Early detection of wildfire risk in the Great Plains: merging machine learning, landscape metrics, and rich data sources

by

Brynn Noble

B.S., Eastern Michigan University, 2020

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Division of Biology
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2023

Approved by:

Major Professor
Zak Ratajczak

Copyright

Abstract

The encroachment of woody plants is rapidly shifting tallgrass prairie into evergreen dominated ecosystems, mainly due to exclusion of fire. This increase in woody vegetation increases the potential for forest crown fires, specifically due to expansion of native eastern red cedar (*Juniperus virginiana*; henceforth, ERC), which are more dangerous due to their ability to spread much faster and cast embers far beyond the edge of fires. Many of the places where fire is being excluded are areas of high population density, potentially causing eastern red cedar to become dense surrounding residential areas. In drier and variable climates like the Central Great Plains, the question is not if, but when conditions will allow wildfires to spread. The goal of this project was to determine the spatial variability in forest fire risk in Manhattan, Kansas, as an emerging semi-urban zone that exemplifies exurban expansion into the remaining grasslands of the central Great Plains. This thesis assesses two key questions: 1) how effective are two of the U.S. government's USDA National Agriculture Imagery Program (NAIP) and National Ecological Observation Network (NEON) products for classifying grass-shrub-tree mosaics? and 2) is there an emerging wildland urban interface (WUI) forming around Manhattan KS and if so, does it have high wildfire risk? For chapter 2, we compared accuracies of land use maps created from aerial imagery from two freely available government sources (NAIP and NEON) and two commonly used machine learning techniques (random forests and support vector machines). NEON provides a much greater suite of data products, including hyperspectral and light detection and ranging (LiDAR), but NAIP covers a much larger area. We found that land cover maps created using NEON inputs were more accurate and relied almost entirely on LiDAR. NAIP created maps, however, severely undercounted ERC, indicating that land cover maps created on a larger scale (outside of NEON extent) need some other inputs to accurately detect ERC. We also found very little difference in accuracy between machine learning methods, but random forests ran the model in substantially less time than support vector machines. For chapter 3, we classified land cover using NAIP imagery, aerial imagery captured in the winter, and random forests. We then used this land cover map to analyze the extent and spatial patterns of ERC in Manhattan and thirteen neighborhoods, representing approximately 11,261 homes and out-dwelling units (structures from hereon). Structures in each neighborhood were identified

using FEMA USA Structures polygons. Landscape metrics were calculated based on an 800m buffer of each neighborhood. We found that ERC currently covers 9.1% (2,062 ha) of Manhattan, and ranges from 5-23% cover across neighborhoods. There is currently low connectivity between eastern red cedar patches but high cohesion, meaning that patches of ERC are growing close together but not touching yet. However, the gaps between ERC patches are small enough to disappear in coming years due to the speed of encroachment. We also calculated number of houses within different distances to ERC patches based on three levels of danger: direct flame (within 4m of houses), extreme radiant heat (within 20m of houses), and embers (within 800m of houses). We also looked at three patch sizes within each of those distances: patches $\geq 10m^2$, $\geq 1000m^2$, and $\geq 5000m^2$. All thirteen neighborhoods have over 50% of houses within 4m of ERC patches $\geq 10m^2$, and ten neighborhoods have 75% of houses within 4m of ERC patches $\geq 10m^2$. This indicates that a substantial number of homes are in danger of damage from direct flames of wildfires. Furthermore, seven neighborhoods have 100% of houses within 800m of ERC patches $\geq 5000m^2$, and four more have over 75% of houses within 800m of ERC patches $\geq 5000m^2$, signifying that almost all houses in most neighborhoods are within falling distance of embers and in danger of a spot fire. Therefore, if a wildfire breaks out in or around Manhattan, most structures could be in danger from either from the fire directly, or through rouge embers causing spot fires unless preventative measures are taken.

Table of Contents

List of Figures ... vii
List of Tables ... viii
Acknowledgements ... ix
Chapter 1 - Introduction ... 1
Chapter 2 - NEON's LiDAR increases woody plant detection in a grassland using machine learning ... 8
Abstract ... 9
Introduction ... 11
Methods ... 14
Site Description ... 14
Imagery ... 15
NAIP Imagery ... 16
NEON Imagery ... 16
Machine Learning Methods ... 17
Support Vector Machines ... 17
Random Forests ... 18
Training Data ... 19
Accuracy Assessment ... 20
Run Time and Other Logistics ... 20
Results ... 21
Importance of Different Data Inputs ... 22
Model Run Time and Other Logistics ... 22
Discussion ... 23
Conclusion ... 26
Acknowledgements ... 26
Figures ... 27
Tables ... 31
Chapter 3 - Emergence of encroached wooded WUIs in grassland settlements: An analysis of changing fire risk ... 40

Abstract ... 41
Introduction ... 43
Methods ... 45
Study Site ... 45
Creating Land Use Land Cover Map ... 46
Fire Risk ... 48
Neighborhood Delineation ... 48
Eastern Red Cedar Encroachment ... 48
Quantifying Homes in Danger ... 49
Results ... 51
Eastern Red Cedar Encroachment ... 51
Quantifying Homes in Danger ... 52
Discussion ... 54
Conclusion ... 57
Acknowledgments ... 57
Figures ... 59
Tables ... 63
Chapter 4 - Conclusion ... 64
References ... 68
Appendix A ... 76
Appendix B ... 79

List of Figures

Figure 1.1: Aerial image of an Eastern Red Cedar encroached neighborhood. 6

Figure 1.2: FEMA's version of defensible space around homes (FEMA 2008). 7

Figure 2.1: An estimate of the historical extent of arid and semi-arid Great Plains grasslands (light grey; based on EPA ecoregions), temperate Great Plains grasslands (dark grey; based on EPA ecoregions), the Flint Hills ecoregion (orange), and our study site (black star). The map inset shows an elevation map of our study site, Konza Prairie Biological Station. 27

Figure 2.2: Simple linear form of SVM. Adopted from Burges 1998, which is the type of SVM used in this study. 28

Figure 2.3: A) Aerial imagery from NAIP RGB (red-green-blue; naked eye view); B) visual of values in most important NAIP input for model training; C) visual of values most important NEON input for model training; D-F) visuals of RF classified models for all three images. 29

Figure 2.4: Average importance of each variable in Random Forest model building for: A) NAIP only; B) NEON only; and C) NAIP and NEON together. Mean GINI decrease essentially measures the amount of accuracy lost when that variable is removed. See Table 1 for input definitions. 30

Figure 3.1: Land Use Land Cover Map of Manhattan, Kansas 59

Figure 3.2: Map of neighborhoods used in study. Yellow polygons indicate exterior neighborhoods, light green polygons indicate edge neighborhoods, and dark green polygons indicate interior neighborhoods. 60

Figure 3.3: Percent cover of Eastern Red Cedar. "Average" is across all thirteen neighborhoods, "MHK" is across all Manhattan. Numbers represent neighborhoods. 61

Figure 3.4: Percentage of structures in each neighborhood A) within 4 m of ERC patches ≥ 10 m^2, B) within 20 m of ERC patches ≥ 1000 m^2, and C) within 800 m of ERC patches ≥ 5000 m^2. "Average" is across all thirteen neighborhoods, "MHK" is across all structures in Manhattan. Numbers represent neighborhoods. 62

List of Tables

Table 2.1: Summary of input variables and their descriptions used from each source. 31
Table 2.2: Summary of training points 32
Table 2.3: Accuracy results of all image and ML methods. 33
Table 2.4: Confusion matrix for SVM NAIP; columns represent class of training pixels, and rows represent class of model predicted pixels. 34
Table 2.5: Confusion matrix for RF NAIP; columns represent class of training pixels, and rows represent class of model predicted pixels. 35
Table 2.6: Confusion matrix for SVM NEON; columns represent class of training pixels, and rows represent class of model predicted pixels. 36
Table 2.7: Confusion matrix for RF NEON; columns represent class of training pixels, and rows represent class of model predicted pixels. 37
Table 2.8: Confusion matrix for SVM NAIP+NEON; columns represent class of training pixels, and rows represent class of model predicted pixels. 38
Table 2.9: Confusion matrix for RF NAIP+NEON; columns represent class of training pixels, and rows represent class of model predicted pixels. 39
Table 3.1: Summary of Manhattan neighborhoods used in this study.. 63
Table A.1: Best SVM model inputs calculated in R. Kernel refers to the method of data transformation. Degree is the degree of the polynomial kernel function. Gamma is the kernel coefficient which determines how much curvature will be in the decision boundary. Cost helps control error, a lower cost will accept a lower number of misclassified pixels. . 76
Table A.2: Time to train models. 77
Table A.3: Time to use models to predict (classify) entire study site.. 78
Table B.4: Confusion Matrix of LULC map accuracy 79
Table B.5: Landscape Metrics 80
Table B.6: Proportion of houses in each danger zone 81

Acknowledgements

This study would not have been possible without the help and support of many individuals. First and foremost, I'd like to thank my advisor, Dr. Zak Ratajczak, for letting me pursue this project even though neither of us had any background in machine learning or fire behavior. He helped me become a better and well-rounded scientist by teaching me new skills, enhancing my writing ability, challenging me to think about problems in new ways.

I would also like to thank my committee members, Dr. Dan Sullins and Dr. Audrey Joslin, for their time, expertise, and helpful inputs. They provided critical feedback and advice on the scope and direction of my projects, as well as provide helpful comments on my manuscripts.

I'd also like to acknowledge the many Division of Biology faculty for teaching me skills that helped me grow as a scientist and as a person. Many thanks to Dr. Kim With for teaching me many of the landscape ecology skills necessary for completing this project. Thanks also to Dr. Michi Tobler for teaching me how to make fantastic visuals and create great PowerPoint presentations.

Special thanks to my field technicians, Dylan Darter and Jaclyn Perry, for collecting hundreds of training points for the models. It's an incredibly tedious and boring task to walk endless circles around trees, shrubs, and other land cover types, but they collected invaluable data used to create incredibly high accuracy land cover maps.

Lastly, I would like to thank my husband Sidney Noble for all his moral support. This project would not possible without his encouragement, unwavering support, and willingness to cook dinner every night while I was stuck in the office fixing code.

Chapter 1 - Introduction

Tallgrass prairies in the Central Great Plains are one of the most endangered ecosystems, with less than 1% of the historical range still existing, due to conversion to agricultural land and human development (Samson and Knopf 1994). The only remaining landscape of tallgrass prairie is the Flint Hills of eastern Kansas and Northern Oklahoma, which supports a large pasture-raised cattle industry, intertwined with expanding human settlements into intact grassland ecosystem ecosystems. In the tallgrass prairie that remains, the largest conservation concern is woody encroachment, which is the growth and expansion of woody plants into grassland ecosystems, including shrubs and native conifers such as Eastern Red Cedar (*Juniperus Virginiana*: henceforth, ERC). Expansion can occur at rapid speeds (Twidwell et al. 2013, Galgamuwa et al. 2020, Moser et al. 2013), turning an open prairie into a closed canopy forest in as little as 40 years in the absence of fire (Briggs et al. 2002).

Many impacts of woody encroachment are well-studied, including negative effects on species diversity (Swengel 1996, Lettow et al. 2018, Albrecht et al. 2016), bird populations (Engle et al. 2008, Lautenbach et al. 2017), and groundwater levels (Keen et al. 2022, Zou et al. 2018). ERC encroachment can also change the fire risk from prairie ground fires to forest crown fires due to their volatile nature (Twidwell et al. 2013). Crown fires are fires which exist in the canopy of trees and spread faster, travel further, and are more difficult to control than ground fires (Scott and Reinhardt 2001). Crown fire potential is based on a combination of factors: fuel availability, ladder fuels, canopy density, and weather (Scott and Reinhardt 2001). With branches that grow to the ground, ERC trees provide their own ladder to carry fire from the ground to the crown, and tend to grown into dense, closed-canopy stands, indicating high crown fire potential. My goal was to create a high-resolution land use land cover map in order to quantify the current level of ERC encroachment and to calculate the number of houses exposed to three different fire risks.

Most wildfires occur in the wildland-urban interface (WUI), where humans and human developments intermix with wildland vegetation, and are mostly located in forested ecosystems (Stein et al. 2013). Wildfires costs billions of dollars each year, including the cost of fire response and direct damage (UNEP 2022, Hurst 2022), and millions of dollars more spent on fire prevention and research to determine areas of high wildfire risk (Stein et al. 2013, UNEP 2022). For example, between 2002-2007, Florida spent an average of $500,000 per year on wildfire prevention education (Stein et al. 2013). However, most of the Central Great Plains are not considered part of the WUI (Stein et al. 2013) and are often left out of wildfire studies; but woody encroachment is quickly turning much of the Great Plains into woodlands (Briggs et al. 2002, Ratajczak et al. 2016). Therefore, instead of the common movement of people into the WUI, people are settling in grasslands and a fire prone WUI is potentially being created as humans suppress fires and woody fuel loads grow (Figure 1.1).

Defensible space, or the space around homes which has been modified to stop or slow the spread of wildfire, is often promoted as the best way to reduce risk of damage from wildfire. Defensible space generally refers to three different zones of space at different distances from the house, each with specific instructions to reduce risk of fire damage, such as xeriscaping or thinning trees (Figure 1.2; FEMA 2008, FireSmart Canada 2023). For example, the first zone closest to homes is advised to have no flammable objects, including wood decks, any flammable vegetation, and firewood (FEMA 2008). However, initial anecdotal observations of Manhattan show that there is little to no defensible space around residences (Figure 1.1). The UN Environmental Assembly calls on governments around the world to start spending more time and money on prevention and preparedness to reduce the costs of response and recovery (UNEP 2022); being prepared and taking precautions will be one of the most effective ways to reduce risks associated with wildfires as they continually increasing around the world (UNEP 2022). However, no preparation can or will be done if people are not aware of the risk (McCaffrey et al. 2011). Some areas similar to my study site have already acknowledged this new risk and have begun taking measures to reduce the risk by burning and removing large stands of ERC (Twidwell et al 2013). In order to quantify the current state of risk, we need to be able to map the current extent of ERC on the landscape, and especially near human developments.

Woody encroachment of ERC is already well underway in Texas and Oklahoma, is quickly spreading up through Kansas, and we expect this “green glacier” of trees to continue spreading north into Nebraska (Engle et al. 2008). Furthermore, the number of wildfires in Oklahoma is increasing each year (Donovan et al. 2017), and woody vegetation is more conducive to extreme wildfires (Donovan et al. 2020). Thus, the number of structures lost to wildfires is much greater in Texas and Oklahoma than Kansas (Stein et al. 2013). As ERC continues to encroach into Manhattan neighborhoods, we could see a similar increase in large wildfires and structures lost. Therefore, it is important to start developing monitoring tools that 1) enable proactive management of encroachment through prescribed fire and other techniques; 2) capture the actual extent of wildfire risk for human structures; and 3) engage with local stakeholders to communicate risk and learn from their challenges and experiences. The goal of this thesis is to advance objectives one and two by creating a high-resolution land cover map with high woody plant accuracy and determining the current extent of ERC in proximity to human structures.

Most studies looking at wildfires and many studies tracking woody encroachment use coarse resolution aerial imagery (>10 m^2). However, small patches of ERC (<10 m^2) can be dangerous in the wrong places and newly encroached areas tend to have many small ERC trees spread out that are missed in coarse resolution aerial imagery. Therefore, it is critical to classify ERC extent using finer resolution inputs. Another benefit of this early detection is that smaller trees can be controlled by the reintroduction of controlled surface fires (Briggs et al. 2002). The US government has been investing in fine resolution (< 2 m^2) remote sensing through United States Department of Agriculture (USDA) National Agriculture Imagery Program (NAIP) and more recently, the National Ecological Observatory Network (NEON), which provide different data products and each cost millions of dollars annually. In chapter 2 of this thesis, I compared two common methods of machine learning classification of land cover (random forests and support vector machines) factorially crossed with these two freely available remotely sensed platforms. I found that all models had very high overall classification accuracy (>91%), but NEON-based models correctly classified woody vegetation far better than NAIP models. Specifically, NAIP's accuracy for ERC was 55-83% compared to 78-89% in NEON models. I conclude that NAIP alone is not sufficient to accurately classify woody vegetation.

Using insights from chapter two, I used random forests and two large-scale remote sensing products (NAIP and Kansas NG911—a wintertime mapping campaign) to create a much larger land cover map of Riley County and surrounding areas. Using this data-product, my third chapter aimed to analyze the extent and spatial patterns of ERC in Manhattan and thirteen neighborhoods, representing approximately 11,261 homes and out-dwelling units (structures from hereon). Structures in each neighborhood were identified using FEMA USA Structures polygons. Landscape metrics were calculated based on an 800 m buffer of each neighborhood. I found that ERC currently covers 9.1% (2,062 ha) of Manhattan, and ranges from 5-23% cover across neighborhoods. There is currently low connectivity between ERC patches but high cohesion, meaning that patches of ERC are growing close together but not touching yet. However, the gaps between ERC patches are small enough to disappear in coming years due to the speed of encroachment and realistic assumptions of flame length under extreme drought conditions. I also calculated number of houses within different distances to ERC patches based on three levels of danger: direct flame (within 4 m of houses), extreme radiant heat (within 20 m of houses), and embers (within 800 m of houses). I also looked at three patch sizes within each of those distances: patches ≥ 10 m^2, ≥ 1000 m^2, and ≥ 5000 m^2. These patches represent different types of risk, where small patches are more likely to be near homes, but less likely to burn in crown fire complexes. Averaged across all thirteen neighborhoods, I found that 82.4% of houses are in danger of direct flames, 20.3% are in danger of radiant heat, and 89.7% of houses are in danger of embers. This indicates that a substantial number of homes are in danger of damage from direct flames of wildfires, and almost all houses in most neighborhoods are within falling distance of embers and in danger of a spot fire. Therefore, if a wildfire breaks out in or around Manhattan, most structures could be in danger from either from the fire directly, or through rouge embers causing spot fires unless preventative measures are taken.

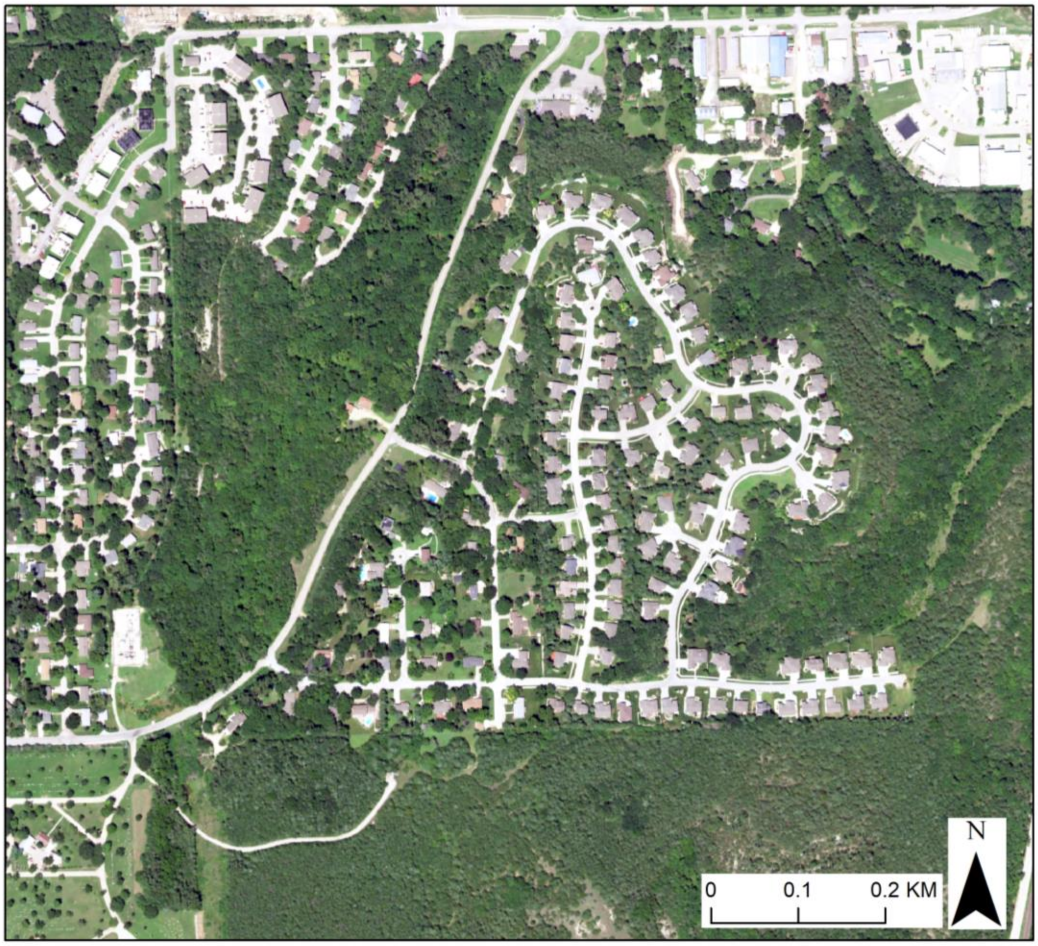

Figure 1.1: Aerial image of an Eastern Red Cedar encroached neighborhood.

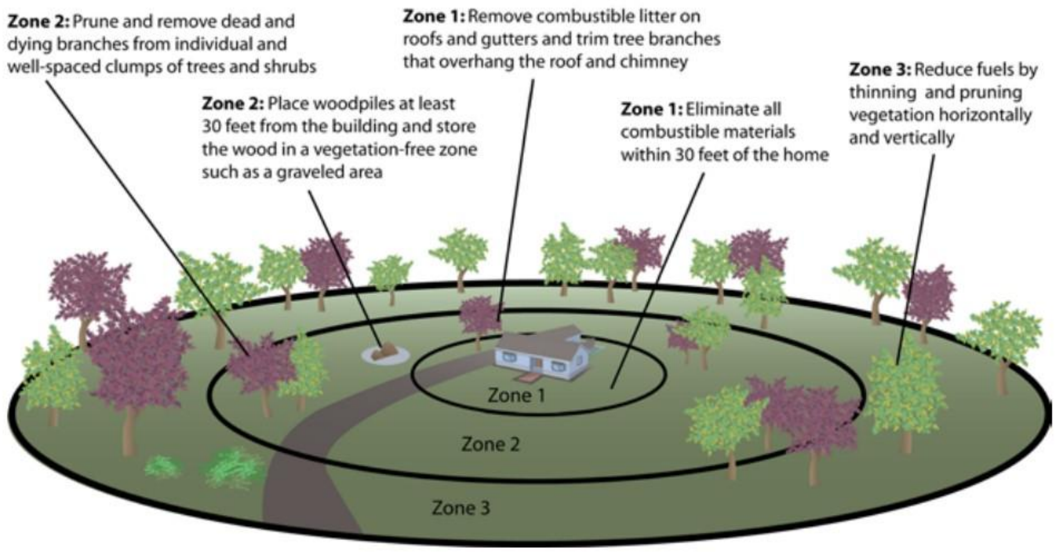

Figure 1.2: FEMA's version of defensible space around homes (FEMA 2008).

Chapter 2 - NEON’s LiDAR increases woody plant detection in a grassland using machine learning

Please note: This paper is formatted for Ecosphere journal

Abstract

Woody encroachment, or invasion of woody plants, is shifting many grasslands and savannas into shrub and evergreen dominated ecosystems. Tracking the pace and extent of woody encroachment is difficult because shrubs and small trees are much smaller than the coarse resolution of common remote sensing platforms (> 10 m^2) and ground-based approaches are slowed by the impassibility of encroaching woody thickets. However, the US government has been investing in fine resolution (< 2 m^2) remote sensing through United States Department of Agriculture (USDA) National Agriculture Imagery Program (NAIP) and more recently, the National Ecological Observatory Network (NEON), which provide different data products and each cost millions of dollars annually. We compared two common methods of machine learning classification of land cover (random forests and support vector machines) factorially crossed with these two freely available remotely sensed platforms to determine if and how much NEON adds to classification accuracy. All models had very high overall classification accuracy (>91%), with the NEON-based models a few percent more accurate than NAIP. A model using both inputs had the highest accuracy. This was mostly due to the models correctly classifying non-woody vegetation. However, there were differences between accuracies of NAIP and NEON models for woody vegetation: compared to NEON, NAIP's accuracy was 55-83% compared to 78-89% for evergreen trees, 83-93% compared to 96-98% for shrubs, and 78-91% compared to 94-97% for deciduous trees. The NEON-based models rely on canopy height (LiDAR) to make classifications, whereas the several bands of light make similar contributions to accuracy in the NAIP models. Finally, models using the same data sources were nearly identical in accuracy, indicating no important difference between random forests and support vector machines in

classifying vegetation. We conclude that the addition of LiDAR through the NEON program will increase our ability to accurately track woody plant encroachment.

Introduction

Woody encroachment, or the expansion of woody plants in grasslands, is negatively affecting grasslands around the world (Briggs et al. 2002, Galgamuwa et al. 2020, Moser et al. 2013, Twidwell et al. 2013). The Flint Hills in Kansas and Oklahoma, United States is witnessing Tallgrass Prairie rapidly being converted to shrublands and woodlands of Eastern Red Cedar (*Juniperus virginiana*; henceforth, ERC) (Briggs et al. 2005, Engle et al. 2008, Meneguzzo and Liknes 2015, Ratajczak et al. 2014). When fire is excluded in tallgrass prairie, ERC can expand, becoming a closed canopy in as little as 40 years (Briggs et al. 2002). In the Flint Hills, the largest remaining landscape of tallgrass prairie, 45% of current grasslands are burned so infrequently that they are likely to transition to woodlands in the next ten to thirty years unless management practices change (Ratajczak et al. 2016). As woody species increasingly takeover grasslands, many grassland obligate species could decline further, such as Monarch butterflies (Swengel 1996), native bees (Lettow et al. 2018), Lesser Prairie Chickens (Lautenbach et al. 2017), and others (Albrecht et al. 2016). Woody encroachment can also have economic impacts by reducing freshwater recharge (Keen et al. 2022) and forage for commercial grazers (Anadon et al. 2014). One of the challenges in studying woody plant encroachment is logistical difficulties. Woody encroachment creates thickets of dense and often thorny vegetation, which can be difficult to pass through, and in some cases, is impassable without altering the vegetation itself. Therefore, remote sensing could allow grassland ecologists to study woody encroachment more accurately, more quickly, and at a larger spatial extent than is possible with on-the-ground approaches. In this study, our goal was to determine what combinations of machine learning approaches and data sources produce the most accurate and fastest combination for remote sensing of woody plant encroachment.

Remote sensing (RS) and machine learning (ML) have long been used to classify land use and land cover (LULC). Until recently, many studies on land-use classification using RS and ML are done at coarse resolutions of >10 m^2 (Allred et al. 2021, Galgamuwa et al. 2020, Kranjčić et al. 2019, Nguyen et al. 2020, Thanh Noi and Kappas 2018). However, woody encroachment can be difficult to track at coarse resolutions because shrubs and smaller trees are often smaller than the minimum grain-size of common satellite-derived data (e.g., >9 m^2; Whiteman and Brown 1998). The growing availability of higher resolution remote sensing from unmanned vehicles, low flying planes, and advanced satellites could rapidly improve our ability to remote sense shrubs and other forms of woody encroachment (Toth and Jóźków 2016). For instance, high resolution remote sensing was recently used to identify millions of small trees across Northern Africa—a region thought to house few trees, because more coarse resolution data-products could not identify the smaller trees of this region (Brandt et al. 2020).

The utility of ML and other classification methods might also be constrained by the types of data available. In the United States, the United States Department of Agriculture (USDA) National Agriculture Imagery Program (NAIP) has provided the most consistent, widespread, and freely available high-resolution RS data in the United States. This product was a large investment, with mixed impact (Maxwell et al. 2017). Recent investments from the U.S. National Foundation have begun to provide another set of high-resolution remote sensing data through the National Ecological Observatory Network’s (NEON) aerial observation platform (AOP) (reviewed by Nagy et al. 2021). NEON’s AOP covers a much smaller spatial extent than NAIP (81 total sites versus the entire continental U.S.), but provides a much wider range of data, including LiDAR, hyperspectral data, and a suite of derived products, such as estimated canopy height and canopy nitrogen. These additions are promising, but because their adoption by ecologists is still nascent, several questions remain unanswered: how accurate are NEON's derived products? Is the learning curve of using these products preventing widespread use? For instance, at one site, a recent study found that some derived NEON products were weakly correlated with ground-based measurements (Pau et al. 2022).

The number of machine learning approaches is large, growing, and becoming more accessible through a growing number of open-source methods and training vignettes. Most ML methods follow the same general workflow: 1) train the model with remote sensed inputs and polygons or pixels of known cover class types; 2) apply the model to a dataset of known points or polygons to assess accuracy; and 3) apply the model to an unknown dataset to classify unknown values (Kranjčić et al. 2019). There are many methods of ML available to classify RS aerial imagery, but random forests (RF) and support vector machines (SVM) have emerged toward the top of the field in the past decade (Sheykhmousa et al. 2020), so we compared these two.

Here we assess the value-added of NEON for remote sensing woody plant encroachment, over the more widely available and longer running NAIP program. This is a timely question given the cost of NEON ($469 million USD initially, with a current operating budget >$70 million USD per year; Mervis 2016, NEON FY 2022 Budget Request). While a larger number of data products and more complex machine learning methods might produce the most accurate product, the inputs are extremely data-heavy, require substantial computing power, and are not accessible to many users. We aim to determine the method and inputs necessary to maximize accuracy while limiting computational effort and using data products that would be available to a reasonably skilled graduate student, post-doc, or professor. Therefore, we restricted our use of NEON data to "off the shelf" data-products, such as NEON's data product "estimated canopy height” based on LiDAR. This decision was motivated by observations that a lack of certain computational skills might be impeding the wider usage of NEON remote sensing products (see Nagy et al. 2021 for a review). We performed a set of factorial analyses that explore the role of more rich data (NAIP RGB, NEON vegetation indices, and NEON LiDAR) and model sophistication (SVMs, RFs).

We hypothesized that:1) NEON would increase accuracy, primarily due to the addition of canopy height estimated using LiDAR; 2) shrubs would have the lower accuracy than grasses and trees, because their height and traits fall in between these two functional groups; 3) ERC would have high accuracy, given its unique leaf type compared to other plant functional groups we considered, which are all deciduous.

Methods

Site Description

Konza Prairie Biological Station (KPBS), is a National Science Foundation long-term ecological research (LTER) site with 3,487 ha of native unplowed tallgrass prairie located in the Flint Hills in northeastern Kansas (Fig. 2.1). KPBS has high seasonal variability with an average high of 26.6º C in July and -2.7º C in January. KPBS receives an average annual rainfall of 835 mm, with 75% falling during the growing season (April-October). The soils are non-glaciated with thin rocky upland soils (mostly from the Florence series), deeper lowlands (often from the Tully series), and complex benches, outcrops, and slopes that connect these two soil types.

KPBS is split into 60 different management units, with replicates spanning 1-, 2-, 3-, 4-, and 20-year fire frequencies, as well as ungrazed, grazed by bison, or grazed by cattle. These different treatments have created a mosaic of contrasting land covers, including areas that are dominated by herbaceous species, shrubs, deciduous trees, or evergreen trees. The herbaceous dominated areas can be floristically diverse, with high grass dominance in areas without bison or cattle, and mosaics of tallgrasses, short-grass grazing lawns, and patches of tall forbs in areas with bison and cattle. Areas dominated by shrubs typically have little to no herbaceous species (Briggs et al. 2002, Ratajczak et al. 2011, Ratajczak et al. 2014) and dominant shrubs (primarily the species *Cornus drummondii*) are all clonal, creating "islands" of ramets that can reach over 10 m in diameter. Heights range from <0.5 m tall for young clonal stems to over 3 m tall for older stems. At the lowest elevations and some intermittent streams, a full riparian forest has become established, dominated primarily by oaks (mostly Chinqapin oak, *Quercus muehlenbergii* and Burr oak, *Quercus macrocarpa*). Outside of these lowlands, the height and continuity of deciduous trees is lower, with species including honeylocust (*Gleditsia triacanthos*), red buds (*Cercis canadensis*), and several elm species. The only evergreen trees known to occur on site is ERC (*Juniperus virginiana*), which is primarily in areas without bison and without frequent fire.

Imagery

Two data sources were used for this project: USDA NAIP and NSF NEON. Each data source was tested alone and then together (NAIP+NEON) for a total of three models for each ML method, resulting in six models. Table 2.1 outlines the inputs used for each image. The images were taken in separate years, but between these two years there was no major change in climate (Ratajczak et al. 2022) and no major fires occurred. Therefore, major changes in vegetation between these two time periods is unlikely.

NAIP Imagery

An image was sourced from the USDA NAIP (USDA 2019a&b, the final product we used can be found at Noble and Ratajczak 2022). The image was captured on July 10, 2019 by low-flying aircrafts and has a resolution of 0.6 m^2. We used bilinear interpolation to transform each image to 2 m^2 pixels and snapped to a common grid with all other images and inputs.

The image sourced from NAIP contains nine bands, four are provided in the imagery (red, green, blue, infrared), and five more were calculated (red neighborhood, green neighborhood, blue neighborhood, infrared neighborhood, and normalized difference vegetation index [NDVI]; Table 2.1). Neighborhood calculations take the average value of all surrounding pixels. The red neighborhood calculation, for example, averages the redness of the pixels immediately surrounding each pixel. We added these neighborhood averages after a first application of machine learning found that some single pixels of deep shadows were misclassified as ERC.

NEON Imagery

NSF NEON imagery was captured in June 2020 by low-flying airplanes and has a resolution of 1 m^2, which was upscaled to the same resolution and grid as NAIP using bilinear interpolation (NEON 2020a&b, the final product we used can be found at Noble and Ratajczak 2022). NEON provides many data products which are derived from physical measurements. For example, NDVI is a product derived from dividing the difference between near-infrared (NIR) and red bands by the addition of NIR and red bands. NEON offers several of these derived products in addition to hyperspectral RS bands. We used 8 derived vegetation indices from this image: enhanced vegetation index (EVI), normalized difference nitrogen index (NDNI), normalized difference lignin index (NDLI), soil-adjusted vegetation index (SAVI), atmospherically resistant vegetation index (ARVI), NDVI, NDVI neighborhood, and canopy height (LiDAR; Table 2.1). Leaf area index was not included because it is calculated using SAVI. NEON's 10-cm RGB was unusable due to distortions along seamlines, however these distortions were not in the derived products.

Machine Learning Methods

Machine learning uses a small set of user inputs (training data) to learn and classify unknown data. We compared two different methods of supervised ML for this project, SVM and RF, as these are the most common methods used in RS today (Thanh Noi and Kappas 2018, Sheykhmousa et al. 2020).

Support Vector Machines

Support Vector Machines (SVM) are a supervised nonparametric classification technique which use a fixed optimal hyperplane to split the data into the desired number of discrete categories (Burges 1998). In the simplest form, a linear line separates two-dimensional data into two categories (Mountrakis et al. 2011; Fig. 2.2), but SVM are popular for their ability to work with high-dimensional data (Sheykhmousa et al. 2020).

Each pixel of a RS image is a series of numbers, one value from each input, which are mapped with each input as a new dimension during SVM model creation. The points which lie the closest to the hyperplane are support vectors and are the most important in determining the decision boundary. While many linear hyperplanes may exist in the data, SVM chooses the largest margin between points, allowing for some misclassification. In real life applications, many datasets do not have a linear break between data; SVM can easily circumvent this problem by shifting the data into a higher dimension using kernels, which separates the data even further to allow for a linear hyperplane to split the data. Furthermore, SVM are often used in the RS field for their ability to be accurate even with small training sets (Thanh Noi and Kappas 2018, Mantero et al. 2005), making them a strong choice for large study areas.

SVM were run in program R (v4.0.5; R Core Team 2021) using the 'e1071' package and model inputs were optimized using the 'best.svm' function (v1.7-6; Mayer et al. 2021; see Table A.1 for final model parameters).

Random Forests

RF are a non-parametric supervised ML technique. The building block of RF are decision trees, which use nodes to split data into smaller and smaller subsets to predict the pixel class. RFs use a bagging approach when building trees, where each decision tree is built with a random selection of input variables, creating a forest of different tree structures to limit overfitting (Evans et al. 2011). The user can set the number of decision trees for each model (ntree) and the number of input variables used to split each node (mtry), but many users rely on the default values of ntree (500) and mtry (square-root of number of inputs; Thanh Noi and Kappas 2018). RF make predictions based on majority voting; each individual decision tree outputs a predicted class and the class which is predicted the most times in the forest is the overall predicted classification (Sheykhmousa et al. 2020). For example, if a RF has 100 trees, and 76 of them predict a pixel as grassland, the model predicts that pixel to be grassland. Adding decision trees can improve accuracy, but it will increase the model run time and required computing capacity. Models with an excessive number of trees yield diminishing returns in accuracy. Lastly, RF models can determine the GINI decrease for each input, which measures the amount of accuracy lost with the removal of an input variable, indicating the relative importance of each.

RF models were run in program R (v4.0.5; R Core Team 2021) using the ‘randomForest’ package and model inputs were optimized using the ‘best.randomForest’ function (v4.6.14; Liaw and Wiener 2002). The NAIP+NEON model had an ntree of 500 and mtry of 4, and both single-source models (NAIP and NEON) had an ntree of 500 and mtry of 3.

Training Data

The study area had five categories of LULC: (1) grassland (herbaceous dominated areas); (2) shrubs; (3) deciduous trees; (4) ERC trees; (5) and other (roads, water, and buildings). Training datasets were created by a combination of ground-truth points collected using high-precision GPS units (with below 2 m error) and computer-drawn polygons. Ground-truth points data collection occurred from June to August 2021 and were collected using a random sampling approach, with a few locations where all vegetation within the area was sampled. Computer-drawn polygons were traced using a combination of the 2020 NEON RGB-10 cm^2 imagery and publicly available 1 m^2 RGB (a 2019 image from Maxar technologies available at Google Earth); neither of these images were used in the SVMs or RFs. Polygons were drawn in locations where species was confirmed in the field or where classes were obvious (in particular, buildings, water, and roads). A total of 3,635 training polygons were collected, resulting in 300,328 2 x 2 m pixels of known vegetation type, totaling 3.42% of the total area (Table 2.2). 70% of the points were used to train the models and the remaining 30% were held for evaluation of the models.

Accuracy Assessment

When assessing the accuracy of each model, four aspects are considered: producer accuracy (PA), user accuracy (UA), overall accuracy (OA), and Kappa. PA refers to the number of pixels correctly classified from the training data. In other words, the proportion of 'grassland' pixels in the evaluation data also appear as 'grassland' in the final classified image. UA refers to the number of pixels which are classified as, for instance, grassland and are actually grassland on the ground (i.e., in the training data). OA is an estimate of accuracy among all predicted cover types and is a ratio of the total number of correctly classified pixels to the total number of pixels. Lastly, the Kappa coefficient measures the accuracies by comparing the classification outcome versus randomly assigning values. Kappa ranges from -1 to 1, with 0 indicating that the model performed on par with randomly assigning values, <0 indicating that the model performed worse than random, and >0 indicating that the model performed better than random.

Run Time and Other Logistics

Run time can become a consideration for some machine learning approaches, requiring PIs and/or students to learn new techniques to complete more computationally intensive tasks. We recorded and report run time for both model training and extrapolation to the remainder of our site, to give an estimate of trade-offs between model accuracy versus computational efficiency. For reference, these models were run on a Dell XPS 8930, with relevant specifications of 64 GB RAM and an Intel® Core™ i9-9900K processor, with 3.6 GHz speed, 8 cores, and the ability to perform 16 threads. Note that program R runs most processes through the RAM.

Results

Our two modelling approaches (SVMs vs RFs) yielded nearly identical OA (<3.2% difference; Table 2.3). In general NEON performed better than NAIP, and the NAIP+NEON models were more accurate than either single-source model. NAIP+NEON had almost no difference between RF (OA: 98.4%, Kappa: 0.973) and SVM (OA: 98.2%, Kappa: 0.969; Table 2.3). NEON also had very little difference in accuracies between RF (OA: 97.9%, Kappa: 0.964) and SVM (OA: 97.3%, Kappa: 0.953; Table 2.3). NAIP had the largest difference between classification method, with RF (OA: 94.3%, Kappa: 0.899) a few percentage points more accurate than SVM (OA: 91.2%, Kappa: 0.842; Table 2.3). Classified maps comparing all three data sources are shown in Fig. 3d-e, where the ML method was RFs for all panels.

For all models, the grassland and “other” categories had values of 95% or above for both UA and PA for all combinations of machine learning methods and data sources (Table 2.3). However, the three categories of woody plants (shrubs, deciduous trees, and ERC) were more difficult to classify accurately. Shrubs and deciduous trees had very high PA and UA accuracies (>94%) in NEON and NAIP+NEON, but NAIP alone performed slightly worse, with PA of 83-93% for shrubs and 78-82% for deciduous trees, and UA of 85-89% for shrubs and 85-91% for deciduous trees (Table 2.3). Deciduous trees were most often misclassified as shrubs, and shrubs were most often misclassified as grassland (Tables 2.4-2.9). All models had the lowest accuracies for ERC, but NAIP performed particularly poorly. PA and UA for NAIP-based ERC were both low, but PA was especially low at 55-61%, compared to 83% UA (Tables 2.3-2.5). This low PA means that the models are undercounting ERC by classifying it as something else (mostly deciduous trees), rather than misclassifying other categories as ERC (Tables 2.4 & 2.5).

Importance of Different Data Inputs

For NAIP, the most valuable input variable was red neighborhood (visualized in Fig. 2.3b), followed by blue neighborhood, red, and blue (Fig. 2.4a). The most valuable input variable for NEON was canopy height (visualized in Fig. 2.3c), followed very far behind by NDVI and NDVI neighborhood (Fig. 2.4b). NAIP+NEON also largely relies on LiDAR, followed by five inputs from NAIP: red neighborhood, blue neighborhood, red localized, blue localized, and green neighborhood. In the combined model, NEON derived products (other than canopy height) provided very little GINI decrease (Fig 2.4c).

Model Run Time and Other Logistics

The amount of time it took to train each model varied greatly, with the shortest time at only 23 minutes for RF NEON, and the longest took 4 hours and 49 minutes to run SVM NAIP (Table A.2). More specifically, RF and SVM models had large differences between training run times, with the longest RF model taking 1 hour and 5 minutes, and the shortest SVM run time at 1 hour and 37 minutes (Table A.2). The time it took the models to predict the entire study site was drastically different between RF and SVM; RF classification took 7:20-10:43 minutes, and SVM classification took 2:05-6:15 hours (Table A.3). For SVM, model run time was negatively correlated with accuracy, with the NAIP+NEON model taking the least amount of time to both train and classify, and was the most accurate (Tables 2.3, A.2, & A.3). However, for RF, the most accurate model (NAIP+NEON) took the longest time to train, but the shortest time to classify (Tables 2.3, A.2, & A.3).

Discussion

Until recently, NAIP was the only widely available high-resolution open source of aerial remote sensing in the U.S. We found that with a few manipulations (use of neighborhoods) and readily available ML methods, NAIP succeeds at identifying grasslands and to some extent shrubs and deciduous trees. However, additional data, such as NEON, greatly increased our ability to correctly identify all forms of woody vegetation and especially ERC. The addition of NEON also made classification less subject to choices of ML methods. Therefore, in the limited locations where NEON is available (81 total sites vs entire continental US for NAIP coverage), NEON is a potential replacement for NAIP or the two data-sources could be used synergistically—NEON for its addition of LiDAR and NAIP for its undistorted red, green, blue, and NIR.

Eastern Red Cedar is a native evergreen encroaching rapidly in tallgrass prairies, negatively impacting species diversity and ecosystem services (Briggs et al. 2002, Limb et al. 2010, Van Auken 2009, Zou et al. 2018). Detecting ERC is important for monitoring and managing the impacts of woody encroachment (Meneguzzo and Liknes 2015). However, ERC had the lowest accuracy in all models, with PA lower than UA (Table 2.3), indicating that ERC is being substantially undercounted (in some cases by almost half) because many ERC pixels are being classified as deciduous trees. This problem was particularly acute when we only used NAIP imagery (Tables 2.4 & 2.5). We hypothesized that ERC would have much higher accuracy, since it has a much different leaf structure and water content than other woody plants in the area—characteristics that are supposed to be measured by NEON derived products (e.g., NDVI, NDLI, NDNI). However, our hypothesis proved incorrect, which has implications for using these models to make predictions of the rate and volume of current and future woody encroachment. Other studies also had difficulty detecting ERC in aerial imagery, specifically when ERC was at low densities (Kaskie et al. 2019, Kaskie et al. 2022). At densities below 30%, Kaskie et al (2019) found <50% accuracy ERC classification using aerial imagery alone. However, using other predictor variables such as slope, aspect, and Euclidean distance to nearest ERC pixel, Kaskie et al (2022) was able to increase accuracy of ERC at low densities (<15%) to 84.7%. ERC density is low across our site, but accuracy was fairly high (78-94%; Table 2.3) in NEON models. Therefore, NEON appears to overcome some challenges of low-density ERC detection. However, more work needs to be done to determine better and more efficient methods to accurately classify ERC at low densities.

We found that NEON imagery adds a substantial amount of accuracy when using ML methods to classify woody vegetation. Comparing the two single-source models (NEON vs NAIP), NEON increased accuracy by 8-43% for ERC, 7-21% for deciduous trees, and 5-15% for shrub cover (measured as percent increases going from NEON to NAIP; Table 2.3). In both models that included NEON data, canopy height data (based on NEON LiDAR), was by far the most influential input for accurately classifying our vegetation classes when using RF models (Fig. 2.4), indicating that LiDAR alone seems to be responsible for the increase in accuracy for woody plants. This was somewhat surprising given that all estimated values of canopy height <2 m are truncated to a value of zero as part of NEON's data cleaning, and some areas dominated by woody vegetation on site are in the range of 1.5 to 2 m height. However, this is consistent with many other studies which also found that LiDAR-based canopy height was among or the most important input for accurately classifying vegetation types (Scholl et al. 2020, Pervin 2022) and that LiDAR increases classification accuracy (Bork and Su 2007, Jin and Mountrakis 2022). Jin and Mountrakis (2022) analyzed 37 studies which compared accuracy using multispectral imagery alone versus classification with LiDAR added and found increases in model accuracy for almost all studies, with accuracy increasing as much as 68%. This again points towards value added from NEON, but specifically from the addition of LiDAR, since NEON vegetation indices other than canopy height played very little role in increasing model accuracy in the NEON+NAIP model (Fig. 2.4). However, NEON also provides a full suite of hyperspectral data, which were not tested here, but could have yet more value added on top of NEON's LiDAR product.

Shrubs and small trees have been difficult to classify in the past because they are often smaller than the resolution of remotely sensed aerial imagery (> 9 m^2; Whiteman and Brown 1998), leading to undercounting of shrubs and small trees (Brandt et al. 2020). Thus, we had hypothesized that shrubs would be among our lowest accuracy by class. However, despite having a higher resolution than previous studies (2 m^2), it appears that the addition of LiDAR (canopy height) is important for overcoming challenges of shrub classification. Contrary to our hypothesis, we found high accuracy for shrubs in all models using LiDAR (96-98%; Table 2.3)—more accurate, in fact, than both deciduous and ERC trees. In the NAIP only models (without LiDAR), shrub accuracy was much lower (83-93%; Table 2.3). While NAIP accuracy is still fairly high, this indicates that LiDAR-based canopy height can boost our ability to detect shrubs and small trees.

Finally, we found that two commonly used ML techniques had mostly similar performance. RF models added a very small amount of accuracy (< 3.2%), but the runtimes of readily available RF model implementations (both training and predicting) were faster than readily available SVM models in program R (Table A.2 & A.3); prediction times for RF took minutes while SVM took hours. RF was also slightly less sensitive to the source of data, whereas SVMs performed quite poorly using only NAIP imagery. Finally, RF also have the advantage of easily parsing the importance of different input variables, which could again, be valuable for using multiple remote sensing platforms synergistically.

Conclusion

Management of woody plant encroachment, including mature ERC stands and clonal shrublands is costly in money, time, and effort (Bidwell et al. 2002). Timely detection of small individual shrubs and trees can allow managers to engage in preventative management while woody plants are still at low densities and less resistant to disturbances, such as fire. Tools like NEON's LiDAR increases accuracy of ERC classification, which can help implement management interventions and identify areas of elevated wildfire risk. Furthermore, increase in all class accuracies when using NEON creates more accurate overall vegetation mosaics, allowing other users to use these derived LULC maps for applications including hydrology (Keen et al. 2022) and habitat use (e.g. Silber et al. *In prep*).

Acknowledgements

We would like to thank my funding sources, the Kansas State University Division of Biology, the Kansas Agricultural Station, and the NSF-funded Konza Long-Term Ecological Research (LTER) grant. We would also like to thank all the amazing people who run Konza, including the prescribed burning volunteers, for helping to maintain the long-term experiments. Finally, a huge thank you to Dylan Darter and Jaclyn Perry, our field technicians who helped collect thousands of training points.

Figures

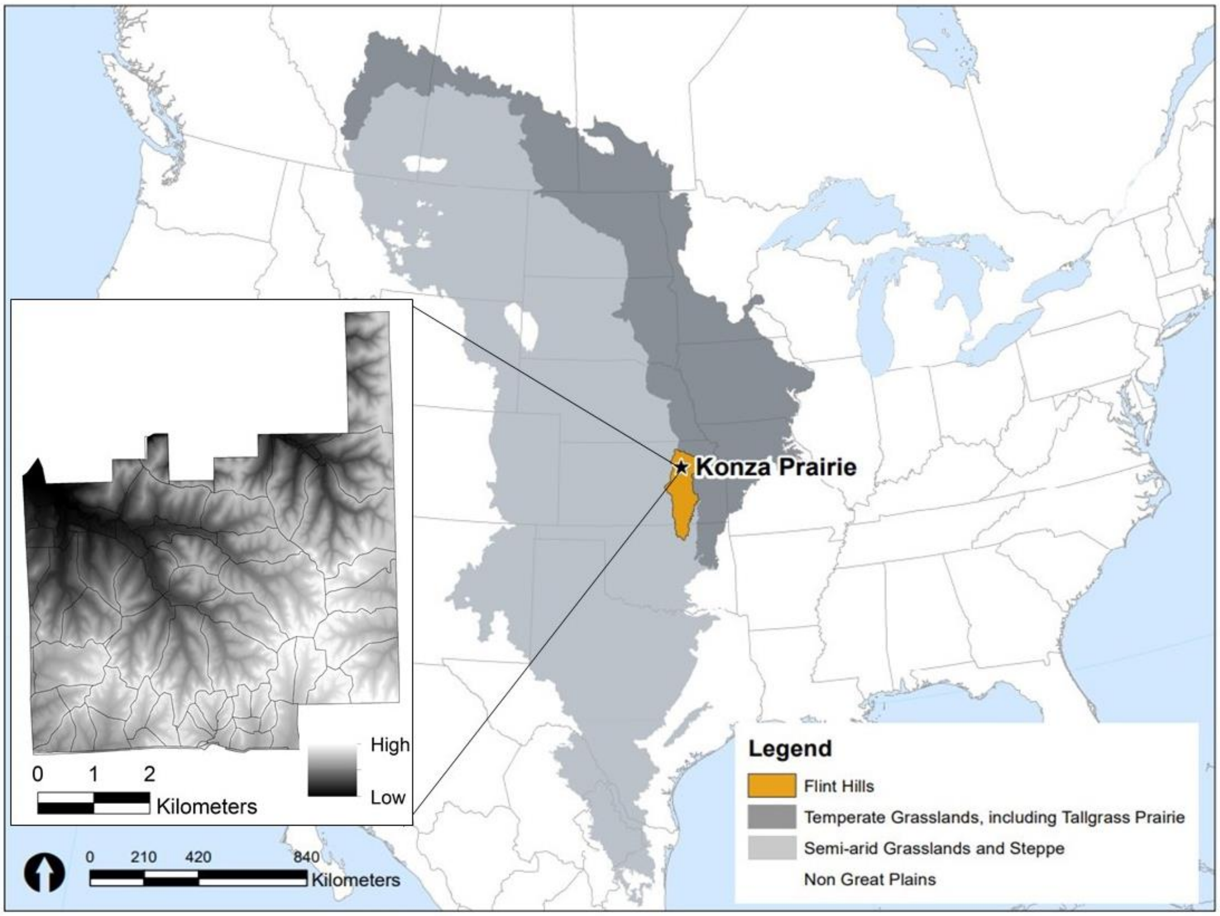

Figure 2.1: An estimate of the historical extent of arid and semi-arid Great Plains grasslands (light grey; based on EPA ecoregions), temperate Great Plains grasslands (dark grey; based on EPA ecoregions), the Flint Hills ecoregion (orange), and our study site (black star). The map inset shows an elevation map of our study site, Konza Prairie Biological Station.

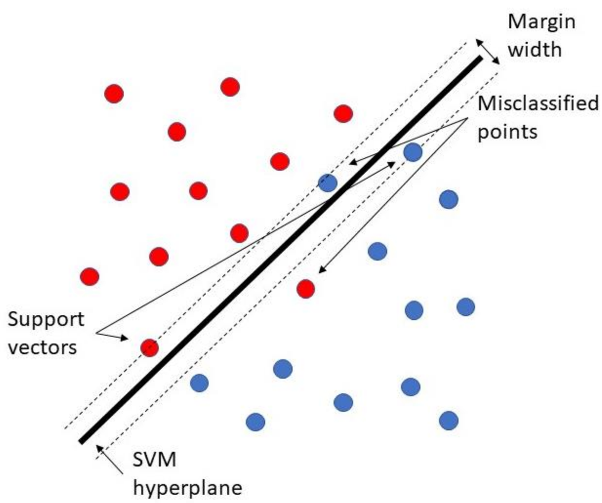

Figure 2.2: Simple linear form of SVM. Adopted from Burges 1998, which is the type of SVM used in this study.

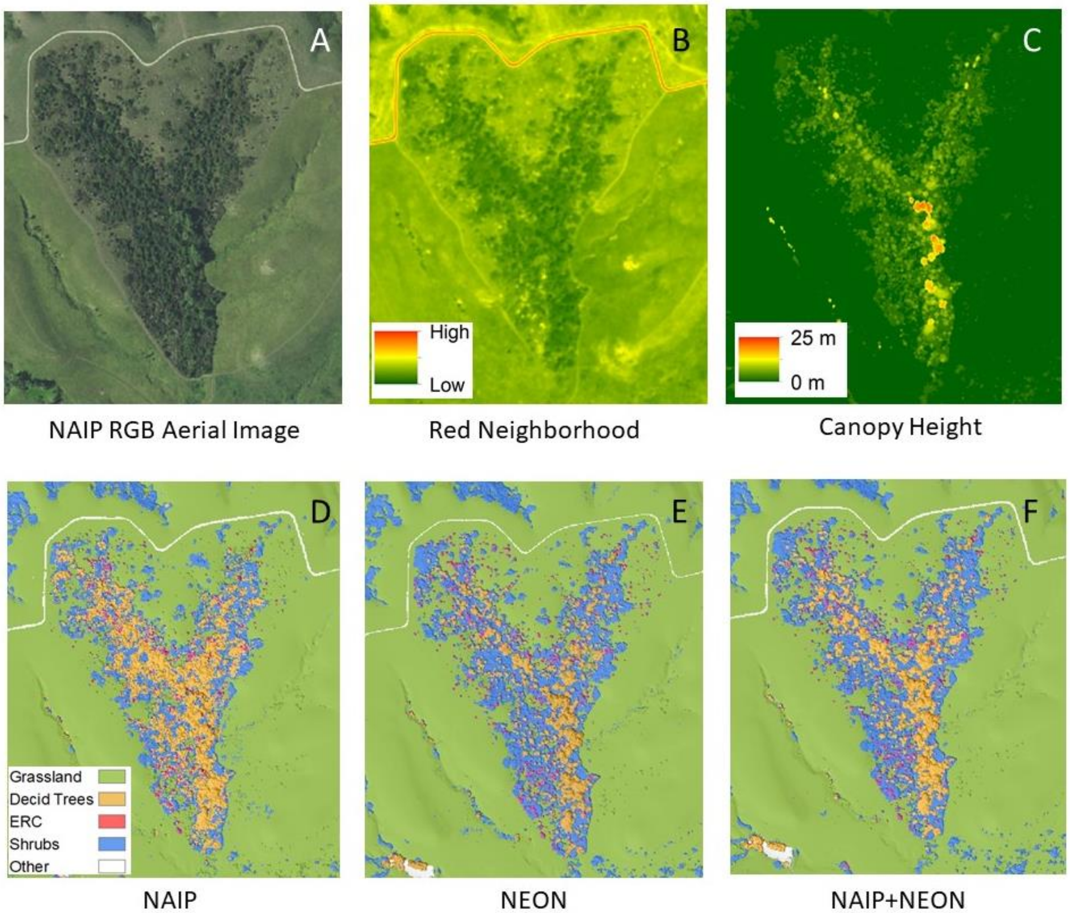

Figure 2.3: A) Aerial imagery from NAIP RGB (red-green-blue; naked eye view); B) visual of values in most important NAIP input for model training; C) visual of values most important NEON input for model training; D-F) visuals of RF classified models for all three images.

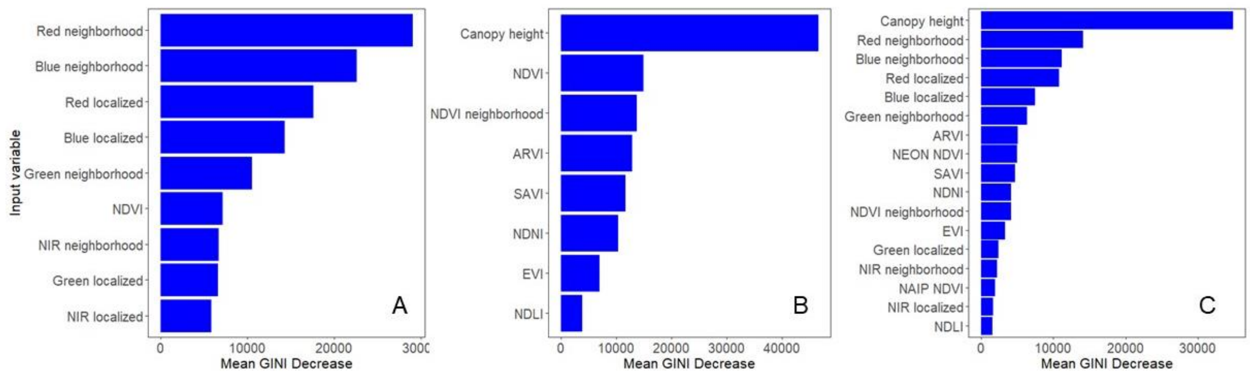

Figure 2.4: Average importance of each variable in Random Forest model building for: A) NAIP only; B) NEON only; and C) NAIP and NEON together. Mean GINI decrease essentially measures the amount of accuracy lost when that variable is removed. See Table 1 for input definitions.

Tables

Table 2.1: Summary of input variables and their descriptions used from each source.

Source	# bands	Inputs	Input Description
USDA NAIP	9	Red	Redness of each pixel
		Green	Greenness of each pixel
		Blue	Blueness of each pixel
		Red Neighborhood	Avg. redness of surrounding pixels
		Green Neighborhood	Avg. greenness of surrounding pixels
		Blue Neighborhood	Avg. blueness of surrounding pixels
		Near-infrared	Value of near-infrared wavelength
		Near-infrared Neighborhood	Avg. values of near-infrared of surrounding pixels
		NDVI	Calculated from NIR and red bands, indicates live green vegetation density
NSF NEON	8	Enhanced vegetation index (EVI)	Similar to NDVI, estimates vegetation greenness and biomass
		Normalized difference nitrogen index (NDNI)	Relative nitrogen concentration in canopy
		Normalized difference lignin index (NDLI)	Uses shortwave IR to estimate lignin content in canopy
		Soil-adjusted vegetation index (SAVI)	Reduces soil brightness in areas where vegetation cover is low
		Atmospherically resistant vegetation index (ARVI)	Reduces atmospheric noise from dust, smoke, rain, etc.
		NDVI	Calculated from NIR and red bands, indicates live green vegetation density
		NDVI neighborhood	Avg. NDVI values of surrounding pixels
		Canopy height (LiDAR)	Height of canopy above bare earth
NEON + NAIP	17	All of the above	

Table 2.2: Summary of training points

Class	# ground-truthed polygons	# computer-drawn polygons	Total polygons	Total Pixels	Total area (m^2)	% of total training
Deciduous Trees	68	620	688	37,215	150,533.5	12.7%
Grass	246	160	406	179,799	719,013.9	60.3%
Easter Red Cedar	51	506	557	5,537	22,751.6	1.9%
Shrubs	341	1578	1919	71,101	285,690.1	24%
Other*	0	65	65	6,676	13,484.3	1.1%
Total	706	2929	3635	300,328	1,191,473	100%

*Other = water, roads, and buildings

Table 2.3: Accuracy results of all image and ML methods.

Source	ML Method	OA	Kappa	Accuracy	Deciduous Trees	Grassland	ERC	Shrubs	Other*
NAIP	SVM	0.912	0.842	PA	0.78	0.98	0.55	0.83	0.95
				UA	0.85	0.95	0.83	0.85	0.98
	RF	0.943	0.899	PA	0.82	0.98	0.61	0.93	0.98
				UA	0.91	0.97	0.83	0.89	0.98
NEON	SVM	0.973	0.953	PA	0.94	0.99	0.78	0.96	0.95
				UA	0.97	0.98	0.89	0.96	0.99
	RF	0.979	0.964	PA	0.95	0.99	0.79	0.98	0.98
				UA	0.97	0.99	0.89	0.96	0.99
NAIP + NEON	SVM	0.982	0.969	PA	0.96	0.99	0.84	0.97	0.98
				UA	0.97	0.99	0.94	0.97	0.99
	RF	0.984	0.973	PA	0.96	0.99	0.83	0.98	0.99
				UA	0.98	0.99	0.92	0.97	0.99

OA: overall accuracy; PA: producer accuracy; UA: user accuracy.

*Other = water, roads, and buildings

Table 2.4: Confusion matrix for SVM NAIP; columns represent class of training pixels, and rows represent class of model predicted pixels.

SVM NAIP	Deciduous Trees	Grassland	Eastern Red Cedar	Shrub	Other	User accuracy
Deciduous Trees	8671	15	500	971	19	0.852
Grassland	260	52908	126	2614	59	0.945
Eastern Red Cedar	122	13	926	38	19	0.828
Shrub	2042	991	134	17776	1	0.849
Other	8	23	2	0	1875	0.983
Producer accuracy	0.781	0.981	0.549	0.831	0.950	

Table 2.5: Confusion matrix for RF NAIP; columns represent class of training pixels, and rows represent class of model predicted pixels.

RF NAIP	Deciduous Trees	Grassland	Eastern Red Cedar	Shrub	Other	User accuracy
Deciduous Trees	9127	18	443	437	8	0.910
Grassland	226	53000	98	1065	20	0.974
Eastern Red Cedar	178	23	1027	11	4	0.826
Shrub	1566	883	117	19886	0	0.886
Other	6	26	3	0	1941	0.982
Producer accuracy	0.822	0.982	0.608	0.929	0.984	

Table 2.6: Confusion matrix for SVM NEON; columns represent class of training pixels, and rows represent class of model predicted pixels.

SVM NEON	Deciduous Trees	Grassland	Eastern Red Cedar	Shrub	Other	User accuracy
Deciduous Trees	10470	12	230	91	23	0.967
Grassland	53	53548	17	794	63	0.983
Eastern Red Cedar	129	2	1325	19	9	0.893
Shrub	447	372	116	20495	4	0.956
Other	4	16	0	0	1874	0.989
Producer accuracy	0.943	0.993	0.785	0.958	0.950	

Table 2.7: Confusion matrix for RF NEON; columns represent class of training pixels, and rows represent class of model predicted pixels.

RF NEON	Deciduous Trees	Grassland	Eastern Red Cedar	Shrub	Other	User accuracy
Deciduous Trees	10495	11	239	49	6	0.972
Grassland	49	53559	10	409	20	0.991
Eastern Red Cedar	139	7	1341	14	6	0.890
Shrub	413	358	98	20927	2	0.960
Other	7	15	0	0	1939	0.989
Producer accuracy	0.945	0.993	0.794	0.978	0.983	

Table 2.8: Confusion matrix for SVM NAIP+NEON; columns represent class of training pixels, and rows represent class of model predicted pixels.

SVM NAIP+NEON	Deciduous Trees	Grassland	Eastern Red Cedar	Shrub	Other	User accuracy
Deciduous Trees	10687	15	184	76	12	0.974
Grassland	31	53728	17	550	27	0.989
Eastern Red Cedar	72	2	1423	15	7	0.937
Shrub	308	194	63	20758	0	0.974
Other	5	11	1	0	1927	0.991
Producer accuracy	0.963	0.996	0.843	0.970	0.977	

Table 2.9: Confusion matrix for RF NAIP+NEON; columns represent class of training pixels, and rows represent class of model predicted pixels.

RF NAIP+NEON	Deciduous Trees	Grassland	Eastern Red Cedar	Shrub	Other	User accuracy
Deciduous Trees	10657	7	215	37	4	0.976
Grassland	25	53722	9	369	14	0.992
Eastern Red Cedar	107	7	1401	10	4	0.916
Shrub	310	202	62	20983	0	0.973
Other	4	12	1	0	1951	0.991
Producer accuracy	0.959	0.996	0.829	0.981	0.989	

Chapter 3 - Emergence of encroached wooded WUIs in grassland settlements: An analysis of changing fire risk

Please note: This chapter is formatted for the International Journal of Wildland Fire

Abstract

Woody encroachment of Eastern Red Cedar (*Juniperus virginiana*; henceforth, ERC), a native conifer, is quickly turning large swaths of tallgrass prairie into woodlands. This encroachment has many negative effects on grassland ecosystems, including a potential to shift the natural fire regime from ground fires to more dangerous forest crown fires. In drier and variable climates like the Central Great Plains, the question is not if, but when conditions will allow wildfires to spread. The goal of this project was to determine the spatial variability in forest fire risk in Manhattan, Kansas, an emerging semi-urban zone that exemplifies exurban expansion into the remaining grasslands of the central Great Plains. This study used fine-resolution aerial imagery to create a land use land cover map, which was then used to determine the extent and spatial patterns of ERC in Manhattan and thirteen neighborhoods, representing approximately 11,261 homes (structures from hereon). We found that ERC currently covers 9.1% (2,062 ha) of Manhattan, and ranges from 5-23% cover across neighborhoods. There is currently low connectivity between ERC patches but high cohesion, meaning that patches of ERC are growing close together but not touching yet. However, the gaps between ERC patches are small enough to disappear in coming years due to the speed of encroachment and considering the long typical flame length of ERC crown fires. We calculated number of houses within different distances to ERC patches based on three levels of danger: direct flame (within 4 m of houses), extreme radiant heat (within 20 m of houses), and embers (within 800 m of houses), and factorially crossed this analysis with calculations of distance to different patch sizes of ERC: patches $\geq 10\ m^2$, $\geq 1000\ m^2$, and $\geq 5000\ m^2$, as patch size could potentially influence fire behavior (e.g. large patches are more likely to produce long-distance embers). We found that a substantial number of houses were at risk for direct flame fire damage, but only from small

patches of ERC, indicating that fire must first spread from large patches to smaller patches for direct flame risk to occur. Furthermore, we found a small risk of radiant heat damage from all size ERC patches (11-20%). Lastly, an average of 90% of houses in neighborhoods are at risk for embers, with seven neighborhoods having 100% of houses within 800 m of ERC patches ≥ 5000 m^2. These large patches are more likely to create extreme fire behavior such as updrafts, signifying that almost all houses in most neighborhoods are within falling distance of embers and in danger of a spot fire. Therefore, if severe wildfire conditions and ignitions are combined in these neighborhoods, most structures could be in danger from either direct flames or wind-blown embers causing spot fires unless preventative measures are taken. This example is quite different than the typical image of wildfire prone wildland urban interface (WUI), where humans move into an area with high fuel loads typical of flammable forests. Instead, the Central Great Plains example here involves humans moving into a low fuel load system, sustained by frequent ground fires, which is switching to a high fuel load system after humans create settlements and suppress fire. The end result is still quite similar—human settlements are intertwined with vegetation that poses high wildfire risk, especially due to embers and spot fires.

Introduction

Wildfire is a national and international phenomenon which endangers humans, infrastructure, and in some cases, key ecosystems functions. Each decade, forest fires are breaking records for size, severity, and damage caused, a trend which is only expected to increase (UNEP 2022), especially in grassland biomes (Donovan et al. 2017). While wildfires are necessary for maintaining many ecosystems, they can cause damage when they encounter human systems. Wildfires cause damage to houses, businesses, powerlines, and water systems. In the U.S. alone, the amount of money spent on firefighting has increased more than 170% in the past decade, to $1.9 billion spent annually (UNEP 2022). In 1985, the average cost to fight a wildfire was $2,905, but has since grown 1,228% to $38,575 per wildfire in 2020 (Hurst 2022). Wildfires close to humans have many negative health effects, including damage to respiratory and cardiovascular systems from smoke inhalation (UNEP 2022). In many cases, humans are moving into woodlands already at risk for wildfire, and these risks have been well studied (Parisien et al. 2020, UNEP 2022). However, in the Great Plains and other grasslands, people appear to be moving into grasslands and excluding fire and/or large grazers, resulting in transitions to woodlands that could increasing the risk of wildfire (Log and Gjedrem 2022, Ratajczak et al. 2016, Mariani et al. 2022).

Woody encroachment—the expansion of woody plants in grasslands and savannas—is occurring in the Great Plains at rapid speeds (Briggs et al. 2002, Twidwell et al. 2013, Galgamuwa et al. 2020, Moser et al. 2013). Much of this invasion is Eastern Red Cedar (*Juniperus virginiana*; henceforth, ERC), the most widespread native conifer in the United States (Briggs et al. 2002). Just a century ago, the Central Great Plains was largely devoid of ERC, but as homesteaders settled the Great Plains, they planted ERC by the millions as windbreaks to help protect soils from desiccation and wind erosion, and to provide food, nesting sites, and thick cover for wildlife, as well as the wood itself being economically important to the Great Plains (Meneguzzo and Liknes 2015). Furthermore, even without active planting, an open prairie can turn into a closed canopy ERC forest in as little as 40 years without fire, due to seed sources from riparian forests (Briggs et al. 2002). One study found that the growing stock volume of ERC increased by 15,000% in Kansas between 1965-2010 (Moser et al. 2013), while another focusing only on riparian areas in Kansas found ERC increases of 139-539% from 1986-2017 (Galgamuwa et al. 2020). In areas that remain grasslands, approximately 45% are burned so infrequently that they are likely to transition to ERC woodlands in the next ten to thirty years (Ratajczak et al. 2016). What remains unknown is how close ERC expansion is to human settlements and their potential future effects.

Wildfires are becoming more common in the Great Plains, increasing in frequency more than any other ecosystem (Donovan et al. 2017), possible due to the increasing woody encroachment (Donovan et al. 2020) overlapping with extremely dry and windy weather (Donovan et al. 2017). Recent years with extreme drought and/or wind speeds were associated with especially large increases in wildfire incidence (Donovan et al. 2017), and these conditions are projected to become more common (Cook et al. 2015). Woody vegetation is more likely to burn in wildfires than other land types in the Great Plains (Donovan et al. 2020), and are more conducive to extreme wildfires due to their ability to create crown fires, which are fires that exist in the canopy of trees and spread faster, travel further, and are more difficult to control than ground fires (Scott and Reinhardt 2001). The danger for a crown fire increases with high tree density, connectivity, and fuel availability. When grassland burns, the flame lengths can vary between <0.1m to 3.4 m, while crown fires in ERC woodlands can produce flame lengths of >14 m (Twidwell et al 2013). This large flame length means that fires can jump over breaks in the trees (e.g., roads, grassland where woody encroachment is still semi-sparse, etc.), allowing the fire to continue in new areas. Furthermore, ERC has a much longer spotting distance than grassland fires—under wildfire conditions, ERC woodlands can send embers that could start new fires up to 5 km away, whereas grasslands in the same conditions can only send embers up to 1.3 km away (Donovan et al. 2023).

Most studies looking at wildfires and many studies tracking woody encroachment use coarse resolution aerial imagery, with at best 10 m^2 but often >250 m^2. However, small patches of ERC (<10 m^2) can be dangerous in the wrong places and newly encroached areas tend to have lots of small ERC trees spread out that are missed in coarse resolution aerial imagery. Coarse-grain products also fail to represent the connectivity of fire-prone vegetation, which is major oversight because small amounts of ERC could represent disproportionally large potential for wildfires if ERC trees are aggregated. Therefore, this study utilizes fine resolution aerial imagery to 1) track the current extent and spatial distribution of woody encroachment; and 2) determine if there an emerging fire-prone WUI forming around Manhattan KS and quantify how many houses are at risk for different types of wildfire danger.

Methods

Study Site

Manhattan is a city in eastern Kansas, United States, located on a floodplain at the junction of the Kansas and Big Blue rivers. Although historically a tallgrass prairie, Manhattan gets an average of 835 mm of rainfall each year and can support a woodland in the absence of fire. Prior to European colonization, the area was controlled by a series of Native American groups, such as the Kaw Nation and Osage, who purposely increased fire occurrence (Stambaugh et al. 2013). These fires probably occurred every 2 to 10 years, with an average fire return interval of around three years, which at the time, was sufficient to avoid widespread expansion of woody plants (Stambaugh et al. 2013). After European-Americans seized control of the area in the 1850s, the city grew quickly but the fire was almost entirely excluded from the system. Within the first 11 years of settlement, a public university and railroad were established, and the town became a significant mining supply town, allowing the population to grow substantially. Manhattan's population has continued to grow every decade since its founding, and has a current population of over 98,000 people in the Manhattan metropolitan area (U.S. Census Bureau 2015). An average of 17 new houses built each year within the city limits (Planning and Development Environmental Health Annual Report 2021) and even more being built on the edges of town. Fire has since been reintroduced into many public lands and private rangelands in the surrounding region, but is still absent from areas around dense settlements.

Creating Land Use Land Cover Map

We created a land use land cover (LULC) map using supervised random forest machine learning and two different sources of aerial imagery: USDA NAIP and Kansas NG911 (USDA 2019a&b, Kansas NG911 Coordinating Council 2023). The USDA NAIP imagery was captured on July 10, 2019 by low-flying aircrafts and has a resolution of 0.6 m^2. We used four inputs from NAIP, red, green, blue, and near infrared, and calculated five more bands, neighborhood values for each given band and normalized difference vegetation index (NDVI). Neighborhood values are calculated by taking the average of all adjacent pixels, which can help reduce errors caused by shadows. NDVI is a commonly used vegetation index which quantifies vegetation greenness.

The Kansas NG911 image was also captured by a series of low flying planes from February 26-April 11, 2021 and has a resolution of 1 ft^2. This resulted in a true color image with red, green, and blue bands and we calculated neighborhood values for each band, using a total of six bands. An added value of this winter image is that ERC is the only evergreen native to the state of Kansas and houses covered by tree canopies are visible in this image. We used bilinear interpolation to transform each image to 1.8 m^2 pixels that shared a common grid, which helped reduce model run times and excessive pixilation of small features like shadows.

We collected training polygons for ten categories of land use using a combination of in situ and ex situ methods: buildings, pavement, gravel, deciduous trees, ERC, shrubs, grasslands, water, agriculture, and other. Training data in some land use types were cropped to equalize across all types of land cover, except for two classes (gravel and other) which had the lowest cover, totaling 208,920 pixels, or about 677,000 m^2, in the final data set. 70% of pixels were used to train the model while the other 30% was used to evaluate model performance. Random forest models were run in program R (v4.0.5; R Core Team 2021) using the 'randomForest' package (v4.6.14; Liaw and Wiener 2002), with 400 trees (the 'ntree' parameter) and 4 input variables assessed at each node of each tree (the 'mtry' parameter). Figure 3.1 shows the final LULC map. This follows our approach in Chapter 2 (Noble and Ratajczak, *in review*), with the exception that in that effort we had LiDAR available, whereas in this effort the added product to NAIP imagery is a winter image. The LULC map has an overall accuracy of 93.3%, with the most important land types, ERC and buildings, performing especially well. ERC has a producer accuracy, which measures the proportion of evaluation pixels classified correctly in the LULC map, of 94.8%, and a user accuracy, which measures the proportion correctly classified pixels in the LULC map based on our evaluation pixels, of 96.7%. Buildings have a producer accuracy of

96.3% and a user accuracy of 94.9% (Table B.1). Therefore, our two most important land cover types had very high overall accuracy and low potential for both errors omission and commission.

Fire Risk

Neighborhood Delineation

We identified and analyzed thirteen neighborhoods in and around Manhattan (Fig. 3.2), delineated based on average age and house price. We intentionally tried to leave out non-residential buildings such as schools or businesses, but did not crop or segment neighborhoods to reach this goal. The thirteen neighborhoods do not encompass all houses in and around Manhattan, but were drawn to include sprawling suburbs and other neighborhood types. Four neighborhoods were drawn well within the confines of the city (neighborhoods 1, 3, 10, and 13), six on the edge of the city (neighborhoods 5, 6, 8, 9, 11, and 12), and three well outside the city (neighborhoods 2, 4, and 7) (Table 3.1). Average home value and age were calculated from a random sample of twenty houses per neighborhood except for neighborhood 1, which we could only find information on 18 houses) using information gathered from Zillow. In general, neighborhoods further from the urban core ten to be newer and have a higher proportion of single-family suburban homes. Each neighborhood was then buffered 800 m to include relevant surrounding vegetation that could generate wildfires that pose risks to each neighborhood; the LULC map was cropped to each buffered neighborhood. Landscape metrics were calculated for Manhattan and for each neighborhood individually.

Eastern Red Cedar Encroachment

Landscape metrics were run in program R (v4.0.5; R Core Team 2021) using the 'landscapemetrics' package (Hesselbarth et al. 2019) to analyze the following landscape metrics for ERC in Manhattan and each neighborhood: percent cover, average patch size, cohesion, contiguity, and average Euclidean distance to the nearest neighbor. Percent cover measures the proportion of each class relative to the landscape. Average patch size is the average size of all patches for each class, measured in square meters. Patches are defined as sets of contiguous vegetation of the same type—in our case 1.8 x 1.8 m ERC pixels. Cohesion is an aggregation metric that explains the spatial distribution of patches; the unit is a percentage and values ranging from 0 to 100, with values close to 0 being that patches are isolated and values close to 100 being that patches are entirely aggregated together. Contiguity measures the spatial connectedness of patches; it is a unitless metric and values range from 0 to 1, with 0 being that each patch is a single pixel and 1 being that the entire class is one single patch. Average Euclidean nearest neighbor measures the average distance between patches, measured in meters. All landscape metrics were run in the queen's rule, with 8-connectedness between pixels.

Quantifying Homes in Danger

All ERC pixels from the LULC map were transformed into polygons, with all touching ERC pixels representing a single patch of ERC and transformed into a single polygon. Structure polygons were soured from FEMA (FEMA 2022) and structures were cropped in each unbuffered neighborhood. Note that these polygons were separate from those used to train the random forest model. Structures include houses and other large buildings, but do not include smaller structures such as sheds or detached garages.

There are three different ways that houses are at risk from a wildfire: direct flame, extreme radiant heat, and embers. These different risks are found at different distances to homes. Some studies suggest that danger from direct flame can be within 10 m of a home, danger from radiant heat can be up to 30 m, and danger from falling embers (i.e., spotting distance) can be up to 5000 m in extreme conditions (FireSmart Canada 2023, Donovan et al. 2023), but values have not yet been calculated for ERC so we used more conservative values as not to overestimate wildfire risk. Therefore, we analyzed the number of houses within three distances from ERC patches: 4 meters to estimate potential exposure to direct flames, 20 meters for radiant heat, and 800 meters for embers and other forms of spot fires.

The exposure to these three types of fire risk likely varies with ERC patch size. All size patches of ERC can create flames and put houses in danger of direct flame risk. However, patch sizes of < 10 m^2 were likely to be map errors or single trees which are difficult to catch on fire, therefore we disregarded any ERC patches less than ten square meters in size. Medium patch sizes (≥1000 m^2) can pose risks from direct flames and are likely the minimum patch size required to burn hot enough to create danger from radiant heat ≤20 m away from structures. Finally, large patches (≥5000 m^2) are the most likely to burn due to their high connectivity, and the resulting wildfires are more likely to burn hot enough to spread to smaller patches due to their higher fuel availability. Large crown fire can also produce updrafts and that spread embers further than embers produced by smaller burning patches. We analyzed the six factorial combinations of patch size and distance from structures: ERC patches ≥10 m^2, ≥1000 m^2, and ≥5000 m^2 within 4 m of structures, patches ≥1000 m^2 and ≥5000 m^2 within 20 m of structures, and patches ≥5000 m^2 within 800 m of structures.

Results

Eastern Red Cedar Encroachment

ERC covers 9.1% of Manhattan, and ranges from 5-23% across neighborhoods (Fig. 3.3, Table B.2). Patches of ERC average 95.4 m^2 in Manhattan, with a range of average cover of 39.5 to 176.9 m^2 among neighborhoods. Cohesion values are high and had little variation across all neighborhoods, ranging from 89-98% with an average of 94.7%, indicating that ERC tends to aggregate on the landscape. Contiguity, however, is quite low across neighborhoods with low variability, with values ranging from 0.22-0.27, meaning that patches of ERC tend to be disconnected (Table B.2). However, the average Euclidean nearest neighbor for ERC patches ranges from 6.6-8.2 m, indicating that, on average, gaps between patches are small (Table B.2). Therefore, while ERC connectivity is low, trees tend to grow in the same areas with small gaps between patches. Furthermore, flame lengths of ERC can be longer than these gap distances (Twidwell et al. 2013), suggesting that patches of ERC still tend to be functionally connected throughout much of Manhattan and its outlying neighborhoods.

The most encroached neighborhoods were on the edge of town. ERC covered an average of 13.5% of edge neighborhoods, versus an average of only 10% of interior neighborhoods and 6.6% in exterior neighborhoods (Fig. 3.3). The neighborhood with the worst encroachment, Neighborhood 9 located on the edge of Manhattan, has ERC coving 23% of the landscape, with an average patch size of 176.9 m^2, and a cohesion value of 98.8 (Table B.2).

Surprisingly however, the neighborhoods on the outside of town have the least encroachment. Two of the three have ERC only covering around 5% of the landscape (Fig. 3.3, Table B.2). However, average patch size in outside neighborhoods varied greatly; neighborhood

7 has the smallest average patch size across neighborhoods, while neighborhood 4 has the second largest average patch size.

Quantifying Homes in Danger

On average, there is a high risk for direct flame and embers, but low risk for radiant heat. Currently, there are 82,363 patches of ERC 10 m^2 or larger within our drawn boundary of Manhattan. Across all thirteen neighborhoods, 84.3% of houses are within 4m of ERC patches ≥ 10m^2 (Fig. 3.4, Table B.3). All thirteen neighborhoods have over 50% of houses within 4m of ERC patches ≥ 10m^2, and ten neighborhoods have 75% of houses within 4m of ERC patches ≥ 10m^2. This indicates that a substantial number of homes are in danger of damage from direct flames of wildfire in neighborhoods.

There are 1,419 patches of ERC 1000 m^2 or larger in Manhattan. An average of 10.9% of houses are within 4m of ERC patches ≥ 1000m^2 for all thirteen neighborhoods, with values ranging from 0-31% of houses. An average of 20.3% of houses are within 20m of ERC patches ≥ 1000m^2 across neighborhoods, with only one neighborhood with over 50% of houses within 20 m of ERC patches ≥ 1000 m^2, and eight neighborhoods with less than 25% of houses within 20 m of ERC patches ≥ 1000 m^2 (Fig. 3.4).

There are 345 patches of ERC 5000 m^2 or larger in Manhattan. Only 7% of houses in neighborhoods are within 4 m of large patches, but values range from 0-23% between neighborhoods (Table B.3). 11% of houses are within 20 m of ERC patches ≥ 5000 m^2, and neighborhood three has 50% of houses in this distance. There is an average of 89.7% of houses are within 800m of ERC patches ≥ 5000 m^2 across all neighborhoods, with seven neighborhoods having 100% of houses within 800m of ERC patches ≥ 5000 m^2, and four more have over 75%

of houses within 800m of ERC patches ≥ 5000 m^2, signifying that almost all houses in most neighborhoods are within falling distance of embers and in danger of a spot fire (Fig. 3.4). However, large patches of ERC tend to currently exist outside of neighborhoods, so radiant heat and direct flame risk is low when only considering large ERC patches.

Direct flame risk was high when looking at all size patches of ERC (10 m^2 or larger), but low when only considering medium (1000 m^2 or larger) and large (5000 m^2) size patches. An average of 84% of houses in all neighborhoods are within 4 m of an ERC patch 10 m^2 or larger, whereas an average of only 11% of houses are in danger of direct flame from ERC patches of 1000 m^2 or larger, and only 7% of houses are within 4 m of ERC patches 5000 m^2 or larger. This means that a crown fire must spread to smaller patches first in order to pose a danger. However, given the high cohesion and small distances between ERC patches, fire spread between patches is possible.

Fire risk to homes varied between location. Edge neighborhoods had the highest average percentage of houses within direct flame risk for ERC patches 10 m^2 or larger (90%), exterior neighborhoods had the highest average percentage of houses within radiant heat and ember risk (25.3% and 91.9%, respectively, Fig. 3.4). Interior neighborhoods had the lowest average percentage of houses within both direct flame and ember risk (77.3% and 87.1%, respectively, Fig. 3.4).

The average number of houses at risk for all combinations of patch size and distance to houses across neighborhoods was higher than for the average of all Manhattan (Fig. 3.4, Table B.3). For radiant heat risk, neighborhoods have an average of 20.3% of houses within 20 m of ERC patches ≥ 1000 m^2, but only 10.5% of all buildings across Manhattan meet that criteria (Fig. 3.4). Furthermore, direct flame risk is 3.5% higher and ember risk is 8.6% higher in neighborhoods compared to Manhattan (Fig. 3.4). Therefore, it appears that dense residential areas have a disproportionally higher fire risk.

Discussion

Aerial photographs and on the ground accounts from the early 1900s suggest that Manhattan and its metropolitan area were largely devoid of woody vegetation, except along rivers and deep values (Abrams 1986, Bragg and Hulbert 1976, Briggs et al. 2002). Over 100 years later, shrubs and ERC trees have encroached much of the area, leading to higher fire danger. The most encroached neighborhood has high fire risk for all types of fire danger (neighborhood 9), but this was variable between neighborhood location and across type of fire danger (Tables B.2&B.3). Neighborhoods outside of Manhattan had relatively low encroachment, but had the highest percentage of houses at risk of radiant heat and ember damage. Interior neighborhoods had the lowest average patch size of ERC, but highest proportion of houses within 4 m of ERC patches ≥ 5000 m^2 (10%), compared to edge neighborhoods (5%) and exterior neighborhoods (4%). Our results reiterate the importance of landcover configuration, which is only measurably with high-resolution high accuracy remote sensing.

Much of the Great Plains is not considered part of the Wildland-Urban Interface (WUI), which refers to areas where human settlements are intermixed with flammable wildland vegetation (Radeloff et al. 2018, Stein et al. 2013). A recent study found an increase in WUI area in Texas and Oklahoma (Radeloff et al. 2018), where woody encroachment of Juniper species is further along (Engle et al. 2008). However, the increase in WUI area was due to increases in housing density and population, rather than changing vegetation (Radeloff et al. 2018).

However, our results indicate that much of Manhattan and its neighborhoods are fairly well forested by a fire-prone ERC with high functional connectivity and a substantial number of houses are at risk for multiple types of forest fire danger, indicating that a WUI now exists in Manhattan. Most WUIs in the United States are created by people moving into forests, but in this case, humans are settling in grasslands and creating a WUI through fire suppression, allowing woody fuels to increase, a relatively new and understudied phenomenon.

Our results indicate that a significant percentage of structures are at risk of damage from wildfires, but there are easy and fairly low-cost ways to reduce and mitigate risk. Common guides to reduce risk of damage to homes suggest three zones of defensible space (FEMA 2008, FireSmart Canada 2023), with different actions to reduce risk in each zone. Reducing radiant heat risk requires management of land up to 20 m from houses, including tree removal and trimming low branches, which is expensive (FireSmart Canada 2023). Reducing the risks of direct flame and embers, however, are much easier and less costly because both focus on the zone immediately surrounding homes (within 4 m). While embers can come in from up to 800 m away, making changes on or near the home reduces the risk of home damage, such as adding mesh covers on chimneys and vents, keeping gutters clear of dry leaves, and moving flammable materials (wood piles, etc.) away from structures. Similarly, reducing risk of direct flame damage requires reducing flammable surfaces within a small barrier of houses. Therefore, while risk of damage from wildfires is high across Manhattan, there are easy methods to reduce this risk.

Values for critical patch size and connectivity required for an ERC crown fire to occur and spread are unknown. However, wildfires are more likely to occur and spread in extremely dry and windy conditions (Donovan et al. 2023, Reid et al. 2010) and our estimates are assuming those extreme conditions. Therefore, high fire danger for houses is only relevant in extreme wildfire conditions, but extreme wildfire conditions (e.g., very dry and windy) are expected to occur more often in the Great Plains with a changing climate (Cook et al. 2015). It's also important to note that this study is probably undercounting homes at risk by using conservative values for distances to homes at which fire is dangerous. For example, some studies suggest that danger from direct flame can be within 10 m of a home, danger from radiant heat can be up to 30 m, and danger from falling embers (i.e., spotting distance) can be up to 5000 m in extreme conditions (FireSmart Canada 2023, Donovan et al. 2023), but values have not yet been calculated for ERC. However, even using conservative values and possibly undercounting risk, we still found a substantial proportion of Manhattan homes in danger.

The trend of woody encroachment increasing fire prevalence is occurring across the Great Plains (Stein et al. 2013, Donovan et al. 2017). In the southern Great Plains, where ERC encroachment is more established (Engle et al. 2008), the number of large wildfires is increasing (Donovan et al. 2017), disproportionately burning woody vegetation (Donovan et al. 2020). Riley County, which encompasses our study site, has seen an increase in wildfires with several just in April 2022, but no structure damage thus far (Riley County Fire District No. 1 2023*a*, 2023*b*, 2023*c*, 2023*d*). However, as wildfire conditions continue to increase in frequency and intensity, wildfires will increase and structure damage becomes more and more likely.

Many areas outside of the Great Plains are also seeing woody encroachment into WUIs with increased fire frequencies and structure damage (Filkov et al. 2020, Log and Gjedrem 2022, Mariani et al. 2022). Wildfires in Australia are common, but becoming much more frequent and intense due to encroachment of shrubs (Mariani et al. 2022), leading to higher rates of destruction of lives and property (Filkov et al. 2020). Furthermore, in Portugal, areas with shrub encroachment burn disproportionately higher than any other land type, including native conifer forests (Moreira et al. 2009). Even in places where wildfires are rare, such as an island in Norway, similar woody encroachment of a native juniper due to fire suppression has resulted in a wildfire which damaged several structures (Log and Gjedrem 2022).

Conclusion

Woody encroachment is quickly spreading through Manhattan KS, particularly in edge neighborhoods. Patches of ERC are growing close together on the landscape, with gaps between patches smaller than possible flame lengths of crown fires, and likely to disappear in coming years with the current speed of encroachment, indicating high functional connectivity. Most houses in Manhattan neighborhoods are close to ERC patches and within distances to direct flames and embers if a wildfire breaks out. However, there are low-cost ways to reduce this risk by reducing flammability of surfaces immediately surrounding homes and covering any entry points to protect from embers. While we only focused on metropolitan area, given the universality of woody encroachment and settlement expansion across the Central Great Plains, we expect our results apply to other towns and cities throughout the region or in similar grassland landscapes.

Acknowledgments

We would like to thank my funding sources, the Kansas State University Division of Biology, the Kansas Agricultural Station, and the NSF-funded Konza Long-Term Ecological Research (LTER) grant. We would also like to thank all the amazing people who run Konza, including the prescribed burning volunteers, for helping to maintain the long-term experiments.

Finally, a huge thank you to Dylan Darter and Jaclyn Perry, our field technicians who helped collect thousands of training points.

Figures

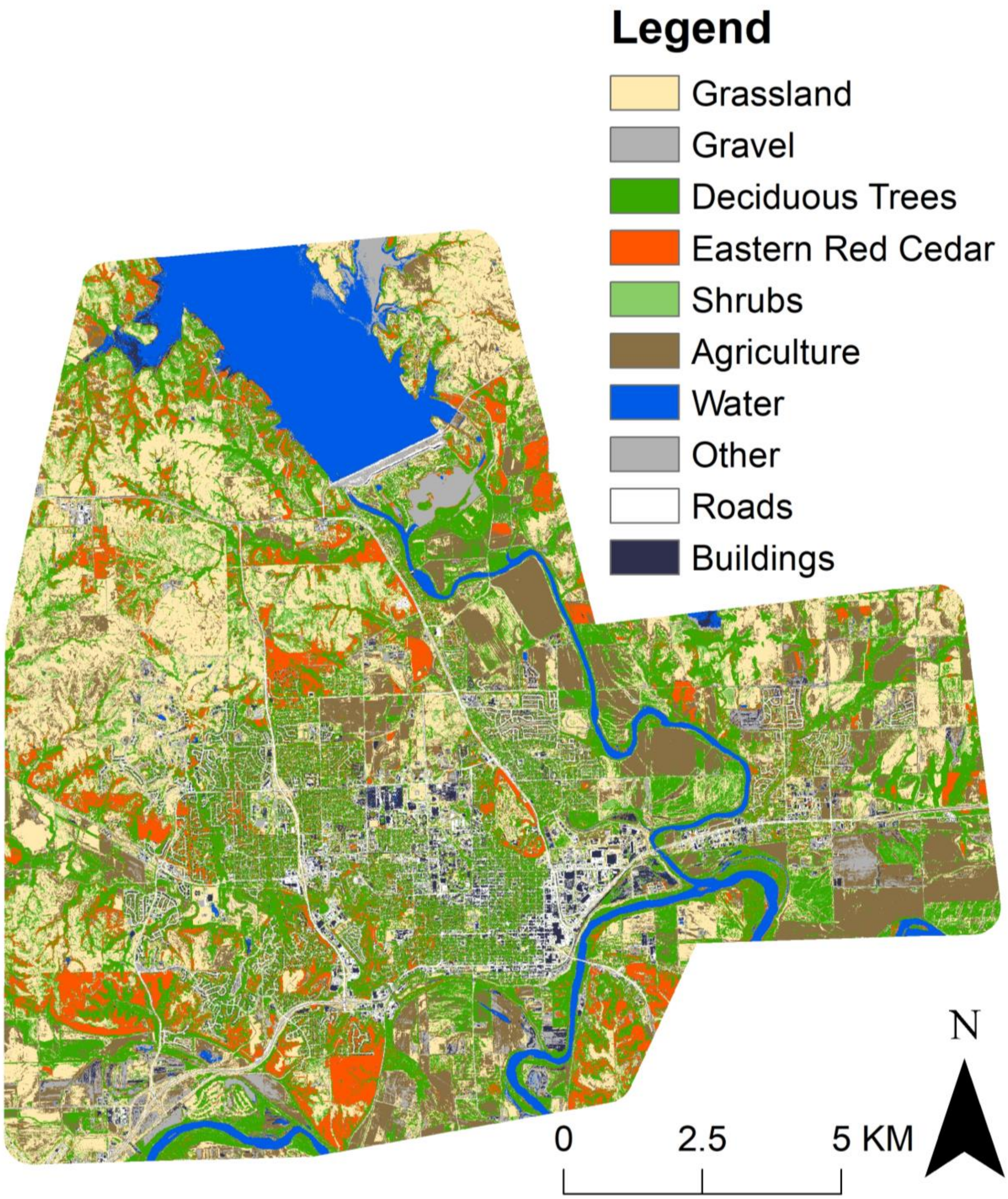

Figure 3.1: Land Use Land Cover Map of Manhattan, Kansas

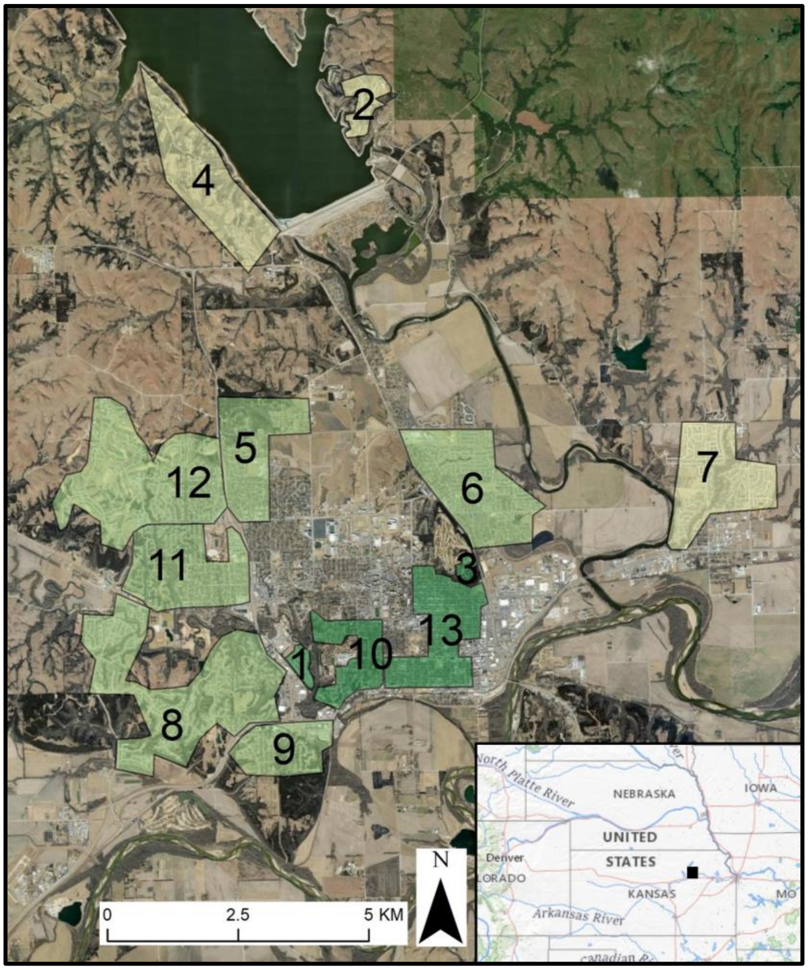

Figure 3.2: Map of neighborhoods used in study. Yellow polygons indicate exterior neighborhoods, light green polygons indicate edge neighborhoods, and dark green polygons indicate interior neighborhoods.

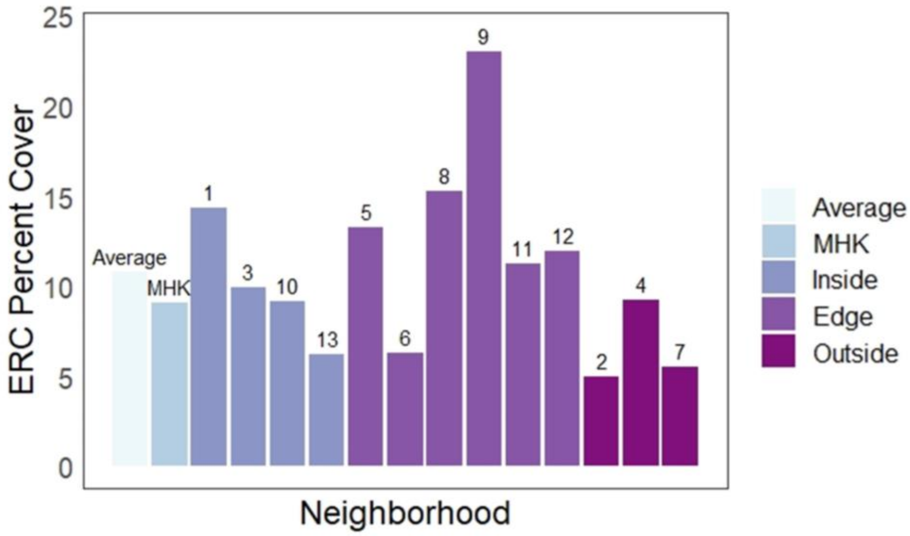

Figure 3.3: Percent cover of Eastern Red Cedar. "Average" is across all thirteen neighborhoods, "MHK" is across all Manhattan. Numbers represent neighborhoods.

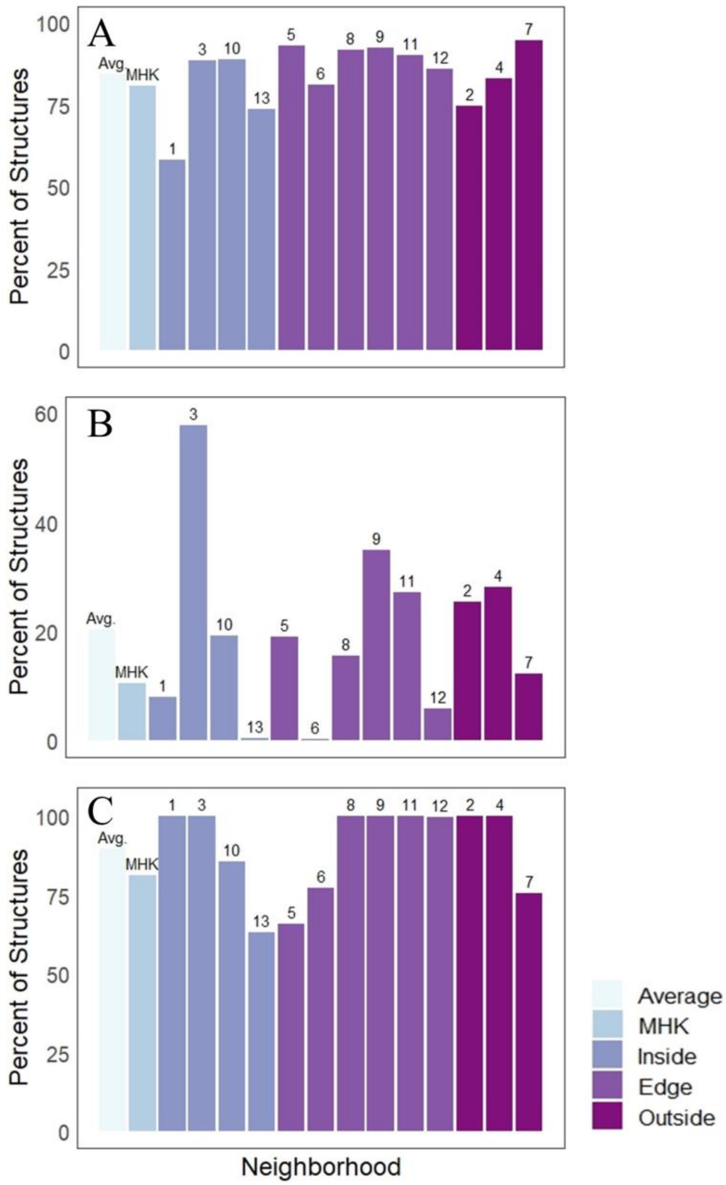

Figure 3.4: Percentage of structures in each neighborhood A) within 4 m of ERC patches ≥ 10 m^2, B) within 20 m of ERC patches ≥ 1000 m^2, and C) within 800 m of ERC patches ≥ 5000 m^2. "Average" is across all thirteen neighborhoods, "MHK" is across all structures in Manhattan. Numbers represent neighborhoods.

Tables

Table 3.1: Summary of Manhattan neighborhoods used in this study

Neighborhood	Area (ha)	# Houses	House Density	Location	Avg Age	Avg. price
1	24.6	289	11.7	Interior	21.3	180 272
2	67.5	63	0.9	Outside	24.5	485 615
3	17.7	26	1.5	Interior	65.2	294 800
4	347.7	286	0.8	Outside	47.5	268 828
5	266.5	728	2.7	Edge	42.4	337 995
6	361.4	2214	6.1	Edge	34	184 170
7	304.7	1080	3.5	Outside	10.7	389 725
8	608.5	1442	2.4	Edge	15.3	398 943
9	169.2	547	3.2	Edge	34.2	282 272
10	143.7	736	5.1	Inside	76.4	322 200
11	315.9	1285	4.1	Edge	36.1	421 200
12	548.4	1107	2.1	Edge	15.1	437 078
13	264.2	1758	6.7	Inside	80.7	259 430
Manhattan	17552.8	17634	1.004	--	--	--

Chapter 4 - Conclusion

Woody encroachment of Eastern Red Cedar (*Juniperus virginiana*; henceforth ERC) is quickly spreading from the southern Great Plains northward into Kansas in a “green glacier” of trees, and is expected to continue infilling throughout Kansas and spread up through Nebraska (Engle et al. 2008). The intensity of woody plant encroachment in Oklahoma is several decades ahead of most of Kansas, and in Oklahoma the number of wildfires in Oklahoma is increasing each year (Donovan et al. 2017). This increase in large wildfires in Oklahoma tracks ERC encroachment because woody vegetation is more conducive to extreme wildfires (Donovan et al. 2020). In contrast, grasslands tend to be burned under prescribed conditions, as part of a reciprocal relationship between local economies and maintaining tree-free ecosystems. Consequently, the number of structures damaged in wildfires is much greater in the southern Great Plains than in areas where woody vegetation is less established (Stein et al. 2013). As ERC continues to encroach into the Flint Hills, including Manhattan neighborhoods, we could see a similar increase in large wildfires and structures lost. This issue would be especially acute during years that combine dry conditions and high wind speeds (Donovan et al. 2017).

Most studies on wildfires or woody encroachment use coarse resolution aerial imagery (>10 m^2). However, the U.S. government now provides access to free high-resolution remote-sensed aerial imagery which can be used to paint a more accurate picture of the landscape. We used USDA NAIP and NSF NEON to create a land use land cover (LULC) map with 2 m^2 resolution. We found that NEON inputs alone were more accurate than NAIP-only inputs (97.9% vs. 94.3%, respectively), but most of that accuracy relied on LiDAR (Figure 2.4). Accuracy was highest when inputs were used in conjunction with each other (98.4%), and again relied heavily on NEON’s LiDAR, but the next five inputs used for determining land cover were NAIP inputs.

However, NAIP alone is not enough to accurately classify land use. We also found no difference in accuracy between using random forest and support vector machines for classification, but random forest models took several minutes to run while support vector machines took several hours to run. Therefore, for large-scale map projections support vector machines will only be viable with largescale cloud or cluster computing, and even then, could prove time consuming.

Using insights from chapter 2, I created a much larger land cover map of Riley County and surrounding areas using random forests and two large-scale remote sensing products: NAIP and Kansas NG911, a fine-scale winter aerial image. Using this map, my third chapter aimed to analyze the extent and spatial patterns of ERC in Manhattan and thirteen neighborhoods, representing approximately 11,261 homes. I found that ERC encroachment is well underway in Manhattan; ERC currently covers 9.1% (2,062 ha) of Manhattan, and ranges from 5-23% cover across neighborhoods. Connectivity between ERC patches is currently low, but cohesion of ERC is high, meaning that these stands of ERC tend to grow together but are not touching yet. However, the average gap between ERC patches is low (7 m), which is small enough for the ERC flame lengths to jump (especially in extreme conditions), and could disappear in coming years due to the speed of encroachment.

We also quantified how many houses were in three different ranges of fire risk. We found a substantial number of houses at risk for direct flame, but only for small patches of ERC, which are unlikely to be part of a crown fire complex. An average of 84% of houses in all neighborhoods are within 4 m of an ERC patch 10 m^2 or larger, whereas an average of only 11% of houses are in danger of direct flame from ERC patches of 1000 m^2 or larger, and only 7% of houses are within 4 m of ERC patches 5000 m^2 or larger. This indicates that danger of damage from direct flames of wildfires is high but only if a wildfire first spreads to smaller patches.

Furthermore, there is a very high danger for spot fires, which are fires created by falling embers. Seven of thirteen neighborhoods have 100% of houses within 800m of large ERC patches 5000 m^2 or larger, and an additional four neighborhoods have over 75% of houses within ember falling distance of large ERC patches.

Much of the Great Plains is not considered part of the Wildland-Urban Interface (WUI), which refers to areas where human settlements are intermixed with flammable wildland vegetation (Stein et al. 2013). However, our results indicate that much of Manhattan and its neighborhoods are fairly well forested by a fire-prone ERC with high functional connectivity and a substantial number of houses are at risk for multiple types of forest fire danger, indicating that a WUI now exists in Manhattan. Most WUIs in the United States are created by people moving into forests, but in this case, humans are settling in grasslands and creating a WUI through fire suppression, allowing woody fuels to increase, a relatively new and understudied phenomenon.

A commonly recommended method to reduce wildfire risk to homes in the WUI is defensible space, which is the space around homes which has been improved to stop or slow the spread of wildfire. Defensible space generally refers to three different zones of space at different distances from the house, each with specific instructions to reduce risk of fire damage, such as xeriscaping or thinning trees (FEMA 2008, FireSmart Canada 2023). The most expensive actions, thinning trees and trimming branches further away from homes, are to reduce risk of radiant heat. However, only a low percentage of houses in Manhattan neighborhoods are at risk for this damage (20%). Actions to reduce direct flame or ember damage, on the other hand, are relatively low-cost and only require action to the house and the zone immediately surrounding the house, such as adding mesh covers to chimneys or vents, moving firewood piles away more than 10 m away, and removing any trees or shrubs. Despite risk currently being high, there are relatively easy methods to reduce this risk. Wildfires are already beginning to increase around neighborhoods in Manhattan, with several burning just in April, 2023 (Riley County Fire District No. 1 2023a, 2023b, 2023c, 2023d). Luckily, no homes or other structures have been damaged in wildfires thus far, but it is better to be proactive and take measures to reduce risk so that we can limit the damage to structures in the future.

While our study focused on a single city in Kansas, U.S., this issue of increasing forest fire risk due to woody encroachment is not limited to the southern Great Plains; grasslands across the globe are being encroached by potentially flammable woody species, creating new WUIs (Archer et al. 2017). Wildfires are already becoming more frequent and severe around the globe (UNEP 2022), and could become even more so as woody encroachment shifts grasslands into fire-prone woodlands. As wildfire activity increases, the risk and danger to humans and human structures continues to grow as well, even in places thought to have low fire danger.

References

Abrams, Marc D. 1986. Historical development of gallery forests in northeast Kansas. Vegetatio 65(1):29-37.

Albrecht, M. A., R. E. Becknell, & Q. Long. 2016. Habitat change in insular grasslands: woody encroachment alters the population dynamics of a rare ecotonal plant. Biological Conservation, 196:93-102.

Allred, B. W., B.T. Bestelmeyer, C. S. Boyd, C. Brown, K. W. Davies, M.C. Duniway, L. M. Ellsworth, T. A. Erickson, S. D. Fuhlendorf, T. V. Griffiths, and V. Jansen. 2021. Improving Landsat predictions of rangeland fractional cover with multitask learning and uncertainty. Methods in Ecology and Evolution 12(5):841-9.

Archer, S.R., Andersen, E.M., Predick, K.I., Schwinning, S., Steidl, R.J., Woods, S.R. 2017. Woody Plant Encroachment: Causes and Consequences. In: Briske, D. (eds) Rangeland Systems. Springer Series on Environmental Management. Springer, Cham.

Anadón, J. D., O. E. Sala, B. L. Turner, and E. M. Bennett. 2014. Effect of woody-plant encroachment on livestock production in North and South America. Proceedings of the National Academy of Sciences 111(35):12948-53.

Bork, E. W., and J. G. Su. 2007. Integrating LIDAR data and multispectral imagery for enhanced classification of rangeland vegetation: A meta analysis. Remote Sensing of the Environment 111:11-24.

Bidwell, T. G., J. R. Weir, and D. M. Engle. 2013. Eastern redcedar control and management-best management practices to restore Oklahoma's ecosystems. Oklahoma State University. Division of Agricultural Sciences and Natural Resources. Cooperative Extension Service. https://hdl.handle.net/11244/50056.2

Bragg, T. B., and L.C. Hulbert. 1976. Woody plant invasion of unburned Kansas bluestem prairie. Journal of Range Management, 29(1):19-24.

Brandt, M., C. J. Tucker, A. Kariryaa, K. Rasmussen, C. Abel, J. Small, J. Chave, L. V. Rasmussen, P.Hiernaux, A. A. Diouf, L. Kergoat, O. Mertz, C. Igel, F. Gieseke, J. Schöning, S. Li, K. Melocik, J. Meyer, S. Sinno, E. Romero, E. Glennie, A. Montagul, M. Dendoncker, and R. Fensholt. 2020. An unexpectedly large count of trees in the West African Sahara and Sahel. Nature 587:78-82.

Briggs, J., G. Hoch, and L. Johnson. 2002. Assessing the Rate, Mechanisms, and Consequences of the Conversion of Tallgrass Prairie to Juniperus virginiana Forest. Ecosystems 5:578–586.

Briggs, J. M., A. K. Knapp, J. M. Blair, J. L. Heisler, G. A. Hoch, M. S. Lett, and J. K. McCarron. 2005. An ecosystem in transition: causes and consequences of the conversion of mesic grassland to shrubland. BioScience 55(3):243-25

Burges, C. J. 1998. A tutorial on support vector machines for pattern recognition. Data mining and knowledge discovery, 2(2):121-167.

Cook, B.I., Ault, T.R. and Smerdon, J.E., 2015. Unprecedented 21st century drought risk in the American Southwest and Central Plains. Science Advances, 1(1):e1400082

Donovan, Victoria. M., Carissa L. Wonkka, and Dirac Twidwell. 2017. Surging wildfire activity in a grassland biome, Geophysical Research Letters 44:5986–5993.

Donovan, Victoria M., Carissa L. Wonkka, David A. Wedin, and Dirac Twidwell. 2020. Land-Use Type as a Driver of Large Wildfire Occurrence in the U.S. Great Plains. Remote Sensing 12:1869.

Donovan, Victoria M., Dillon T. Fogarty, and Dirac Twidwell. 2023. Spot-fire distance increases disproportionately for wildfires compared to prescribed fires as grasslands transition to *Juniperus* woodlands. PLOS One 18(4): e0283816.

Engle, D.M., B. R. Coppedge, and S. D. Fuhlendorf. 2008. From the dust bowl to the green glacier: human activity and environmental change in Great Plains grasslands. In Western North American Juniperus Communities (pp. 253-271). Springer, New York, NY.

Evans, J. S., M. A. Murphy, Z. A. Holden, and S. A. Cushman. 2011. Modeling species distribution and change using random forest. In Predictive species and habitat modeling in landscape ecology (pp. 139-159). Springer, New York, NY.

FEMA. Defensible Space: Home Builder's Guide to Construction in Wildfire Zones. Technical Fact Sheet No. 4. September 2008.

FEMA. 2022. USA Structures [in en]. https://gis-fema.hub.arcgis.com/pages/usa-structures. Accessed: 3-28-2023

Filkov, Alexander I., Tuan Ngo, Stuart Matthews, Simeon Telfer, and Trent D. Penman. 2020. Impact of Australia's catastrophic 2019/20 bushfire season on communities and environment. Retrospective analysis and current trends. Journal of Safety Science and Resilience 1:44-56.

FireSmart Canada. 2023. FireSmart Begins at Home Guide. https://firesmartcanada.ca/wp-content/uploads/2023/04/Begins-at-Home-Guide-WEB.pdf

Galgamuwa, G. P., J. Wang, and C. J. Barden. 2020. Expansion of Eastern Redcedar (*Juniperus virginiana L.*) into the Deciduous Woodlands within the Forest–Prairie Ecotone of Kansas. Forests, 11(2):154.

Hesselbarth, M.H.K., Sciaini, M., With, K.A., Wiegand, K., Nowosad, J. 2019. landscapemetrics: an open-source R tool to calculate landscape metrics. - Ecography 42:1648-1657(ver. 4.0.5).

Hurst, Andrew. "Wildfire Statistics: Wildfire Damage & Home Insurance Claims by State." Value Penguin. 2022. https://www.valuepenguin.com/homeowners-insurance/wildfire-statistics#cost

Jin, H. and G. Mountrakis. 2022. Fusion of optical, radar and waveform LiDAR observations for land cover classification. ISPRS Journal of Photogrammetry and Remote Sensing 187:171-190.

Kansas NG911 Coordinating Council. 2023. 2021 NG911 Imagery. https://hub.kansasgis.org/datasets/KSDOT::2021-ng911-imagery

Kaskie, K. D., M. C. Wimberly, and P.r J. Bauma. 2019. Rapid assessment of juniper distribution in prairie landscapes of the northern Great Plains. International Journal of Applied Earth Observation and Geoinformation 83:101946.

Kaskie, K. D., M. C. Wimberly, and P. J. Bauman. 2022. Predictive Mapping of Low-Density Juniper Stands in Prairie Landscapes of the Northern Great Plains. Rangeland Ecology & Management 83:81-90.

Keen, R.M., J. B. Nippert, P. L. Sullivan, Z. Ratajczak, B. Ritchey, K. O'Keefe, and W. K. Dodds. 2022. Impacts of Riparian and Non-riparian Woody Encroachment on Tallgrass Prairie Ecohydrology. Ecosystems: 1-12. DOI: https://doi.org/10.1007/s10021-022-00756-7.

Kranjčić, N., D. Medak, R. Župan, and M. Rezo. 2019. Support vector machine accuracy assessment for extracting green urban areas in towns. Remote Sensing, 11(6):655.

Lautenbach, J. M., R. T. Plumb, S. G. R.n, C. A. Hagen, D. A. Haukos, and J. C. Pitman. 2017. Lesser prairie-chicken avoidance of trees in a grassland landscape. Rangeland Ecology & Management 70:78-86.

Lettow, M. C., L. A. Brudvig, C. A. Bahlai, J. Gibbs, R. P. Jean, and D. A. Landis. 2018. Bee community responses to a gradient of oak savanna restoration practices. Restoration Ecology, 26(5):882-890.

Liaw, A., and M. Wiener. 2002. Classification and Regression by randomForest. R News 2(3):18-22. R package version 4.6.14. https://CRAN.R-project.org/doc/Rnews/

Limb, R. F., D. M. Engle, A. L. Alford, and E. C. Hellgren. 2010. Tallgrass prairie plant community dynamics along a canopy cover gradient of eastern redcedar (*Juniperus virginiana L.*). Rangeland Ecology & Management, 63(6):638-644.

Log, T.; Gjedrem, A.M. 2022. A Fire Revealing Coastal Norway's Wildland–Urban Interface Challenges and Possible Low-Cost Sustainable Solutions. International Journal of Environmental Research and Public Health 19:3038.

Mantero, P., G. Moser, and S. B. Serpico. 2005. Partially supervised classification of remote sensing images through SVM-based probability density estimation. IEEE Transactions on Geoscience and Remote Sensing, 43(3):559-570.

Mariani, Michela, Simon E Connor, Martin Theuerkauf, Annika Herbert, Petr Kuneš, David Bowman, Michael-Shawn Fletcher, Lesley Head, A Peter Kershaw, Simon G Haberle, Janelle Stevenson, Matthew Adeleye, Haidee Cadd, Feli Hopf, and Christy Briles. 2022. Disruption of cultural burning promotes shrub encroachment and unprecedented wildfires. Frontiers in Ecology and the Environment 20(5):292-300.

Maxwell, A. E., T. A. Warner, B. C. Vanderbilt, and C. A. Ramezan. 2017. Land Cover Classification and Feature Extraction from National Agriculture Imagery Program (NAIP) Orthoimagery: A Review. Photogrammetric Engineering & Remote Sensing 83(11):737–747.

McCaffrey, Sarah M., Melanie Stidham, Eric Toman, and Bruce Shindler. 2011. Outreach Programs, Peer Pressure, and Common Sense: What Motivates Homeowners to Mitigate Wildfire Risk? Environmental Management 48:475-488.

Meneguzzo D. M. and G. C. Liknes. 2015. Status and Trends of Eastern Redcedar (*Juniperus virginiana*) in the Central United States: Analyses and Observations Based on Forest Inventory and Analysis Data. Journal of Forestry 113:325–334.

Mervis, J. 1 June 2016. Completing troubled ecological observatory will cost $35 million more: NEON says additional funds will mean finishing all 81 sites by end of 2017. Science. doi: 10.1126/science.aaf5829. https://www.science.org/content/article/completing-troubled-ecological-observatory-will-cost-35-million-more.

Meyer, D., E. Dimitriadou, K. Hornik, A. Weingessel and F. Leisch. 2021. e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. R package version 1.7-6. https://CRAN.R-project.org/package=e1071.

Moreira, Francisco, Pedro Vaz, Filipe Catry, and Joaquim S. Silva. 2009. Regional variations in wildfire susceptibility of land-cover types in Portugal : implications for landscape management to minimize fire hazard. International Journal of Wildland Fire 18 :563-574.

Moser, W.K., M. H. Hansen, R. L. Atchison, B. J. Butler, S. J. Crocker, G. Domke, C. M. Kurtz, A. Lister, and P. D. Miles. 2013. Kansas' forests 2010. In Resource Bulletin; NRS-85; USDA Forest Service, Northern Research Station: Newtown Square, PA, USA, p. 63.

Mountrakis, G., J. Im, and C. Ogole. 2011. Support vector machines in remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing, 66(3):247-259.

Nagy, R. C., J. K. Balch, E. K. Bissell, M. E. Cattau, N. F. Glenn, B. S. Halpern, N. Ilangakoon et al. 2021. Harnessing the NEON data revolution to advance open environmental science with a diverse and data-capable community. Ecosphere 12(12):e03833.

NEON (National Ecological Observatory Network). 2020a. Ecosystem structure (DP3.30015.001), RELEASE-2022. https://doi.org/10.48443/hd92-rr80. Dataset accessed from https://data.neonscience.org on December 8, 2022.

NEON (National Ecological Observatory Network). 2020b. Vegetation indices - spectrometer - mosaic (DP3.30026.001), RELEASE-2022. https://doi.org/10.48443/8mc6-pp97. Dataset accessed from https://data.neonscience.org on December 8, 2022.

Nguyen, H. T. T., T. M. Doan, E. Tomppo, and R. E. McRoberts. 2020. Land Use/land cover mapping using multitemporal Sentinel-2 imagery and four classification methods—A case study from Dak Nong, Vietnam. Remote Sensing, 12(9):1367.

Noble, B. and Z. Ratajczak. 2022. WPE01 Assessing the value added of NEON for using machine learning to quantify vegetation mosaics and woody plant encroachment at Konza Prairie ver 1. Environmental Data Initiative. https://doi.org/10.6073/pasta/a7b40e41080460bb1123dcc7b6d4d942 (Accessed 2022-12-08).

Parisien, Marc-André, Quinn E. Barber, Kelvin G. Hirsch, Christopher A. Stockdale, Sandy Erni, Xianli Wang, Dominique Arseneault, and Sean A. Parks. 2020. Fire deficit increases wildfire risk for many communities in the Canadian boreal forest. Nature Communications, 11:2121.

Pau, S., J. B. Nippert, R. Slapikas, D. Griffith, S. Bachle, B. R. Helliker, R. C. O'Connor, W. J. Riley, C. J. Still, and M. Zaricor. 2022. Poor relationships between NEON Airborne Observation Platform data and field-based vegetation traits at a mesic grassland. Ecology 103(2):e03590.

Pervin, R., S. M. Robeson, and N. MacBean. 2022. Fusion of airborne hyperspectral and LiDAR canopy-height data for estimating fractional cover of tall woody plants, herbaceous vegetation, and other soil cover types in a semi-arid savanna ecosystem. International Journal of Remote Sensing 43(10):3890-3926.

R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. V4.0.5. URL https://www.R-project.org/.

Radeloff, Volker C., David P. Helmers, H. Anu Kramer, Miranda H. Mockrin, Patricia M. Alexandrea, Avi Bar-Massada, Van Butsic, Todd J. Hawbaker, Sebastián Martinuzzi

Alexandra D. Syphard, and Susan I. Stewart. 2009. Rapid growth of the US wildland-urban interface raises wildfire risk. PNAS 115(13):3314-3316

Ratajczak, Z., J. B. Nippert, J. C. Hartman, and T. W. Ocheltree. 2011. Positive feedbacks amplify rates of woody encroachment in mesic tallgrass prairie. Ecosphere 2(11):121.

Ratajczak, Z., J. B. Nippert, J. M. Briggs, and J. M. Blair. 2014. Fire dynamics distinguish grasslands, shrublands and woodlands as alternative attractors in the Central Great Plains of North America. Journal of Ecology, 102(6):1374-1385.

Ratajczak, Z., J. M. Briggs, D. G. Goodin, L. Luo, R. L. Mohler, J. B. Nippert, and B. Obermeyer. 2016. Assessing the Potential for Transitions from Tallgrass Prairie to Woodlands: Are We Operating Beyond Critical Fire Thresholds? Rangeland Ecology and Management 69:280-287.

Ratajczak, Z., S. L. Collins, J. M. Blair, and J. B. Nippert. 2022. Reintroducing bison results in long-running and resilient increases in grassland diversity. PNAS 119(36):e2210433119.

Reid, A. M., Fuhlendorf, S. D., & Weir, J. R. 2010. Weather variables affecting Oklahoma wildfires. Rangeland Ecology & Management, 63(5), 599-603.

Riley County Fire District No 1. "Two out-of-control fires burn thousands of acres in Riley County." 10 April 2023a. https://www.rileycountyks.gov/CivicAlerts.aspx?AID=1582

Riley County Fire District No 1. "Riley County signs disaster declaration for wildfire near Indian Hill Road." 18 April 2023b. https://www.rileycountyks.gov/CivicAlerts.aspx?AID=1586

Riley County Fire District No 1. "More than 50 firefighters battle wildfire near Tuttle Creek lakeshore in Northern Riley County" 19 April 2023c. https://www.rileycountyks.gov/CivicAlerts.aspx?AID=1587

Riley County Fire District No 1. "Several fires break out in Riley County over the weekend." 24 April 2023d. https://www.rileycountyks.gov/CivicAlerts.aspx?AID=1588

Riley County Planning and Development Environmental Health Annual Report. 2021. https://www.rileycountyks.gov/DocumentCenter/View/1607/2021-Annual-Report-pdf?bidId=

Stambaugh, Michael C., Richard P. Guyette and Joseph Marschall. 2013. Fire History in the Cherokee Nation of Oklahoma. Human Ecology 41(5):749-758.

Samson, Fred, and Fritz Knopf. 1994. Prairie Conservation in North America. BioScience 44(60):418-421.

Scholl, V. M., M. E. Cattau, M. B. Joseph, and J. K. Balch. 2020. Integrating National Ecological Observatory Network (NEON) airborne remote sending in in-situ data for optimal tree species classification. Remote Sensing 12:1414.

Scott, Joe H., Elizabeth D. Reinhardt. Assessing Crown Fire Potential by Linking Models of Surface and Crown Fire Behavior. Research Paper RMRS-RP-29. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. September 2001.

Sheykhmousa, M., M. Mahdianpari, H. Ghanbari., F. Mohammadimanesh, P. Ghamisi, and S. Homayouni. 2020. Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 13:6308-6325.

Stein, Susan M., James P. Menakis, Mary A. Carr, Sara J. Comas, Susan I. Stewart, Helene Cleveland, Lincoln Bramwell, and Volker C. Radeloff. 2013. Wildfire, wildlands, and people: understanding and preparing for wildfire in the wildland-urban interface—a Forests on the Edge report. General Technical Report RMRS-GTR-299. Fort Collins, CO. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 36 p.

Swengel, A. B. 1996. Effects of fire and hay management on abundance of prairie butterflies. Biological conservation 76(1):73-85.

Thanh Noi, P. and M. Kappas. 2018. Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. Sensors 18(1): 18.

The National Ecological Observatory Network (NEON): FY 2022 Budget Request to Congress. https://www.nsf.gov/about/budget/fy2022/pdf/66j_fy2022.pdf

Toth, C., and G. Jóźków. 2016. Remote sensing platforms and sensors: A survey. ISPRS Journal of Photogrammetry and Remote Sensing 115:22-36.

Twidwell, D., W. E. Rogers, S. D. Fuhlendorf, C. L. Wonkka, D. M. Engle, J. R. Weir, U. P. Kreuter, and C. A. Taylor Jr. 2013. The rising Great Plains fire campaign: citizens' response to woody plant encroachment. Frontiers in Ecology and the Environment, Online Special Issue 11.

United Nations Environment Programme (UNEP). 2022. Spreading like Wildfire – The Rising Threat of Extraordinary Landscape Fires. A UNEP Rapid Response Assessment. Nairobi.

United States Census Bureau. "Annual Estimates of the Resident Population: April 1, 2010 to July 1, 2014 – United States – Metropolitan and Micropolitan Statistical Area; and for Puerto Rico." March 2015.

USDA Geospatial Data Gateway. 2019a. NAIP Orthorectified Kansas Color Infrared 161. https://nrcs.app.box.com/v/naip/file/579854034024

USDA Geospatial Data Gateway. 2019b. NAIP Orthorectified Kansas Natural Color 161. https://nrcs.app.box.com/v/naip/file/578936034378

Van Auken, O. W. 2009. Causes and consequences of woody plant encroachment into western North American grasslands. Journal of environmental management 90(10):2931-2942.

Whiteman, G., and J. R. Brown. 1998. Assessment of a method for mapping woody plant density in a grassland matrix. Journal of Arid Environments 38:269-282.

Zou, C. B., D. Twidwell, C. H. Bielski, D. T. Fogarty, A. R. Mittelstet, P. J. Starks, and B. S. Acharya. 2018. Impact of eastern redcedar proliferation on water resources in the Great Plains USA—current state of knowledge. Water 10(12):1768.

Appendix A

Table A.1: Best SVM model inputs calculated in R. Kernel refers to the method of data transformation. Degree is the degree of the polynomial kernel function. Gamma is the kernel coefficient which determines how much curvature will be in the decision boundary. Cost helps control error, a lower cost will accept a lower number of misclassified pixels.

Image	Kernel	Degree	Gamma	Cost
NAIP	Radial	3	0.1111	1
NEON	Radial	3	0.125	1
NAIP+NEON	Radial	3	0.0588	1

Table A.2: Time to train models.

Model Run Time	NAIP	NEON	NAIP+NEON
SVM	4:49:00	1:43:00	1:37:00
RF	0:30:00	0:23:00	1:05:00

Table A.3: Time to use models to predict (classify) entire study site

Model Run Time	NAIP	NEON	NAIP+NEON
SVM	6:15:25	3:05:08	2:04:59
RF	0:09:59	0:10:43	0:07:20

Appendix B

Table B.4: Confusion Matrix of LULC map accuracy

	Agriculture	Buildings	Deciduous Trees	Grassland	Gravel	ERC	Other	Pavement	Shrub	Water	User Accuracy
Agriculture	6722	11	61	147	2	18	62	0	1	5	**0.956**
Buildings	0	5956	5	0	42	12	57	196	0	6	**0.949**
Deciduous Trees	80	15	7100	70	25	340	87	13	443	0	**0.869**
Grassland	170	1	168	5978	7	39	61	0	257	0	**0.895**
Gravel	0	40	26	7	3769	4	36	114	0	0	**0.943**
ERC	15	0	211	16	1	8488	21	0	22	3	**0.967**
Other	29	32	17	8	27	3	1941	13	2	6	**0.934**
Pavement	0	130	11	0	193	3	27	6088	0	0	**0.944**
Shrub	1	0	503	149	0	46	29	0	5560	0	**0.884**
Water	0	1	0	0	0	0	0	0	0	6486	**0.999**
Producer Accuracy	**0.958**	**0963**	**0.876**	**0.938**	**0.927**	**0.948**	**0.836**	**0.948**	**0.885**	**0.997**	**0.933**

Table B.5: Landscape Metrics

Neighborhood	Location	Area buffered (ha)	% ERC Cover	Avg. Patch Size (m^2)	Cohesion	Contiguity	Average Euclidian nearest neighbor
1	Interior	403	14.3	73	94.6	0.25	6.4
2	Outside	575.1	4.9	89	94.6	0.26	8.2
3	Interior	350.1	9.9	55	95.1	0.22	7.1
4	Outside	1337.8	9.3	128	96.8	0.27	7.5
5	Edge	1090.1	13.3	84	97.2	0.25	6.5
6	Edge	1232.5	6.3	48	94.1	0.23	7.2
7	Outside	1152.7	5.5	47	88.7	0.25	6.8
8	Edge	2022.7	15.3	107	97.1	0.25	6.7
9	Edge	796.9	23.0	177	98.8	0.25	6.5
10	Inside	867.4	9.2	49	91.0	0.25	6.7
11	Edge	1124	11.2	60	94.0	0.25	6.6
12	Edge	1641.2	11.9	103	97.0	0.25	7.1
13	Inside	1138.2	6.2	40	91.3	0.23	7.3
Average	-	-	10.8	81	94.7	0.25	6.9
Manhattan	-	2033.7	9.1	95	97.3	0.24	7.3

Table B.6: Proportion of houses in each danger zone

Distance from structures (m)	≤4	≤4	≤4	≤20	≤20	≤800
ERC Patch size (m^2)	≥10	≥1000	≥5000	≥1000	≥5000	≥5000
1	0.58	0.01	0	0.08	0.03	1
2	0.75	0.11	0.03	0.25	0.06	1
3	0.88	0.31	0.23	0.58	0.5	1
4	0.83	0.18	0.1	0.38	0.22	1
5	0.93	0.14	0.12	0.19	0.15	0.66
6	0.81	0	0	0.002	0.0005	0.77
7	0.95	0.08	0.001	0.12	0.004	0.77
8	0.92	0.07	0.03	0.16	0.06	1
9	0.92	0.23	0.18	0.35	0.25	1
10	0.89	0.09	0.02	0.19	0.04	0.85
11	0.9	0.18	0.02	0.27	0.09	1
12	0.86	0.03	0.004	0.06	0.02	0.99
13	0.74	0.001	0.001	0.005	0.005	0.63
Avg.	0.84	0.11	0.07	0.2	0.11	0.9
MHK	0.81	0.06	0.02	0.1	0.045	0.81