Qin et al. 2019). Wang et al. (2017, 2019) analyzed the inter-annual profile of EVI, LSWI, and NDVI for evergreen forest (junipers) and non-evergreen forests (oak, bottomland hardwood). It was concluded that mean NDVI value of winter season (December, January, and February) can best separate the evergreen forest from non-evergreen forests. In this study, we will develop and test a new approach to identify evergreen forest out of the result forest map, in the hope of enriching the literature on evergreen forest mapping.

2. Materials and methods

2.1. Study area

The SGP of the United States comprises the three states Kansas, Oklahoma and Texas (Fig. 1). This region is featured by various transitional regimes (Bagley et al. 2017). Along the south-north direction, it ranges from subtropical climate to mild temperate climate (Scholtz et al. 2018). Mean annual precipitation (MAP) increases from 242 mm in the west to 1814 mm in the east (Fig. S1a). Following the MAP trend, surface soil moisture ranges from 2.9 mm to 25.4 mm (Fig. S1b). The elevation rises from sea level in the southeast to 2463 m in the west (Fig. S1c). From northeast to southwest, mean daytime land surface temperature increases from 287 K to 313 K (Fig. S1d). According to US Level III Ecoregions, the SGP covers a wide variety of ecoregions, such as...
Chihuahuan Deserts in the western endpoint of Texas, Cross Timbers crossing northern Texas and central Oklahoma, and Flint Hills in eastern Kansas (Omernik and Griffith 2014).

Woody plant encroachment in the SGP started with overgrazing and fire suppression during the middle to late 1800 s that came along with the European settlement (Walker and Janssen 2002). Other secondary factors such as tree seed spreading by livestock, rising atmospheric CO$_2$ level, and increased precipitation intensity accelerated the encroachment (Kulmatiski and Beard 2013; Archer et al. 2017). As a result, much of the original grassland and open savanna is now occupied by woodland and forest (Barger et al. 2011). Nevertheless, the trajectory of woody plant encroachment and human intervention differs across the SGP. The encroachment in Kansas and Oklahoma was more recent than in Texas (Box 1967). This is probably due to the widespread cultivation in Kansas and Oklahoma following the European settlement (Wilcox et al. 2018). While much of the cultivated land was later returned to grassland for sustainability reason, a lot of cropland still exist in these two states. In terms of encroachment management, prescribed fire has long been a popular tool in the Tallgrass Prairie. In other parts of the SGP, however, people are just getting used to prescribed fire along with its recent success in addressing the more and more severe encroachment (Taylor et al. 2012; Twidwell et al. 2013). These discrepancies in the encroachment and management history resulted in a complex mosaic of landscapes with nonuniform encroachment stage and tree stature (Hughes et al. 2006; Scholtz et al. 2018).

As mentioned in the introduction section, the adverse effect of the encroachment is evident across the SGP. More importantly, the encroachment (especially junipers) is continuing, since the vast majority of this region does not reach the upper bound of woody plant cover imposed by climatic conditions (Yang et al. 2016, 2020; Scholtz et al. 2018). Therefore, detailed forest map, especially evergreen forest map of the SGP, is in urgent need for forest resource investigation, targeted restoration effort, identification and conservation of remnant grasslands.

2.2. Data and preprocessing

2.2.1. ALOS-2 PALSAR-2 data

Three annual mosaics (2015–2017) of ALOS-2 PALSAR-2 L-band data available in Google Earth Engine data catalog were utilized in this study. Each mosaic contains a co-polarized wave of HH and a cross-polarized wave of HV, both having a spatial resolution of 25 m. While the HH signal is mainly indicative of double bounce scattering associated with tree trunks, buildings or inundated vegetation, the HV signal is primarily a reflection of volume scattering related to tree leaves and branches (Watanabe et al. 2006). These global ortho-rectificated mosaics were clipped to the SGP (Fig. 1). The digital number (DN) stored in the two polarization bands was respectively converted to backscatter gamma-naught ($\gamma^0$) value (unit: decibel) for further analysis. The following Eq. (1) was applied in Google Earth Engine (Shimada et al. 2009).

$$\gamma^0 = 10 \times \log_{10}(DN^2) - 83 \tag{1}$$

2.2.2. Landsat 7/8 data

The atmospherically corrected and orthorectified surface reflectance data from Landsat 7 ETM+ and Landsat 8 OLI sensors were used. Both sensors have a revisit cycle of 16 days and acquire reflectance data of earth surface at broad wavebands of blue, green, red, near-infrared and shortwave infrared. Specifically, the surface reflectance data of the SGP from 2015 to 2017 were accessed from Google Earth Engine. Bad-quality observations of the two sensors due to cloud, cloud shadow and snow...
were masked out by functions `cloudMaskL457` and `maskL8sr` respectively, with pixel quality band (pixel_qa).

2.2.3. NLCD2016 and CDL2016

The National Land Cover Database 2016 (NLCD2016) was developed by the U.S. Geological Survey (USGS) to keep the NLCD products (1992, 2001, 2006, and 2011) up-to-date (Yang et al. 2018). It provides reliable land cover information (e.g. developed, wetlands) at 30 m resolution for the entire United States. This product has been widely applied in land cover studies (Homer et al. 2020). In this study, NLCD2016 was used as a reference to check the result forest map, particularly the classification result in non-forest area (e.g. planted/cultivated). The Cropland Data Layer (CDL) provides crop-specific land cover information at moderate resolution (e.g. 56 m), for the continental United States (USDA NASS 2016). It is developed on an annual basis by National Agricultural Statistics Service (NASS) of United States Department of Agriculture (USDA). The CDL2016 of 30 m resolution was used to check possible commission error of cropland in the result forest map.

2.2.4. Validation data

A comprehensive and representative set of validation data was prepared for accuracy assessment of the result forest map and evergreen forest map. Firstly, the study area was divided into 1794 grid cells of a quarter degree by a quarter degree (0.25° × 0.25°) (Fig. S2a). Secondly, in each grid cell, the footprints of one to three Landsat pixels of typical land cover types were randomly selected. The selection was in reference to time-series high resolution imagery in Google Earth and digital orthophoto (1 m) of National Agriculture Imagery Program (NAIP) in Google Earth Engine. The presence (amount) of forest (evergreen/deciduous) validation pixels in each grid cell mainly depends on the occurrence (abundance) of forest (evergreen forest).

The selected footprints of Landsat pixels cover seven land cover types across the SGP, including forest (evergreen/deciduous), shrubland, cropland, grassland, barren land, building, and water. To minimize the effect of geometric accuracy of remote sensing data, each footprint was located within a homogeneous land cover patch (e.g. 3-pixel × 3-pixel window). The corresponding land cover type of each selected Landsat pixel was consistent across the study period of 2015 to 2017. A total of 2339 footprints were obtained, of which 1270 are forest and 1069 are non-forest (Fig. S2b). Among the 1270 forest footprints, 559 are evergreen forest, 484 are deciduous forest, and 227 are either mixed or unknown-type forest due to the lack winter season imagery.

2.3. Workflow of this study

The methodological flowchart of this study is displayed in Fig. 2. The workflow consists of four major steps. First, thresholds of HH, HV, HH/HV, HH-HV were derived for forest (above 2 m in height) in the SGP, resulting in annual PALSAR-2 based forest maps from 2015 to 2017. Second, threshold of annual maximum NDVI was calculated for forest, which was used to filter potential commission error of building and rough surface in the PALSAR-2 based forest maps. Third, a final forest map circa 2016 was generated by compositing the three annual forest maps. Fourth, threshold of seasonal NDVI change of evergreen forest was quantified to identify evergreen forest out of the forest map circa 2016.

2.4. Development of forest map

2.4.1. PALSAR-2 based decision rules for forest

Previous research suggests that PALSAR backscatter from forest of a given region is generally consistent and constrained within a certain range (Shimada et al. 2014; Qin et al. 2015, 2017). This study seeks to establish backscatter thresholds for forest over 2 m high in the SGP, in order to generate forest map for this region. A total of 30 typical forest sites (including both evergreen forest and deciduous forest) were randomly selected across the SGP for the threshold derivation (Fig. S3). They are all hand-drawn quadrilaterals in reference to the very high spatial resolution imagery (2017) available in Google Earth. Their sizes vary according to the corresponding homogeneous forest patches but are generally over 250 m by 250 m. These 30 forest sites consist of 9015 PALSAR-2 pixels.

As shown in Fig. 3, the probability density of the four variables (HH, HV, HH-HV, and HH/HV of the sample forest sites).
Table 1
Thresholds of the variables.

<table>
<thead>
<tr>
<th></th>
<th>Low end</th>
<th>High end</th>
</tr>
</thead>
<tbody>
<tr>
<td>HH</td>
<td>-12.0</td>
<td>-1.0</td>
</tr>
<tr>
<td>HV</td>
<td>-18.0</td>
<td>-7.5</td>
</tr>
<tr>
<td>HH-HV</td>
<td>1.0</td>
<td>11.0</td>
</tr>
<tr>
<td>HH/HV</td>
<td>0.1</td>
<td>0.9</td>
</tr>
</tbody>
</table>

Fig. 4. Probability density of NDVI_{\text{max}} (2017) for the 30 forest training sites.

2.4.2. Annual NDVI data for the SGP.

Each variable for forest. The rounded threshold values (Table 1) were used to generate annual forest maps of 2015 to 2017 based on PALSAR-2 data for the SGP.

2.4.2. Annual NDVI_{\text{max}} threshold

Some other land cover types like barren and developed land (e.g. rough surface, building) could have similar SAR backscattering as forest (Meyer 2019). Therefore, the above PALSAR-2 based decision rules may result in misclassification of rough surface and building as forest. Fortunately, forest tends to have significantly higher annual maximum NDVI (NDVI_{\text{max}}) than rough surface and building (Defries and Townsend 1994). This sharp difference has been utilized to filter the potential commission error (Qin et al. 2015, 2016). But the applied annual NDVI_{\text{max}} threshold varies with observation scale and study area. While Qin et al. (2015) applied a threshold of 0.5 at MODIS scale in China, Qin et al. (2016) used a threshold of 0.7 at Landsat scale in Oklahoma, USA.

To obtain the optimum annual NDVI_{\text{max}} threshold for this study in the SGP, we analyzed the probability density of annual NDVI_{\text{max}} values (2017) for the aforementioned 30 forest training sites. First, an annual NDVI_{\text{max}} layer was generated for the SGP with Landsat 8 data of 2017 (a total of 1471 Landsat 8 scenes) in Google Earth Engine. Second, the annual NDVI_{\text{max}} values were extracted for the training sites, which consist of 6237 Landsat pixels. Third, the probability density graph of the annual NDVI_{\text{max}} values was plotted for the training sites (Fig. 4). Therefore, we calculated 1% percentile (0.50) as the threshold annual NDVI_{\text{max}} value, in order to mask out the potential commission error of rough surface and building in the PALSAR-2 based forest maps. To do so, each PALSAR-2 based forest map (25 m) was resampled to the spatial resolution of annual NDVI_{\text{max}} layer (30 m) by nearest neighbor method. Annual NDVI_{\text{max}} layers of 2015 and 2016 were also created, with a total of 1357 and 1470 Landsat 8 scenes respectively.

2.5. Identification of evergreen forest out of the result forest map

We take advantage of the different seasonal NDVI change (NDVI-change) between evergreen forest and deciduous forest to identify evergreen forest out of the result forest map. As demonstrated by Wang et al. (2017, 2018), deciduous forest has very high NDVI value (<0.8) in summer season, but it drops quite a lot by winter season (<0.4). As for evergreen forest, however, the NDVI value is relatively stable across different seasons (around 0.6) and the seasonal change is much smaller. Given so, it is possible to separate evergreen forest and deciduous forest by a threshold value of seasonal NDVI change. To derive the threshold value, we randomly sampled 26 evergreen forest sites and 20 deciduous forest sites across the SGP (Fig. 5), followed by below steps.

First, a NDVI_{\text{max}} layer was generated for the SGP, with Landsat 7 and Landsat 8 data of 2015 to 2017 (a total of 8009 Landsat scenes) available in Google Earth Engine. Second, a winter mean NDVI (NDVI_{\text{winter}}) layer was created with Landsat 7/8 data of January and February of the three years (a total of 1231 Landsat scenes). To achieve the best separability, December Landsat 7/8 data was not included in the NDVI_{\text{winter}} calculation. This is because according to our observation, some deciduous trees still hold leaves with varying degrees of senescence over December in this region, especially in the southern state Texas. Third, a seasonal NDVI change layer was derived as the difference between the NDVI_{\text{max}} layer and NDVI_{\text{winter}} layer. Fourth, seasonal NDVI change values were extracted for the evergreen forest and deciduous forest sample sites, which consist of 4321 and 5955 Landsat pixels respectively.

As shown in Fig. 5, the probability density curves of seasonal NDVI change for evergreen forest and deciduous forest have very little overlap. The upper and lower percentiles were respectively calculated for the two forest types. The 97%, 98%, and 99% percentiles of evergreen forest are 0.29, 0.31, and 0.337, while the 1%, 2%, 3% percentiles of deciduous forest are 0.338, 0.35, and 0.36. It is clear that the seasonal NDVI change of evergreen forest is statistically and significantly lower than that of deciduous forest. In this study, we choose the rounded value of 0.3 as the threshold value of seasonal NDVI change. Forest pixels with seasonal NDVI change below 0.3 were classified as evergreen forest in the SGP.

3. Results

3.1. PALSAR-2/Landsat 8 based annual forest maps

The PALSAR-2/Landsat 8 based annual forest maps are displayed in Fig. 6. In each forest map, green color represents forest. Red color represents the commission error of rough surface and building introduced by the PALSAR-2 based algorithm but masked out by the annual NDVI_{\text{max}} threshold. These maps show a forest area of 301,092 km$^2$ in 2015, 330,000 km$^2$ in 2016, and 309,655 km$^2$ in 2017. It is evident that most forest distributes in southeast SGP, while most PALSAR-2 related commission error occurs in southwest Texas. There is also significant commission error in the metropolitan areas of Oklahoma City, Dallas-
3.2. Forest map circa 2016

As stated above, there are variations among the three annual forest maps in terms of forest area and distribution. Examination of the discrepancy forest regions, in reference to time-series high resolution imagery in Google Earth, suggests that almost all the variations are just uncertainty rather than forest gain or loss. To minimize these variations, a forest map circa 2016 was generated, in which each forest pixel is identified as such at least twice in the three annual forest maps of 2015 to 2017. The result forest map circa 2016 is displayed in Fig. 7a, which shows a forest area of 308,827 km$^2$. A check of this forest map with National Land Cover Database 2016 (NLCD2016) indicates some commission error of cropland in northwest and southern-most SGP, and some commission error of emergent herbaceous wetland in southeast SGP. Thereafter, these two types of commission error were masked out with NLCD2016 layer, leading to a final forest map circa 2016 (Fig. 7b) that exhibits a forest area of 293,648 km$^2$.

3.3. Validation of the forest map circa 2016

The whole validation dataset from Section 2.2.4 (1270 forest and 1069 non-forest Landsat pixels) was used to assess the accuracy of the final forest map circa 2016 (Fig. 7b). The result confusion matrix is displayed in Table 2. The final forest map circa 2016 has a user’s accuracy of 99.2% and a producer’s accuracy of 88.7% for forest. As for
based on the aforementioned approach (Section 2.5) and the final forest coefficient of 0.87 also suggests a perfect agreement between the final forest map circa 2016 (Section 3.2). It shows an evergreen forest area of 113,861 km², accounting for 38.8% of the total forest area (293,648 km²) in the SGP. It is clear that most of the evergreen forest distributes in central and east Texas. There is also a large patch of evergreen forest in southeast Oklahoma. Some small patches of evergreen forest are scattered across the rest of Oklahoma and eastern half of Kansas.

### Table 3

<table>
<thead>
<tr>
<th>Reference classification</th>
<th>Evergreen forest</th>
<th>Other land cover types</th>
<th>Total</th>
<th>Producer's accuracy</th>
<th>User's accuracy</th>
<th>Overall accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evergreen forest</td>
<td>506</td>
<td>53</td>
<td>559</td>
<td>90.5%</td>
<td>97.3%</td>
<td>96.8%</td>
</tr>
<tr>
<td>Other land cover types</td>
<td>14</td>
<td>1539</td>
<td>1553</td>
<td>99.1%</td>
<td>96.7%</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>520</td>
<td>1592</td>
<td>2112</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

non-forest land cover types, it has a user’s accuracy of 88.0% and a producer’s accuracy of 99.2%. The overall accuracy is 93.5%. The Kappa coefficient of 0.87 also suggests a perfect agreement between the reference and classification data.

#### 3.4. Evergreen forest map circa 2016

The result evergreen forest map circa 2016 is displayed in Fig. 8, based on the aforementioned approach (Section 2.5) and the final forest map circa 2016 (Section 3.2). It shows an evergreen forest area of 113,861 km², accounting for 38.8% of the total forest area (293,648 km²) in the SGP. It is clear that most of the evergreen forest distributes in central and east Texas. There is also a large patch of evergreen forest in southeast Oklahoma. Some small patches of evergreen forest are scattered across the rest of Oklahoma and eastern half of Kansas.

#### 3.5. Validation of the evergreen forest map circa 2016

The reference data of 1069 non-forest Landsat pixels, 559 evergreen forest pixels, 484 deciduous forest pixels (Section 2.2.4) was used to assess the accuracy of the result evergreen forest map circa 2016 (Fig. 8). The confusion matrix is displayed in Table 3. As it is shown, the evergreen forest map has a user’s accuracy of 97.3% and a producer’s accuracy of 90.5% for evergreen forest. For land cover types other than evergreen forest, it has a user’s accuracy of 96.7% and a producer’s accuracy of 99.1%. The overall accuracy of the evergreen forest map is 96.8%. The Kappa coefficient of 0.92 also suggests excellent accuracy of the evergreen forest map.

### 4. Discussion

#### 4.1. Distribution of forest and evergreen forest in the SGP

The distribution of forest in the SGP is the result of the combined force of climate, land use, and woody plant encroachment. As shown in Fig. 7b, the vast majority of forest distributes in southeast SGP. This subregion tends to have higher precipitation and higher surface soil moisture, which to some degree could be associated with the opposite terrain pattern (Fig. S1) (Gu et al. 2021). This coincidence is a reflection of the critical role of precipitation in forest distribution at broad scale (Staver et al. 2011; Hansen et al. 2013). Nevertheless, the forest distribution does not strictly follow the precipitation pattern. The factor of land use (e.g. urban development, pasture) may explain low forest area in some places of southeast Texas. The widespread presence of cultivated cropland and grassland in central Kansas and west Oklahoma can account for the very sparse forest area there to a considerable degree (Fischer et al. 2014). The third factor to consider is woody plant encroachment history. Central Texas has medium level precipitation but dense forest, while east Kansas has high precipitation but low forest area. This is largely because the encroachment in Texas dates back to early twentieth century, while it starts more recently in Kansas (Box 1967; Engle et al. 2008; Twidwell et al. 2016).

As for evergreen forest, there are three major clusters located respectively in central Texas, east Texas, and southeast Oklahoma (Fig. 8). The cluster of evergreen forest in central Texas agrees very well with the reported distribution of junipers (Lyons et al. 2009). The presence of this cluster is mainly caused by long-term encroachment. The other two clusters of evergreen forest in east Texas and southeast Oklahoma are primarily cultivated pine trees (Weng et al. 2018; Shepard et al. 2021). They are generally managed through plantation and are an important source of commercial softwood production in the USA (Edgar et al., 2014; Osvalt et al. 2019). Other than that, small patches of evergreen forest can be found northward in the SGP, due to the steady expansion of junipers (Ratajczak et al. 2016; Twidwell et al. 2016; Wang et al. 2018).

The result forest map and evergreen forest map can meet a range of needs for follow-up research and application. Firstly, forest cover is an important parameter in forest biomass modeling (Matasci et al. 2018). It could also be involved in assessing the impact of woody plant encroachment on climate, evapotranspiration and soil moisture (Cui et al. 2020, 2021; Wang et al. 2021). Other than that, the result evergreen forest map can be used as a baseline to develop historical evergreen forest maps, and consequently to some extent trace the trend of woody plant encroachment across the SGP (Wang et al. 2017, 2018).

#### 4.2. Complementarity of PALSAR-2 and Landsat 8 data in forest mapping

This study exhibits the complementarity of PALSAR-2 and Landsat 8 data in regional scale forest mapping (Zehmnan et al. 2015). While the structure-sensitive SAR data (PALSAR-2) indiscriminately identifies both forest (above 2 m in height) and other land cover types with similar backscatter signal as forest, the phenology-sensitive optical data (Landsat 8) can filter out the non-forest land cover types (Qin et al. 2015, 2016, 2017; Li et al. 2021). In this study, those masked non-forest land cover types are represented by red color in Fig. 6. Most of them occurs with the rocky surface scattered with sparse shrub in southwest Texas (Fig. S5). The additional occurs in the metropolitan areas of Oklahoma City, Dallas-Fort Worth, and Houston, which are featured by dense building.

The PALSAR-2/Landsat 8 based forest map (Fig. 7a) contains certain amount of commission error from cropland and emergent herbaceous wetland, by reference to NLCD2016 data. These two types of commission error occupy an area of 15,179 km², accounting for 4.9% of the...
mapped forest area (308,827 km$^2$). According to Cropland Data Layer 2016, the misclassified cropland is primarily corn in northwest SGP and sugarcane in southern-most SGP. These misclassification is probably because that the high density of the crop and herbaceous vegetation results in high biomass level at pixel scale (25 m) comparable to that of forest, and the L-band backscatter is strongly correlated with above-ground biomass (Carreiras et al. 2012; Nesha et al. 2020). At the same time, these two land cover types are supposed to have high annual $\text{NDVI}_{\text{max}}$ values (>0.5). This study relied on the ancillary data of NLCD2016 to minimize these two types of commission error, to produce the final forest map circa 2016 (Fig. 7b). It highlights the utility of NLCD2016 in land cover mapping study. Whereas it is our plan to develop independent algorithms to address these commission issues in future.

In the final forest map circa 2016 (Fig. 7b), 8 out of the 9 commission error are from shrubland, while the other one is from building (house with lawn). As for shrubland, the misclassification reason should be similar to that of cropland and emergent herbaceous wetland. For the pixel covering house and lawn, it is probable that while the house contributes similar PALSAR-2 backscatter signal as forest, the lawn raises its annual $\text{NDVI}_{\text{max}}$ above the threshold (0.5). In terms of omission error, 62 out of the 484 (12.8%) deciduous forest validation pixels were missed, while 50 out of 559 (8.9%) evergreen forest validation pixels were missed. The lower omission rate of evergreen forest could be partially credited to its relatively stable foliage, leaf water and soil moisture throughout the year (Xiao et al. 2009), and consequently less effect of PALSAR-2 acquisition date (season) (Fig. S6) on its backscatter signal (Huang et al. 2021).

### Table 4
Forest area and evergreen forest area from different datasets (unit: km$^2$).

<table>
<thead>
<tr>
<th>Region</th>
<th>This study Forest</th>
<th>Evergreen</th>
<th>NLCD2016 Forest</th>
<th>Evergreen</th>
<th>FNF2016 Forest</th>
<th>Evergreen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kansas</td>
<td>21,362</td>
<td>535</td>
<td>12,297</td>
<td>59</td>
<td>13,278</td>
<td>N/A</td>
</tr>
<tr>
<td>Oklahoma</td>
<td>64,118</td>
<td>8,519</td>
<td>48,957</td>
<td>7,101</td>
<td>51,697</td>
<td>N/A</td>
</tr>
<tr>
<td>Texas</td>
<td>208,168</td>
<td>104,807</td>
<td>84,489</td>
<td>43,305</td>
<td>143,902</td>
<td>N/A</td>
</tr>
<tr>
<td>Total</td>
<td>293,648</td>
<td>113,861</td>
<td>145,743</td>
<td>50,465</td>
<td>208,877</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Fig. 9. Comparison of forest maps from this study, NLCD2016, and FNF2016 at 4 sample sites.
4.3. Success of the new approach for evergreen forest and deciduous forest separation

This study demonstrates the potential of the new approach - threshold of seasonal NDVI change (0.3) - in identifying evergreen forest out of forest map over the broad environmental gradients (e.g. arid, semiarid, mesic) of the SGP (Fig. S1). Table 3 shows very good user’s, producer’s and overall accuracy for the result evergreen forest map circa 2016 (Fig. 8). Among the very small amount of misclassifications in the evergreen forest map, the vast majority were inherited from the forest map (Fig. 7b). In other words, the new approach developed in this study performed perfectly in separating evergreen forest from non-evergreen forest.

Out of the 559 evergreen forest validation pixels, 509 were captured by the forest map, among which 506 (99.4%) were successfully identified as evergreen forest. On the other side, out of the 484 deciduous forest validation pixels, 422 were captured by the forest map, among which only 8 (1.9%) were misclassified as evergreen forest. All these 8 commission error occur in central to south Texas, where the relatively short and warm winter season is partly to blame for the misclassification. Consequently, the development of the new approach enriches the existing literature on evergreen forest mapping (Xiao et al. 2009; Wang et al. 2017, 2018).

Fig. 10. Comparison of evergreen forest maps from this study and NLCD2016 at 4 sample sites.

4.4. Comparison of forest area and evergreen forest area with other products

As a comparison, a set of forest map and evergreen forest map for the SGP was derived from NLCD2016 (Fig. S7). The Global PALSAR-2/PALSAR Forest/Non-Forest Map of 2016 (hereinafter referred to as FNF2016) was also clipped to the SGP (Fig. S8). The forest area and evergreen forest area of the SGP from this study, NLCD2016, and FNF2016 are summarized at state level in Table 4. As expected, this study shows much higher forest area and evergreen forest area in all the three states. The reason is because while NLCD2016 and FNF2016 apply a forest definition of tree height above 5 m, this study adopts a more inclusive forest definition (Australian) of tree height above 2 m. It proves the additional value of this study’s forest map and evergreen forest map in monitoring woody plant encroachment, since many of the encroaching trees do not reach 5 m high. It also demonstrates the capacity of PALSAR-2 L-band in capturing trees of various heights (above 2 m) simultaneously, as well as the efficiency of the new approach in separating evergreen forest from deciduous forest.

As a demonstration, forest maps from this study, NLCD2016, and FNF2016 were compared at four sample sites of 1 km by 1 km, in reference to 1 m resolution digital orthophotos from NAIP (Fig. 9). The center coordinates of the sample sites A, B, C, and D are (30.445501, −94.088027), (33.154271, −99.399227), (28.032242, −99.058177), and (34.90685, −101.15419), while the acquisition dates of the corresponding orthophotos are 09/21/2016, 10/02/2016, 09/19/2016, and
10/04/2016, respectively. Both the orthophotos and forest maps are displayed in geographic coordinate system of World Geodetic System (WGS) 1984.

Similarly, evergreen forest maps from this study and NLCD2016 were examined for another four sample sites, in comparison to 1 m resolution digital orthophotos (Fig. 10). The center coordinates of the sample sites E, F, G, and H are (37.428547, −98.877772), (34.33655, −95.33615), (29.643552, −100.28300), and (30.690252, −104.160151), while the acquisition dates of the corresponding orthophotos are 07/10/2017, 10/01/2017, 10/11/2016, and 09/09/2016, respectively. Figs. 9 and 10 further confirmed the high accuracy of the result forest map and evergreen forest map, as well as their inclusiveness in terms of forest height. These two figures also to some degree explain the large discrepancy in the amount of forest area and evergreen forest area among the three data sources (Table 4). Location of the 8 sample sites of Figs. 9 and 10 in the SGP can be found in Fig. 59.

5. Conclusion

In conclusion, this study generated the first map of forest, especially evergreen forest above 2 m in height, at 30 m resolution for the southern Great Plains. The forest distribution and underlying reason were analyzed. The result forest map can be an important input for the retrieval of biophysical parameters (e.g., biomass), while the evergreen forest map can be used to trace back the encroachment pattern across the SGP over past several decades. This study also proved the complementarity of PALSAR-2 and Landsat data in mapping forest above 2 m in height. It paves the way to develop time series forest and evergreen maps at global scale in future, and subsequently estimate forest loss and gain over time, as more L-band SAR data (e.g. NASA-ISRO Synthetic Aperture Radar) and optical data (e.g. Landsat 9) are becoming available. Other than that, this study developed a simple but robust approach to separate evergreen forest from deciduous forest over broad environmental gradients.

CRediT authorship contribution statement

Xuebin Yang: Conceptualization, Methodology, Formal analysis, Validation, Writing – original draft. Xiangming Xiao: Conceptualization, Supervision, Funding acquisition, Writing – review & editing. Yuanwei Qin: Writing – review & editing. Jie Wang: Writing – review & editing. Kevin Neal: Validation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study is supported in part by research grants from the US National Science Foundation EPSCoR program (IIA-1920946, IIA-1946093).

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jag.2021.102578.

References


Huang, X., Ziniti, B., Cosh, M.H., Reba, M., Wang, J., Torbick, N., 2021. Field-scale soil moisture retrieval of biophysical parameters (e.g. biomass), while the evergreen forest map can be used to trace back the encroachment pattern across the SGP over past several decades. This study also proved the complementarity of PALSAR-2 and Landsat data in mapping forest above 2 m in height. It paves the way to develop time series forest and evergreen maps at global scale in future, and subsequently estimate forest loss and gain over time, as more L-band SAR data (e.g. NASA-ISRO Synthetic Aperture Radar) and optical data (e.g. Landsat 9) are becoming available. Other than that, this study developed a simple but robust approach to separate evergreen forest from deciduous forest over broad environmental gradients.

CRediT authorship contribution statement

Xuebin Yang: Conceptualization, Methodology, Formal analysis, Validation, Writing – original draft. Xiangming Xiao: Conceptualization, Supervision, Funding acquisition, Writing – review & editing. Yuanwei Qin: Writing – review & editing. Jie Wang: Writing – review & editing. Kevin Neal: Validation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study is supported in part by research grants from the US National Science Foundation EPSCoR program (IIA-1920946, IIA-1946093).

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jag.2021.102578.

International Journal of Applied Earth Observation and Geoinformation 104 (2021) 102578
and interoperability in the context of a large-scale operational forest monitoring system. Remote Sens. Environ. 156, 335–349.


